Bernardo de Azevedo Pinto Castro Maciel

Bachelor of Science

Predicting the Outcome of Cognitive Training in Parkinson's Disease using Magnetic Resonance Imaging

Dissertation submitted in partial fulfillment
of the requirements for the degree of

Master of Science in
Biomedical Engineering

Co-advisers: Dr. Chris Vriend, Assistant Professor,
Amsterdam UMC, vrije Universiteit
Prof. Dr. José Manuel Fonseca, Associate
Professor, NOVA University of Lisbon

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

February, 2021

Predicting the Outcome of Cognitive Training in Parkinson's Disease using Magnetic Resonance Imaging

To Chris, Odile, and Tim, for making this possible.

To Rita, Daniel, Pedro and David, for making it possible.

To my parents, for making me possible.

Acknowledgements

First and foremost, I would like to thank and acknowledge the tremendous help from Dr. Chris Vriend, who deserves all the credit and none of the blame. I pay my most sincere gratitude to Chris for being more than an advisor. Secondly, I am ineffably indebted for the guidance that greatly exceeded my expectations from Prof. dr. Odile van den Heuvel. Thirdly, I must thank Prof. Dr. José Fonseca for the dutiful feedback and advice. And finally, I would like to thank the rest of the ANW team, particularly to Tim van Balkom for all the hospitality, time and availability.

On a personal note, the road to the completion of this work was tumultuous. I have to thank my friends who kindly sheltered me during the beginning of the Covid-19 pandemic. Friends like Rita, Pedro, Daniel, and David, are truly invaluable. They have been a never-ending source of emotional support during an especially rough time. Last but not least, I could not end this section without thanking my parents and paying homage to my grandmother. I do not know who I would be without them.

"We want things we cannot have. We seek to reclaim a certain moment, sound, sensation. I want to hear my mother's voice. I want to see my children as children. Hands small, feet swift. Everything changes. Boy grown, father dead, daughter taller than me, weeping for a bad dream. Please stay forever, I say to the things I know. Don't go. Don't grow."

- Patti Smith, M Train

Abstract

Motivation: Cognitive impairment is an important symptom of Parkinson's Disease (PD), usually having a substantial negative impact on the quality of life of patients, families, and caregivers. Cognitive Training (CT) have been proven effective in halting the process of cognitive decline in PD. However, the efficacy of CT is unpredictable from subject to subject.

Objective: Investigate the possibility of predicting the outcome of CT in PD patients with Mild Cognitive Impairment using structural and functional Magnetic Resonance Imaging (MRI) data.

Methods: Before CT, a sample of 42 PD patients underwent structural and functional MRI. Graph measures were then extracted from their structural and functional connectomes and used as features for random forest (RFo) and decision tree (DT) machine learning (ML) regression algorithms with and without prior latent component analysis (LCA). CT response was evaluated by assessing the outcomes of the Tower of London task pre- and post-treatment. Finally, the 4 ML models were used to predict CT response and their performances were assessed. *Post hoc* analyses were conducted to investigate whether these algorithms could predict age using connectomic measures on a sample of 80 PD patients.

Results: The performances of the aforementioned algorithms did not differ significantly from the baseline performance predicting the subject-specific CT outcome. The performance of the RFo without LCA differed significantly from the baseline performance in the age prediction task for the sample of 80 patients.

Conclusion: Notwithstanding the lack of statistical significance in predicting our

cognitive outcomes, the relative success of the age prediction task points towards the potential of this approach. We hypothesise that bigger sample sizes are needed in order to predict the outcome of CT using ML.

Keywords: Parkinson's Disease; Cognitive Training; Connectomics; Machine Learning.

Resumo

Motivação: O défice cognitivo é um sintoma comum da Doença de Parkinson (DP), tendo frequentemente um impacto negativo substancial na qualidade de vida dos pacientes, das famílias e dos cuidadores. O Treino Cognitivo (TC) tem mostrado promessas na retardação no défice cognitivo nos doentes com DP (dDP). No entanto, a eficácia do TC é imprevisível de sujeito para sujeito.

Objetivo: Investigar a possibilidade de previsão da eficácia do TC em dDP com um défice cognitivo ligeiro, intrumentalizando técnicas de imagem funcional e estrutural por ressonância magnética (IRM).

Métodos: Uma amostra de 42 dDP foi sujeita a estudos imagiológicos com IRM antes do TC. Foram extraídas medidas dos grafos definidos pelos conectomas estruturais e funcionais dos dDP para serem usadas como características para algoritmos de aprendizagem automática. Quatro algoritmos de regressão diferentes foram implementados: árvores de decisão e random forest (RFo) com e sem análise de componentes latentes (ACL) prévia das características. A resposta ao TC foi analisada através do desempenho no Teste da Torre de Londres (TTL) antes e após o TC. Por fim, os 4 modelos de AA foram utilizados para prever a melhoria no desempenho do TTL. Foram realizadas análises post hoc para avaliar a capacidade de previsão de idade dos dDP.

Resultados: Os algoritmos não apresentaram um erro significativamente menor que o utilizado como linha de base para prever a diferença do desempenho dos doentes no TTL antes e após o TC. O algortimo RFo com ACL teve um desempenho significativamente diferente da linha de base a prever a idade dos pacientes no grupo com 80 dDP.

Conclusão: Não foi possível prever as medidas de desempenho cognitivo. No entanto,

o sucesso na previsão da idade dos dDP sugere que esta técnica tem potencial para obtenção de resultados positivos se for utilizada uma maior amostra de estudo.

Palavras-chave: Doença de Parkinson; Treino Cognitivo; Conectómica cerebral; Aprendendizagem Automática.

Contents

List of Figures

List of Tables

Glossary

Angle of Precession	The angle between the axis of rotation and the axis of Precession [1]
Bias-Variance Trade-off	The bias-variance trade-off refers to the a trade-off between the complexity of the model and the amount of variance explained. If the model is too complex, it will fit the data used to construct the model very well but generalize poorly to unseen data. In other words, if the model is too complex, it is prone to Overfitting. Conversely, if the complexity is too low the model will be too simple to explain the variability of the data leading to underfitting [2]
Complex Network Analysis	The description of the proprieties of a system through the calculation of topological measures of its network representation [3]
Configuration	The set of necessary and sufficient Hyperparameters to completely define an algorithm [4]
Connectome	A complete description of the anatomical connections between all elements of a nervous system. This description can be detailed at different resolution scales, from the level of individual neurons and synapses to macroscopic connectivity between large-scale brain regions. The term connectome is also often used to describe the organization of functional connectivity networks [5]

Dementia	Acquired cognitive impairment has become severe enough to compromise social and/or occupational functioning. This diagnostic requires substantial impairment to be present in one or (usually) more cognitive domains. The impairment must be sufficient to interfere with independence in everyday activities. According to the DSM-5 nomenclature, dementia is now referred to as Major Neurocognitive Disorder [6]
Diamagnetic	Diamagnetic materials are repelled by a magnetic field; an applied magnetic field creates an induced magnetic field in them in the opposite direction, causing a repulsive force [7]
Effective connectivity	The causal influence that one neuronal system exerts over another. Typically inferred from a model of the neuronal interactions that cause fluctuations in measured neurophysiological signals. A model of effective connectivity results in a directed graph [5]
Functional connectivity	A statistical dependence between neurophysiological signals. Note: Functional connectivity between a pair of nodes does not necessarily imply that one node is causing activity in the other, or that the nodes share a direct anatomical connection [5]
Hyperparameter	In machine learning, a hyperparameter is a parameter that is set before the learning process begins, as opposed to what is simply defined to as a Parameter. Hyperparameters affect the sensitivity of the algorithms to detecting patterns, the Bias-Variance Trade-off, the trade-off between model complexity and fitting of the data, or may trade-off computational complexity for optimality of fitting. [4]
Lewy Bodies	Abnormal aggregates or clusters of the protein α-synuclein which can be present in the brain in a range of neurologic diseases [8]

Neuroplasticity	The ability of the brain to undergo structural and functional alterations by altering neurotransmission, synaptogenesis and neurogenesis from birth to old age [9, 10]
Nutation	A forced displacement of the axis of a spinning body away from the simple cone-shaped figure, which would be traced by the axis during Precession [1]
Overfitting	A model overfits the training data when it describes features that arise from noise or variance in the data, rather than the underlying distribution from which the data were drawn. Overfitting usually leads to loss of accuracy on out-of-sample data [11]
Paramagnetic	Paramagnetic materials are weakly attracted by an externally applied magnetic field, and form internal, induced magnetic fields in the direction of the applied magnetic field [7]
Parameter	In machine learning, a model parameter is a configuration variable that is internal to the model and whose value can be estimated from the given data, as opposed to what is referred to as a Hyperparameter.They are required by the model when making predictions and their values define the skill of the model on your problem.[4]
Parkinsonian Disorders	Medical conditions of which Parkinsonism is a prominent clinical feature [12]
Parkinsonism	Clinical syndrome characterized by bradykinesia, tremor, rigidity and postural instability [12]
Perfusion	In the context of MRI, perfusion refers to the capillary blood supply to a tissue, measured in ml min^{-1} g^{-1} [13]
Precession	Compound motion of a rotating object about an axis other than its axis of rotation. [1]
Radiomics	High-throughput mining of quantitative features from radiographic images [14]

Structural connectivity	The anatomical connections between neural elements. Equivalent to individual axons and synaptic contacts at the micro scale. At coarser scales, structural connectivity refers to the axonal tracts or white matter fiber bundles between different brain regions [5]

List of Abbreviations, Acronyms and Initialisms

T_1-W	T_1-weighted
ACT	Anatomically-Constrained (probabilistic) Tractography
BNA	Brainnetome Atlas
BOLD	Blood Oxygenation Level Dependent
BOM	Bayesian Optimisation Methods
CI	Confidence Interval
CNA	Complex Network Analysis
CT	Cognitive Training
CV	Cross-Validation
CWT	Continuous Wavelet Transform
DEC FA	Directionally Colour Encoded Fractional Anisotropy
DF-F	Decision Tree Algorithm with Factors as Features
dMRI	Diffusion Magnetic Resonance Imaging
DT	Decision Trees
DT-N	Decision Tree Algorithm with Graph Measures as features
DTI	Diffusion Tensor Imaging
DWI	Diffusion Weighted Imaging
EEG	Electroencephalography

FA	Fractional Anisotropy
FID	Free Induction Decay
fMRI	Functional Magnetic Resonance Imaging
FSL	FMRIB Software Library
i.i.d.	Independent and Identically Distributed
LB	Lewy Bodies
LCA	Latent Component Analysis
LN	Lewy Neurites
MEG	Magnetoencephalography
ML	Machine Learning
MPRAGE	Magnetisation-Prepared Rapid Acquisition Gradient-Echo
MRI	Magnetic Resonance Imaging
MRT	Mean Reaction Time
MSE	Mean Square Error
NMR	Nuclear Magnetic Resonance
pBOLD-ts	Preprocessed Blood Oxygenation Level Dependent Time-Series
PCAT	Percentage of Correctly Answered Trials
PD	Parkinson's Disease
PD-MCI	Parkinson's Disease Related Mild Cognitive Impairment
PD-D	Parkinson's Disease Related Dementia
PGSE	Pulsed Gradient Spin Echo
QoL	Quality of Life
RCI	(Iverson) Reliable Change Index
RCT	Randomised Controlled Trial
REM	Rapid Eye Movement
RF	Radio-frequency

RF-N	Random Forest Algorithm with Graph Measures as features
RF-F	Random Forest Algorithm with Factors as Features
RFo	Random Forrest
RMSE	Root Mean Square Error
SE	Spin-Echo
SIFT	Spherical-deconvolution Informed FilTering
sMRI	Structural Magnetic Resonance Imaging
SVM	Support Vector Machines
ToL	Tower of London
TPE	Tree-structured Parzen Estimator

Symbols

All the symbols listed below are used recurrently. The symbols that are only used on a single context are defined in the main text upon their instantiation.

$\vec{B_0}$	Static Magnetic Field
$\vec{B_1}$	Non-stationary Magnetic Field
$\vec{M_0}$	Macroscopic/Net Magnetisation
$\vec{\mu}$	Magnetic Dipole Moment
$\vec{J}$	Total Angular Momentum
$\vec{I}$	Nuclear Spin/Intrinsic Angular Momentum

T_1	Longitudinal relaxation time
T_2	Ideal transverse relaxation time
T_2^*	Real transverse relaxation time
TR	Repetition Time
TE	Time to Echo
TI	Inversion Time
θ	Angle of Precession
ω_0	Larmor Frequency
γ	Gyromagnetic Ratio
$\Delta\epsilon$	Energy Difference
T	Temperature
k_B	Boltzmann Constant
$\hbar$	Reduced Plank Constant
b	b-value

τ_0	Pre-intervention time point
τ_1	Post-intervention time point
ν_1	Frequency Band of $[0.06, 0.12]$ Hz
ν_2	Frequency Band of $[0.01, 0.06[$ Hz
F_1	Functional connectome derived from the calculation of wavelet coherence with a bandpass of ν_1
F_2	Functional connectome derived from the calculation of wavelet coherence with a bandpass of ν_2
S_{FA}	Structural connectome derived from the fraction of anisotropy map
S_S	Structural connectome derived from fibre count

$\mathscr{D}$	Data Set
$\mathcal{L}$	Tuning Set
$\mathcal{L}^*$	Training/Learning Set
$\mathscr{T}$	Test Set
$\mathscr{T}^*$	Validation Set
$\mathscr{H}$	Hypothesis Space

Chapter 1

Introduction

Parkinson's Disease is a progressive neurodegenerative disease characterised by a plethora of symptoms. It's a complex and heterogeneous disease that is infamous for its typical motor symptoms, i.e. tremor at rest, rigidity, bradykinesia and postural instability. Furthermore, Parkinson's Disease (PD) also comprises a wide range of non-motor symptoms which include but are not limited to autonomic dysfunction, sensory and sleep abnormalities, and cognitive impairment. [15]

Cognitive impairment is an important symptom of PD, usually having a substantial negative impact on the quality of life of patients, families, and caregivers [16]. The cognitive dysfunctions in PD range from mild cognitive impairment to dementia and include visuospatial impairment, attentional set-shifting difficulties, working memory impairment, decline of attention ability, inflexibility in thinking, and difficulty planning [17]. Parkinson's Disease Related Mild Cognitive Impairment (PD-MCI) is an umbrella term referring to an abnormal decline in multiple cognitive abilities which is either reported by the patient (or informant) or observed by the caregiver or clinician, and is not caused by other diseases [18, 19]. Studies have reported that 84% of all PD patients suffer from cognitive impairments [15] and 25% show signs of significant cognitive impairments at diagnosis [6, 18]. Moreover, PD-MCI has been described (and often is considered) as a degree of cognitive impairment that is not normal for age and a potential prodromal state to Parkinson's Disease Related Dementia (PD-D) [20]. In this way, PD-MCI is thought

of as an intermediate state between normal cognition and dementia in PD [16]. In fact, PD-D is estimated to affect 50% of PD patients (with varying times of onset and severity). [18, 21] The risk of developing dementia has been shown to be six times higher in PD patients than in the general elderly population [22]. Therefore, the study of treatment for cognitive impairment in PD may provide means to alleviate the current burden of PD-MCI and PD-D.

Several approaches have been explored to relieve symptoms and halt cognitive decline. In particular, non-pharmacological interventions such as Cognitive Training (CT) have been proven effective in halting the process of cognitive decline in PD [23] as well in other diseases such as Alzheimer's disease [24]. CT is thought to have beneficial effects on cognition due to the activation of neuroplastic mechanisms [16]. Some studies have suggested that PD patients benefit from tailored short-term training of executive functions (e.g. attention, planning) [17] and that CT improves memory and overall cognition in PD patients [25]. Despite the promises of efficacy of CT in PD, no large Randomised Controlled Trial (RCT)s have been conducted hitherto [25]. Two large RCT protocols were published. However, these were directed to patients that have already developed PD-MCI [16] (n = 111, on intervention group and, n = 111, control group) or PD-D (n = 15 on cognitive rehabilitation therapy group, n = 15 on relaxation therapy group, n = 15 on the control group) [25].

This dissertation will follow from the COGTIPS study, which is the largest (n = 140) RCT conducted on PD patients to date and to our knowledge. COGTIPS aims to provide evidence for the efficacy of an easily-accessible, home-based online cognitive training protocol, to investigate the potential long-term effects of computerised CT and to shed light on the underlying cognitive mechanisms of PD. Neuroimaging data will be instrumental to reach this goal [26].

Neuroimaging data provides a non-invasive method of investigation of the human brain connectivity on both functional and structural levels [5]. In particular, this data allows the study of several important proprieties of the brain through its representation as a complex network. Furthermore, it is possible to infer quantifiable measures related to the flux of information of an individual's brain (viz. integration, segregation, motif, centrality, and resilience) from these networks. [3]. There is growing body of literature

about differences in neural network connectivity (connectomic) on PD when compared against healthy subjects. Studies [27–31] suggest that lower measures of integration and segregation of the brain's connectivity system are related to cognitive decline (when compared with healthy controls).

Currently, the effect of a given treatment is heterogeneous and unpredictable from subject to subject [25]. However, some previous studies suggest that complex network analysis of neuroimaging data can be a powerful tool in treatment outcome prediction [32–36]. Adding to the evidence of network changes after CT [10, 37–39], this dissertation aims to establish whether connectomic measures can be used to predict subject-specific treatment outcome in patients with Parkinson's Disease. In order to reach our goal, we will apply latent factor analysis and machine learning tree-based regression models to data extracted from structural and functional Magnetic Resonance Imaging (MRI) scans to predict cognitive performance after training in a sample of 42 PD patients.

Chapter 2

Theoretical concepts

2.1 Parkinson's Disease

Parkinson's Disease is a neurological neurodegenerative disease. It is the most common form of Parkinsonism. A person diagnosed with PD should display idiopathic parkinsonism, which must not suspected to be secondary to any other causes [40]. Patients with PD experience significant comorbidities, including increased rates of infections, cardiac and gastrointestinal disorders, and fall-related injuries, in addition to (other) motor and non-motor symptoms [41].

The motor clinical findings in parkinsonian patients as described in [40] are: tremor, rigidity, hypokinesia, abnormal gait and posture. See figure 2.1 for visual cues. The parkinsonian **tremor** is present at rest, increases with emotional stress and decreases with activity. The tremor frequency usually falls in the interval of [4, 6] Hz and it can be on one limb, on both limbs of one side or on all four limbs. It is usually associated with a hand movement called "pill-rolling" (cf. figure 2.1). **Rigidity**, or hypertonia, affects both agonist and antagonist muscles around a particular joint. This increase in the resistance of passive movement causes the infamous flexed posture as well as cogwheel-like movement (severe interruptions in the flow of the movement). **Hypokinesia**, bradykinesia or akinesia, is the a slowness of voluntary movement and a reduction in automatic movement. This symptom manifests itself in many observable features such as in general abnormal

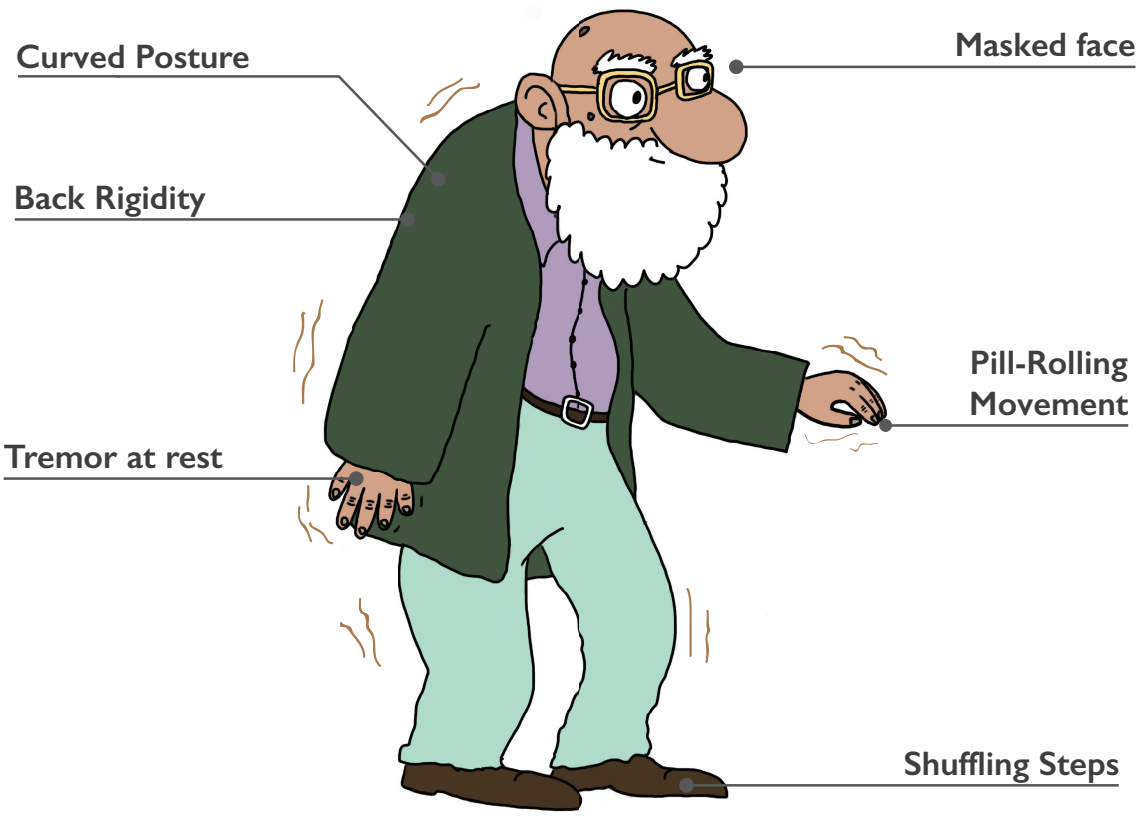

Figure 2.1 – Parkinson's motor symptoms. See text and legend in image for more in-depth description.

slowness of movement, infrequent blinking, soft voice, immovable expression (mask face) and small handwriting, among others. The parkinsonian **gait** is characterised by a flexed posture, slow starts, small and shuffling steps, little arm swing, difficulty to stop and turn. In advanced disease stages, the posture may be so altered that there is involuntary acceleration in order to prevent falls (festination). Typically the motor symptoms have an asymmetric onset.

The non-motor symptoms of PD include but are not limited to autonomic dysfunction, mood disorders (namely depression and anxiety), sleep disorders (typically Rapid Eye Movement (REM) sleep disorders), sensory abnormalities and neuro-cognitive dysfunctions [15]. The latter will be discussed on section 2.1.3. Close monitoring of signs and symptoms like hyposmia, constipation, anxiety, depression, and REM sleep behaviour

disorder is fundamental. These signs may be noticeable in a preclinical stage spanning several years before the onset of symptoms of motor dysfunction [40].

2.1.1 Neuropathology

The pathology of PD is complex and still not fully understood. Idiopathic parkinsonism, or Parkinson's disease, is considered a proteinopathy. This syndrome is characterised by the misfolding and aggregation of α-synuclein thus being also referred to as a synucleinopathy [40].

α-Synuclein is a one-hundred-and-forty-aminoacid-long protein that seems to be involved in intercellular communication and mitochondrial function, among other functions. It is a natively unfolded protein found in presynaptic terminals. Abnormal cytological locations of this protein (i.e. outside presynaptic terminals) are thought to influence its conformation, making it more likely to form neurotoxic aggregates [42], called Lewy Bodies.

Lewy Bodies (LB) are the most common neuropathological finding in PD. These are reported to be present in around 80% of patients [12]. An (initially small) abnormal protein folding can spread from neuron to neuron in something referred to as a prion-like spread. However, LB disease is not unique to PD it is also a common clinical finding in other Parkinsonian Disorders [12]. Similarly, Lewy Neurites (LN) are abnormal projections in diseased neurons. They contain granular material and abnormal α-synuclein filaments similar to those found in LB. In early-stage PD, the LB pathology is considered to be relatively limited to the brainstem and limbic regions. With disease progression and with the development of PD-D, the LB and LN also spread to the neocortex [6].

One of the most well-known models for the pathophysiology of the motor manifestations of PD connects these symptoms to dopamine depletion effects. In general lines, this model describes the motor symptoms as a consequence of decreased availability of dopamine in regions and pathways of the cortico-basal ganglia-thalamocor-tical motor circuit [43][1]. With the LB and LN propagation throughout the neocortex, the non-motor

[1] This a gross simplification of the pathophysiology of PD and pathways. For more detailed information on the direct and indirect pathways, cf. [43], section 2.2.

symptoms become more notorious. In fact, studies present evidence for the motor symptoms onset when the loss is 30% of substantia nigra dopamine neurons or 50 to 70% of nigrostriatal dopaminergic axonal terminals in the striatum [43]. Conversely, the non-motor symptoms can already be present in a preclinical stage thus being of paramount importance in the early-onset detection of PD [40].

2.1.2 Epidemiology and Burden

PD is about 1.4 times more frequent in men than in women and the prevalence increases with age [44] (See figure 2.2). The global prevalence is conservatively estimated to be around 0.3% and it increases sharply with age to more than 3% in those older than 80 years [42]. The incidence of Parkinson's disease seems to be highly variant globally [44]. In particular seems to vary in subgroups defined by race, location, ethnicity, genotype or environment [42]. Parkinson's disease is the neurological disease with the fastest growing burden [45], which has more than doubled over 26 years, from 2.5 million patients in 1990 to 6.1 million patients in 2016 [44, 46]. Moreover, the increase in mortality is bigger than the increase in prevalence [44]. There is a multitude of factors that can account for this increase. To name a few, better study methods and greater general awareness for the disease lead to better estimates of the prevalence. Also, increasing life expectancy (also of PD patients) leads to a higher prevalence. Finally, the increase of environmental factors tied to the growing industrialisation of the world greatly influences incidence rates [44].

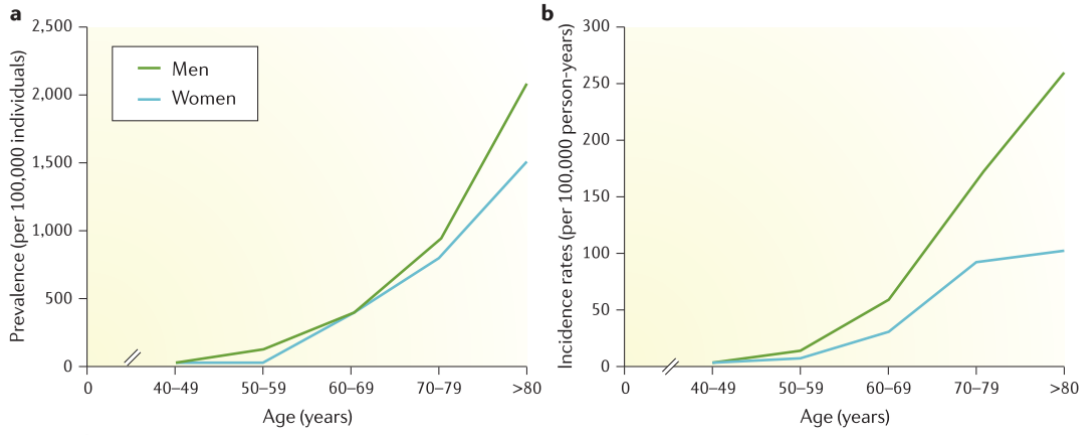

Figure 2.2 – Incidence and prevalence of Parkinson disease. **a** | Prevalence of Parkinson disease in men and women per 100,000 individuals. **b** | Incidence rate of Parkinson disease per 100,000 person-years. Image from [42].

There is no doubt that a prevalent neurodegenerative disorder with significant impact on Quality of Life (QoL) of patients and relevant comorbidies would entail a huge economical burden. PD is associated with substantial excess costs in direct medical costs, indirect productivity losses, non-medical costs, and disability income. Yang et al. [41] estimated that in 2017 in the US the total economic burden was of $51.9B (≈€43.9B) with $25.4B (≈€21.4B) from direct medical care costs and $26.5B (≈€22.4B) from indirect (non-medical) costs.[2] The European Parkinson's disease Association estimated [47] the economic burden of PD to be of almost €14B in 2011.

2.1.3 Cognitive decline in Parkinson's Disease

The rate, severity and time of onset of cognitive impairment is an important factor for differential diagnosis of different Parkinsonian Disorders [6]. Studies have reported that PD patients have a six-fold increase in the probability of developing Dementia when compared with the general population [22]. Seventy eight percent of PD patients are expected to develop dementia in the first eight years after the onset of the motor symptoms and this percentage goes up to 84% if we consider the first twenty years after onset [20]. Adding to that, one out of four PD patients suffers from significant cognitive impairment at the time of diagnosis [18, 21] and that 25% to 50% of non-demented PD patients suffer from some cognitive impairment in the course of their disease [20]. Henceforth a cognitive impairment on non-demented PD patients not secondary to any other cause we will referred to as PD-MCI.

Dementia is an acquired, generalised, and usually progressive impairment of cognitive functions with preserved level of consciousness (unlike in coma and confusional states) [40]. PD-D and PD-MCI are closely related. PD-MCI represents a degree of cognitive impairment that is not normal for age [20]. It is a potential prodromal state to PD-D being often considered an intermediate state between normal cognition and dementia in PD [20].

Definition of the clinical phenotype specific to PD-MCI is complex because this is a highly heterogeneous syndrome. Suboptimal performance of cognitive domains such as executive function, psychomotor speed, visuospatial abilities, language, and memory has been noted in non-demented PD as well as dysfunction in executive function. The

[2] B stands for billion

clinical features of PD-MCI are heterogeneous with and single and multiple-domain impairment and with amnestic or non-amnestic characteristics. Non-amnestic single domain cognitive impairment predominates in PD-MCI. Moreover, PD-MCI is usually associated with older than average age at screening and onset of motor symptoms, male sex, depression, and more severe motor symptoms [20].

Notwithstanding the heterogeneity of clinical features, the Movement Disorder Society task force on PD-MCI has defined an operative definition for this condition to standardise research and clinical practice. The formal diagnosis of PD-MCI is based on the accomplishment of four key features [48][3]:

- diagnosis of Parkinson's disease;
- evidence of a gradual cognitive decline reported by the patient, relative, or clinician;
- validation of this decline through PD-validated scales of global cognitive performance and/or comprehensive neuropsychological assessment;
- the cognitive decline is not severe enough as to significantly interfere with functional independence.

2.2 Magnetic Resonance Imaging

Undoubtedly, the development of non-invasive neuroimaging techniques is one of the major factors contributing to the understanding of some of the inner dynamics of the human brain. Namely, the complex brain network representation and the mapping of brain connectivity (at increasingly smaller scales) are allowing new and exciting advances in neurosciences [49]. In particular, the developments in MRI have made possible structural and functional imaging and data acquisition with increasing spatial and temporal resolutions [3]. MRI will be the imaging modality instrumentalised in this dissertation work. However, other techniques have been successfully used to extract connectomes. Two popular examples are Electroencephalography (EEG) and Magnetoencephalography (MEG).

[3]Only inclusion criteria listed here. See [48] for exclusion criteria.

This section is based on chapters 3 and 8 of [13] and chapters 23 and 24 of [50]. References to more detailed content follow as footnotes where relevant information was omitted because it was considered out of scope for this dissertation work.

2.2.1 Nuclear Magnetic Resonance

Nucleons (protons and neutrons) have an intrinsic magnetic momentum, also called spin. This property finds no adequate classical parallel but it can be thought of as analogous to the angular momentum of a spinning sphere. To simplify this discussion, we will henceforth consider a single proton; i.e. H^+.[4]

Regarding the nature of the atomic magnetic interaction, it is important to state that a non-zero angular momentum is always associated with a non-zero magnetic dipole moment, $\vec{\mu}$, through the following relation:

$$\vec{\mu} = \gamma \vec{J} = \gamma \hbar \vec{I} \tag{2.1}$$

where γ is a proportionality constant called gyromagnetic ratio, $\hbar$ is the reduced Planck constant ($\hbar = 1.054 \cdot 10^{-34}$ J·s), and $\vec{I}$ and $\vec{J}$ are the nuclear intrinsic and total angular momenta, respectively. This means that a charged particle (i.e. a proton), which has a non-zero nuclear intrinsic angular momentum $\vec{I}$, can be conceptualised behaving like a tiny rotating magnet.

Let $\vec{M}$ be the net/macroscopic magnetisation defined by the theoretical vectorial sum of $\vec{\mu}$ for all the particles in a given sample (see figure 2.4). In the vast majority of biological tissues, the directions of the microscopic dipole moments of particles are randomly distributed. Therefore, $\vec{M}$ is usually zero. With the application of a static magnetic field, $\vec{B_0}$, the individual magnetic moments of the particles align with this magnetic field, creating a non-zero $\vec{M}$. However, due to the quantisation of the angular moment, the proton $\left(\|\vec{I}\| = \pm\frac{1}{2}\right)$ can't perfectly align with $\vec{B_0}$. Classically, provided that $\vec{\mu}$ and $\vec{B_0}$ do not have the same direction, the torque exerted by $\vec{B_0}$ is non-zero, provoking a change in the angular moment $\vec{J}$. Thus, $\vec{\mu}$ exhibits a Precession movement. This precession occurs with a frequency ω_0, called the Larmor frequency and described by the Larmor equation

[4] This simplification is not naive taking into account that the MRI signal is mostly due to hydrogen atoms (which nuclei are a single proton). Hydrogen is abundant in the human body, which is 70% to 80% H_2O, and produces a relatively strong signal when compared against other atoms.

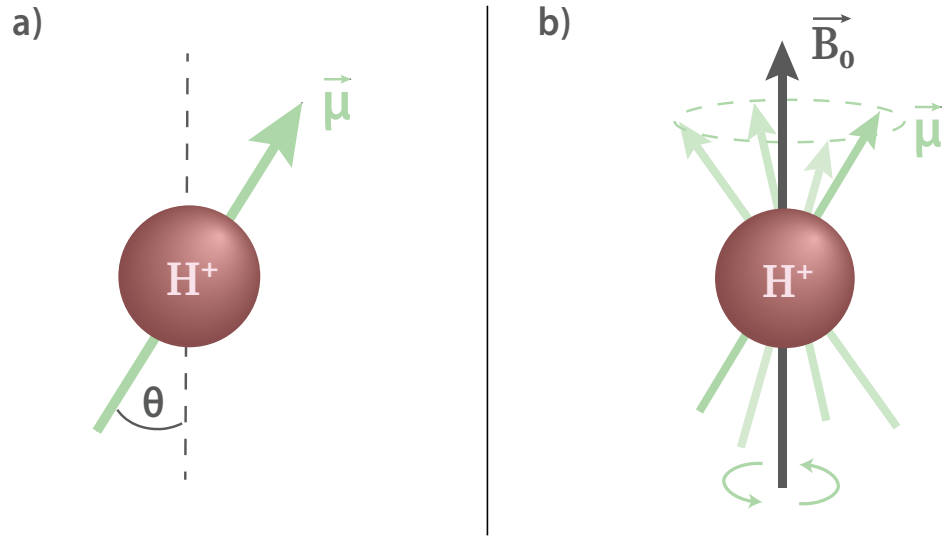

Figure 2.3 – **a** | Classical representation of the proton as a vector defined by its magnetic dipole moment, $\vec{\mu}$. The angle between the axis of rotation and $\vec{\mu}$ is defined as θ, the angle of precession. **b** | Classical representation of the precession movement caused by the interaction of the static uniform magnetic field $\vec{B_0}$ with the proton's dipolar magnetic moment.

(B_0 is the magnitude of $\vec{B_0}$). [5]

$$\omega_0 = \gamma \vec{B_0} \tag{2.2}$$

Additionally, quantum mechanic theory postulates that the interaction between $\vec{\mu}$ and $\vec{B_0}$ is stable when aligned with the latter both in parallel and anti-parallel orientations[6] (See figure 2.3). For a large number of protons, the population of both of these energy states (i.e. the number of protons that are oriented parallel and anti-parallel to $\vec{B_0}$) is given by the Boltzmann ratio, derived from the Boltzmann distribution:

$$\frac{N_{\text{up}}}{N_{\text{down}}} = \exp\left(\frac{\Delta\epsilon}{k_B\, T}\right) \tag{2.3}$$

where N_{up} and N_{down} represent the number of protons oriented parallel and anti-parallel to $\vec{B_0}$, respectively, $\Delta\epsilon$ is the difference of energy between the two states, T is temperature and k_B[7] is the Boltzmann constant.

[5] For more detailed derivation of the Larmor equation, cf. [13] pp. 137-139.

[6] Discussions of the Zeeman effect are considered out of scope for the purposes of this thesis. For more detailed information, cf. [50] pp. 207.

[7] $k_B = 1.39 \cdot 10^{-23}$ J·K^{-1}

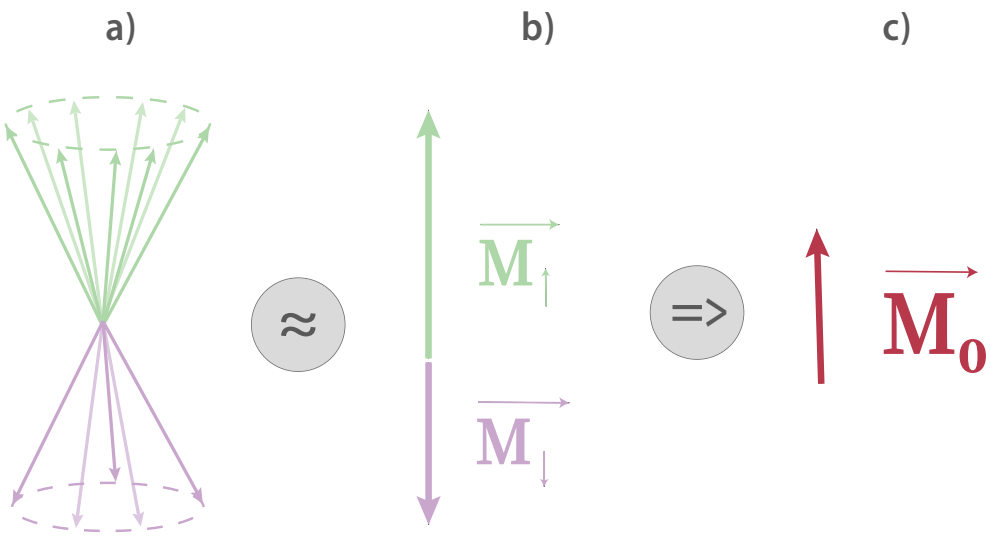

Figure 2.4 – Visual representation of the net magnetisation vector. **a** | A classical approach to the precession of the magnetic dipolar moments of the different atoms in a sample in a static magnetic field. In green the magnetic moments that are oriented parallel to the field and in pink anti-parallel. **b** | A simplification of representation of the different magnetic dipoles as the magnetisation vectors. The different magnetisation vectors for both energy states a vector with the direction of the field and magnitude given by the summation of the dot product between every individual dipole moment and the magnetic field. **c** | Because there are more atoms in the lower energy state (represented by green) at the body temperature (as described by equation 2.3), the net magnetisation is a non-null vector.

Using equation 2.3 to calculate the Boltzmann ratio for protons at body temperature (e.g. $T = 310\text{K}$) on a uniform magnetic field of 1 T, we have a surplus of ≈ 3 ppm of atoms in the lower energy state. For this reason, as the atoms align according to a preferential direction, their dipoles sum constructively. This effect accounts for a small but potentially measurable macroscopic net magnetisation to which we call $\overrightarrow{M_0}$ (see figure 2.4).

The precession phase of $\vec{\mu}$ from the different nuclei in the sample under the influence of field $\overrightarrow{B_0}$ is random. Therefore, the only non-null component of the net magnetisation vector $\overrightarrow{M_0}$ is along the direction of the magnetic field. See figure 2.4 b). We define the direction of the uniform static magnetic field $\overrightarrow{B_0}$ as the z-axis and the component of $\overrightarrow{M_0}$ along that direction as $\overrightarrow{M_z}$. Likewise, the two orthogonal components of this vector (that define the precession plane xOy) are defined as $\overrightarrow{M_x}$ and $\overrightarrow{M_y}$. We will now proceed our analysis using a classical interpretation of the net magnetisation, $\overrightarrow{M_0}$, as described before. Henceforth, we will consider the vector representation without the arrow as the

magnitude of such vector, e.g. $M_z \equiv \|\overrightarrow{M_z}\|$.

2.2.1.1 Interaction with a Radiofrequency Pulse

Let us consider what would happen to this system if we apply a Radio-frequency (RF) pulse with frequency ω_0. This external magnetic field will be referred to as B_1. This pulse induces Nutation - the change of the precession angle θ. In particular, we will focus on the situation of θ equals 90°. B_0 is usually in the order of magnitude of 10^0 T while M_0 can be roughly estimated around 10^{-5} T for the human head[8]. This would make the detection of $\overrightarrow{M_0}$ impossible in the direction of $\overrightarrow{B_0}$. To bypass this, $\overrightarrow{M_0}$ is detected by magnetic induction on a coil perpendicular to B_0, where B_0 has no influence and the signal from M_0 is maximum if we would tilt it to $\theta = 90°$. So, we arrive to the conclusion that a signal induced in a coil perpendicular to $\overrightarrow{B_0}$ is dependent on θ. Furthermore, the signal will be dependent on B_0 by the means of ω_0 (equation 2.2) because a higher precession frequency would mean a bigger variation in the magnetic flux thus inducing a stronger voltage according to Lenz's Law:

$$\epsilon = -\frac{\partial \phi_B}{\partial t} \tag{2.4}$$

on which ϵ is the electromotive force (measured in V) and ϕ_B is the magnetic flux.

Interacting with the spins using ω_0 also leads to the synchronisation of their precession movements. However, the nuclei experience the magnetic fields of their neighbours and lose coherence over time. So, M_x and M_y are non-null immediately after the interaction with B_1 but tend to zero over time. In this way, their magnitude is related to the precession movement and evolves over time as described by the Bloch equations (equations 2.5, 2.6 and 2.7).

$$M_x(t) = M_0 \sin(\omega_0 t) \cdot \exp\left(\frac{-t}{T_2}\right) \tag{2.5}$$

$$M_y(t) = M_0 \cos(\omega_0 t) \cdot \exp\left(\frac{-t}{T_2}\right) \tag{2.6}$$

where T_2 is a constant referred to as transverse relaxation time. Moreover, as mentioned B_1 causes Nutation. However, the magnetisation vector will tend to its position of equilibrium in alignment with B_0:

$$M_z(t) = M_0 \left[1 - \exp\left(\frac{-t}{T_1}\right)\right] \tag{2.7}$$

[8] Cf. [13] pp. 140.

where T_1 is the longitudinal relaxation time.

We now have a notion of how to generate a measurable signal using Nuclear Magnetic Resonance (NMR). In the next subsection we will discuss briefly how this signal translates into a picture.

2.2.2 From Proton to Picture

As discussed previously, it is possible to interact with atoms (with non-null nuclear spin) using non-stationary magnetic fields, i.e. B_1. The simplest possible sequence is a single pulse. If applied with the correct duration, this field can provoke a 90-degree nutation of M_0. To the magnetisation signal produced after a 90-degree pulse we call Free Induction Decay (FID) (figure 2.5). It is characteristically the product of a sinusoid and an exponen-

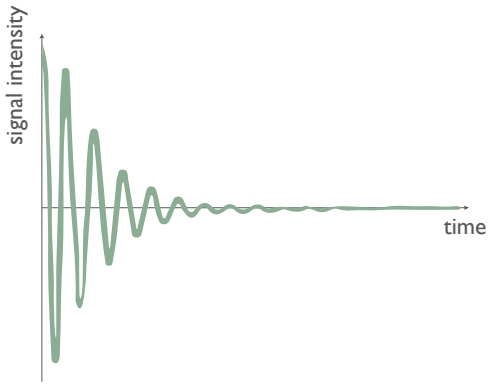

Figure 2.5 – Representation of FID plot. Axes in arbitrary units.

tial curve parametrised by T_2. However, $\overrightarrow{B_0}$ is not completely homogeneous. Magnetic field inhomogeneities across the sample (e.g. presence of ions, polar molecules, etc) introduce local variations in the frequency of precession within the sample volume which cause an artificial reduction of the transverse relaxation time. While T_2 represents what we will henceforth referred to as ideal transverse relaxation time, to this new constant we call T_2^*, the real transverse relaxation time.

T_2 describes the dephasing of spins due to spin-spin interaction with their neighbours and T_2^* accounts also for the dephasing produced by inhomogeneities of the static magnetic field within a sample. However, using only a FID pulse, there is no way to identify the extent of alteration introduced by field inhomogeneities since the field is distorted

by those two effects simultaneously. Adding to that, technical limitations regarding the MRI equipment make it impossible to collect the FID reliably.[9] In order to circumvent this problem, we use another sequence called Spin-Echo (SE).

As we discussed, with time the amplitude of the magnetic resonance signal is reduced because of interactions among spins and magnetic field inhomogeneities. The 90-degree pulse first tips these spins into the transverse plane. After the 90-degree pulse, a 180-degree pulse is applied. The 180-degree pulse allows refocusing of spins (of stationary nuclei) whose phases have been scattered by constant field distortions and inhomogeneities. At time TE, the spins are synchronised. See figure 2.6 for reference and more colloquial explanation.

The time between two different iterations of the whole SE sequence[10] is TR (repetition time). This value is also an important scan parameter because it controls the "initial angle" of $\overrightarrow{M_0}$ and thus its status after nutation.

As made evident by the previous explanations, the signal is highly influenced by the sequence of RF pulses applied to the sample as well as its T_1, T_2 and T_2^*, and by the chosen acquisition times (i.e. TE, TR). Provided that different body tissues have different relaxation times, it is possible to construct images with enhancement of different tissues. This is done by the choice of TE and TR when compared against the relaxation times of the tissues we wish to highlight. The strength intensity of the MRI signal during a SE sequence is given by the following expression:

$$S = N_{\mathrm{H}} \cdot \left[1 - 2 \cdot \exp\left(-\frac{\mathrm{TR} - 0.5 \cdot \mathrm{TE}}{T_1}\right) + \exp\left(-\frac{\mathrm{TR}}{T_1}\right)\right] \cdot \exp\left(-\frac{\mathrm{TE}}{T_2}\right) \tag{2.8}$$

where the N_{H} is a scale factor representing the proton/spin density of the tissue, i.e. the number of spins. The relation between the different parameters and the outcome image is summarised in figure 2.7.

Discussions of spatial encoding are considered out of the scope of this dissertation and not essential to understand the content. Suffice it to say that the spatial encoding of the signal and subsequent transform to an image is done by using a third magnetic field: the gradient field. This field gets its name from the fact that its strength is variable

[9] Cf. [1] pp. 369.

[10] Not only in between the SE sequence, but in between iteration of any sequence that was chosen for the scan.

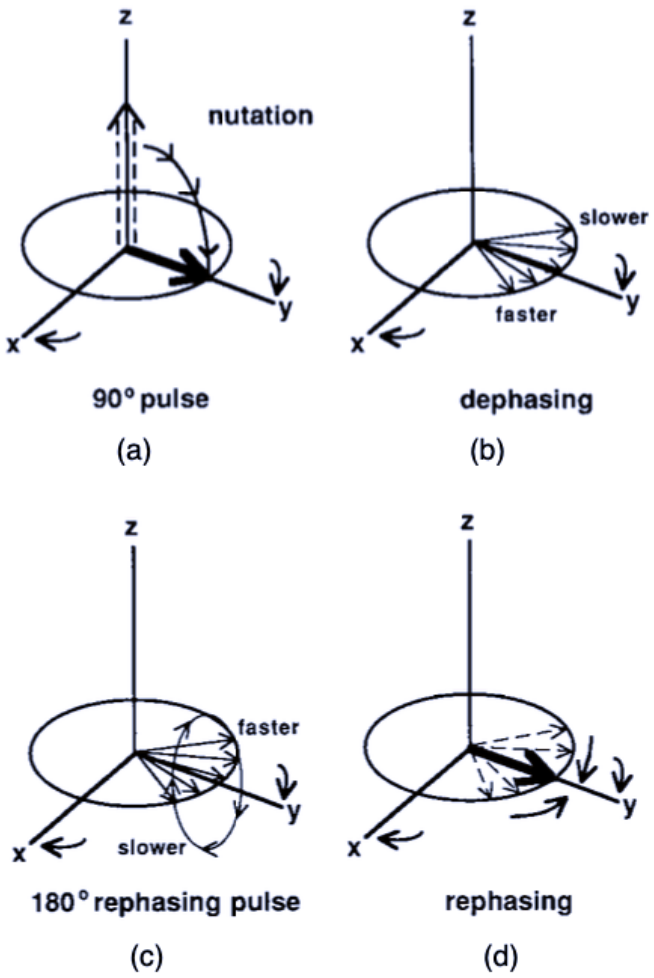

Figure 2.6 – Spin-Echo Sequence Representation. In the spin-echo pulse sequence the 90-degree pulse (A) causes nutation of the bulk magnetisation about the x-axis. Dephasing then occurs (B) as slower components of the magnetisation lag and faster components advance. A 180-degree rephasing pulse (C) causes nutation about the y-axis and reverses the order of fast and slow components. The components then merge (D) to produce the maximum signal. Diagrams A to D are shown in the rotating frame of reference. The faster spins initially rotate toward the viewer and the slower spins rotate away. Image from [1].

through space. Therefore it has a unique value in different points in space, influencing the spins differently. This differentiation makes it possible to identify "where the signal comes from"[11].

[11] For a more detailed discussion of the gradient field and of the spatial encoding techniques cf. [50] pp. 368-373.

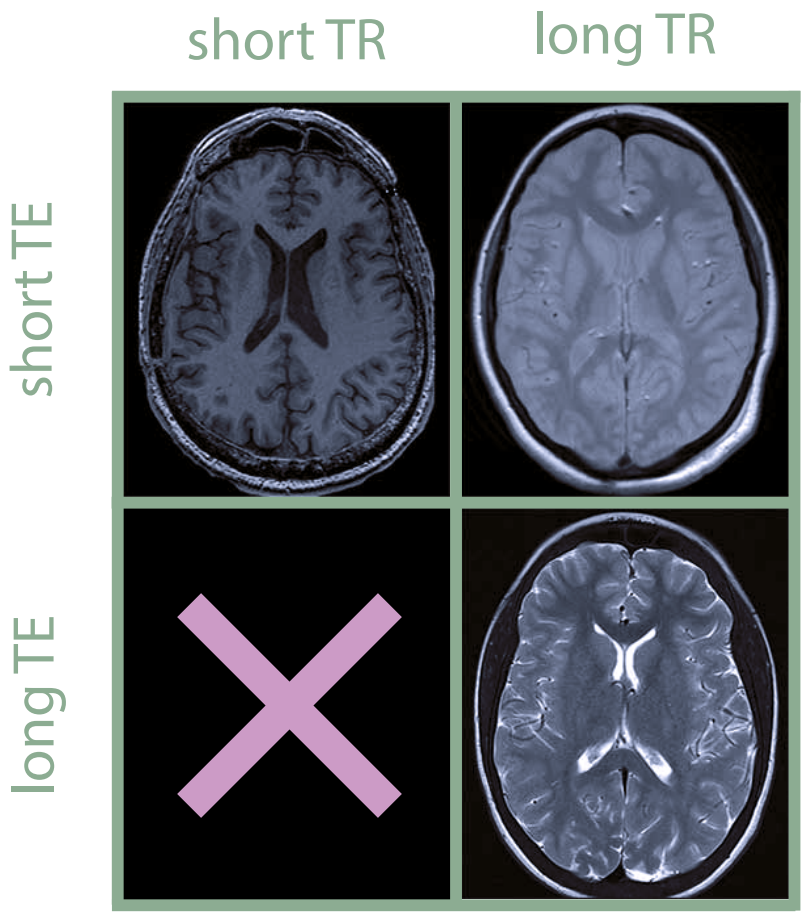

Figure 2.7 – Relationship between time to echo (TE), repetition time (TR) and contrast of the MRI. With short values for TR and TE we get a T_1-weighted image (top left). Long TR and short TE, a proton-density-weighted image is generated (top right). Long TE and short TR produces images with very low contrast-to-noise ratio thus being useless. Finally, long TR and short TE, allows the scan of a T_2-weighted image (bottom right). In this context, "long" and "short" TR and TE are defined in relation to the T_1 and T_2 of the tissue we wish to image, respectively. Images adapted from [52].

2.2.3 Diffusion MRI

Diffusion Magnetic Resonance Imaging (dMRI) is a family of techniques that use information derived from the diffusion of water molecules in the brain to measure the integrity of white matter fibre bundles and with that the strength of connections between brain areas. In particular, dMRI measures the net displacement of water molecules in a given (short) amount of time. Diffusion Tensor Imaging (DTI) is a model based on the quantification of the anisotropy of the water displacement [51]. In practice, this technique comprises the calculation of the diffusion tensor of each voxel (equation 2.9). The inference of the diffusion tensor is based on the measurement of the changes of the MRI signal in

various directions over a period of time and subsequent comparison with a non-diffusion-weighted MRI image. Then, it is possible to estimate the principal direction of diffusion by calculating of eigenvalues and eigenvectors of the diffusion tensor. These principal directions on each voxel provide the means to perform fibre tracking (or tractography) on the brain, yielding information about cerebral tissue microstructure [52, 53].

Pure water is said to diffuse freely, i.e. water molecules have equal probability of wandering off in any direction at any point in time. Hence, free diffusion leads to a zero-mean Gaussian distribution of particle displacements. However, in the brain, diffusion of water molecules is hindered by structures such as white matter fibres and neurons. The diffusion restricted by cylindrical fibres results in the existence of a preferential direction for diffusion: along the fibres [13, 51, 53].

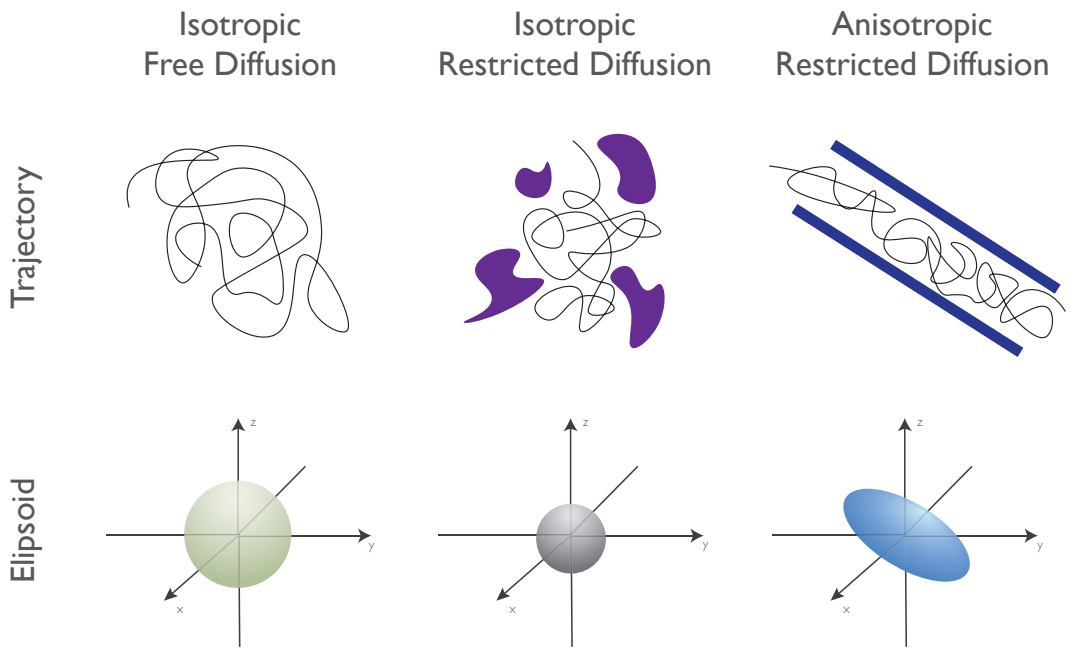

Figure 2.8 – Comparison between free and restricted diffusion. **Top row**| Representation of the diffusion trajectory performed by a particle in the three following situations: Isotropic Free Diffusion (left), Isotropic Restricted Diffusion (middle) and Anisotropic Restricted Diffusion (right). In the middle representation, the purple bodies represent randomly scattered impermeable objects. On the right, the blue rectangles represent regular impermeable barriers (e.g. axonal fibres). **Bottom row**| The ellipsoids associated with the trajectory in the same column.

When diffusion is anisotropic, it is insufficient to refer to the diffusion as a scalar.

Therefore, we use a second order symmetric tensor, the diffusion tensor.

$$D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix} \tag{2.9}$$

The diffusion tensor can be easily derived from diffusion-weighed measurements performed along different directions. The more measurements in different direction, the more accurate the estimation of D. The three diagonal elements represent diffusion coefficients measured along each of the principal laboratory axes. The six off-diagonal terms reflect the correlation of random motions between each pair of principal directions. If the diffusion is isotropic, the tensor is a scalar matrix[12]. We obtain a natural visual representation of the diffusion tensor by plotting an ellipsoid with the length of principal axes described by the tensor eigenvalues, which we will refer to as λ_1 for the principal axis and λ_2 and λ_3 for the secondary axes [54]. See figure 2.8 for the relationship between the type of diffusion and the shape of the correspondent diffusion ellipsoid.

It is also important to quantify directional anisotropy. The Fractional Anisotropy (FA) is a normalised measure for how much the eigenvalues of the diffusion tensor differ.

$$\mathrm{FA} = \sqrt{\frac{3}{2}} \cdot \sqrt{\frac{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2}{\lambda_1^2 + \lambda_2^2 + \lambda_1^3}} \tag{2.10}$$

where $\bar{\lambda}$ is the average of the three eigenvalues. Note from the equation above that an FA value of zero corresponds to perfectly isotropic diffusion and a value of one diffusion only along the principal axis. By multiplying the FA values from different voxels in the dMRI scans and a directional (colour) encoded map, we can obtain a visual representation of the directions of brain tracts (figure 2.9). One voxel can yield millions of fibres; therefore, this method is highly prone to trace spurious tracts (i.e. false positives). It is possible to mitigate some of these false positive connections if the anisotropy of diffusion is accounted for. Weighing the image with the FA map allows us not to trace fibres on voxels showing near-to-isotropic diffusion, which probably correspond to false positive white mater tracts. The resulting map we call Directionally Colour Encoded Fractional Anisotropy (DEC FA) maps. [55]

One of the most common methods for weighing images on diffusion uses the Pulsed Gradient Spin Echo (PGSE) sequence. First, the excitation pulse (figure 2.10a) rotates

[12] By scalar matrix it is meant that $D = \alpha I, \alpha \in \mathbb{R}$.

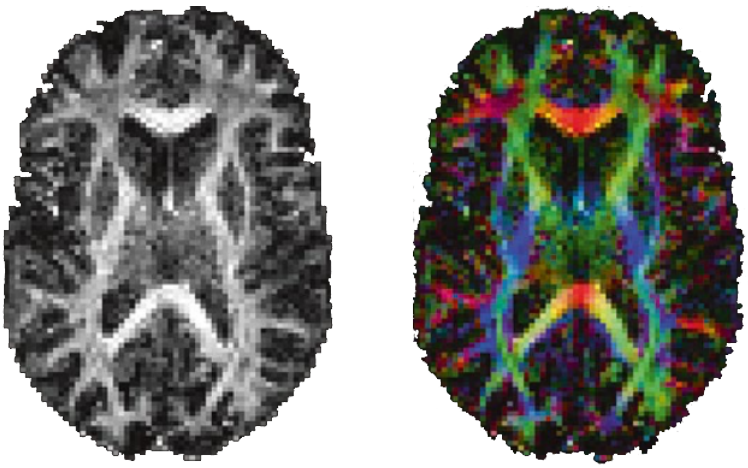

Figure 2.9 – **Left**| FA map. **Right**| DEC FA map. The colours represent the orientation of the (normalised) eigenvectors. As the eigenvector itself is specified relative to the original (Ox, Oy and Oz) axes, the meaning of the colours is: pure red the Ox axis direction, pure green green is Oy, and blue is Oz. Any colour in between is a combination of x, y and z in the adequate proportions. Image adapted from [55].

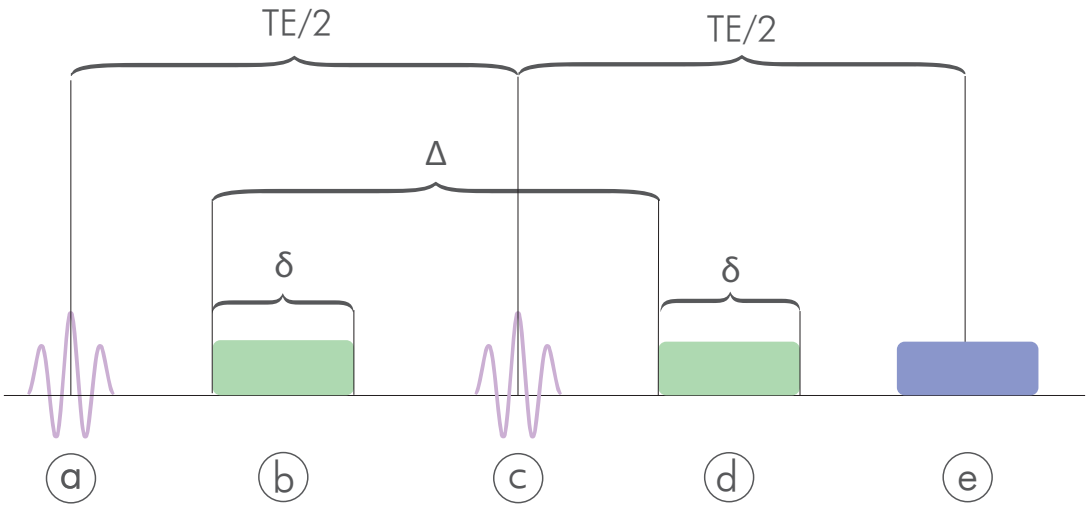

Figure 2.10 – PGSE schematic. **a** | Excitation Pulse. **b** | Gradient Pulse. **c** | Refocusing Pulse. **d** | Rephasing Gradient Pulse. **e** | Acquisition.

$\vec{M}$ onto the transverse plane. Next, the first pulsed gradient (figure 2.10b) dephases the spin magnetisation due to the variation of the frequency of precession along the gradient. Then, the refocusing pulse (figure 2.10c) rotates the magnetisation about the yy axis. Finally, the second pulsed gradient (figure 2.10d) rephases magnetisation differently for molecules that have changed position during the diffusion time. On the one hand, the net magnetisation of static molecules will be completely rephased. On the other, the spins of moving molecules have experienced a varying field due to the gradient, thus having

different frequencies of precession. Therefore, their signal will be attenuated because the pulse won't completely rephase the spins. The signal attenuation by this effect is described by the Stejskal-Tanner equation:

$$\frac{S}{S_0} = \exp\left[-\gamma^2 G^2 \delta^2 \left(\Delta - \frac{\delta}{3}\right) D\right] \tag{2.11}$$

where γ is the gyromagnetic ratio, G is the gradient strength, δ is the gradient duration, Δ is the time between the start of the first and second gradients, D, S is the signal, S_0 is the signal without diffusion weighting gradients. It is common to define a quantity $b = \gamma^2 G^2 \delta^2 \left(\Delta - \frac{\delta}{3}\right)$ [mm^2/s] [51]. The bigger the b-value, the greater the signal attenuation due to diffusion. By the comparison of the attenuated image with a non-attenuated one (called the $b = 0$ or zero b-value image) we can calculate how much voxel-wise diffusion has occurred in time TE thus having the possibility of generating an image with contrast dependent on the amount of diffusion [55].

In the context of this work, DTI refers to the model used to characterise the diffusion. Diffusion Weighted Imaging (DWI) refers to the final image resultant from from the application of DTI model to (multi-shell) dMRI data.

2.2.4 Functional MRI

Using Functional Magnetic Resonance Imaging (fMRI) it is possible to explore the variations in local Blood Oxygenation Level Dependent (BOLD) response. Imaging the BOLD response exploits the different magnetic proprieties of oxygenated and non-oxygenated haemoglobin caused by localised changes in Perfusion. Oxyhaemoglobin is Diamagnetic while deoxyhaemoglobin is Paramagnetic. Therefore, deoxyhaemoglobin has a shorter T_2^* than oxyhaemoglobin. Furthermore, the local variation of the ratio between the concentrations of oxyhaemoglobin and deoxyhaemoglobin allows to create an image modulated by the blood's oxygenation level. However, as we are measuring changes in blood flow, the fMRI signal (or, in layman's terms, brain activation) is a deconvolution of the haemodynamic response function from the BOLD signal [13]. See figure 2.11 for visual representation.

The fMRI image is of great interest after subjecting the brain to specific stimuli (i.e. block/event-related design) as it enables the inference of function through measures of brain activity. To clarify, the distinct patterns of signal strength variation after a specific

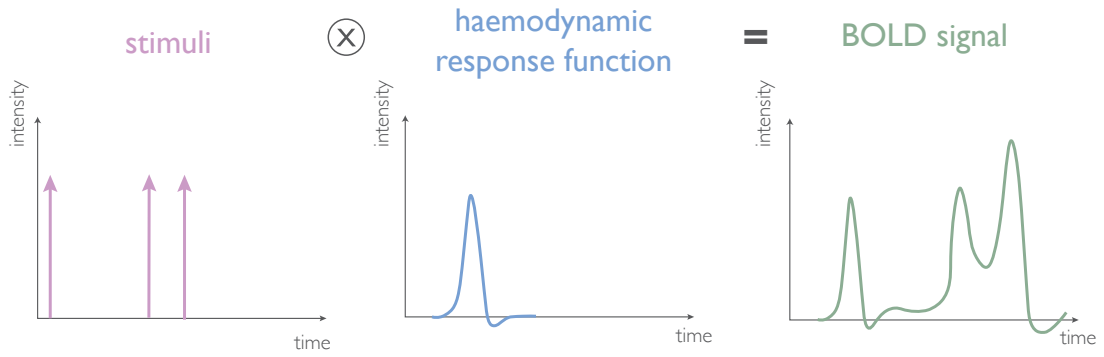

Figure 2.11 – Blood Oxygenation Level Dependent Response and the Haemodynamic Response Function Mock Example. The BOLD signal can be represented mathematically as a convolution between the stimulus and the haemodynamic response. Therefore, by deconvolution of the haemodynamic response function and the BOLD signal, we can study localised patterns of brain activation.

stimulus are associated with haemodynamic changes which are resultant of increased local neural activity due to said stimulus. Conversely, spontaneous correlations at rest have also been observed leading to the development of default (task-negative) networks. [52]

2.3 Connectomics

The conceptualisation of the brain as a network of connected (specialised) units has deep historical roots in neuroscience. Naturally, several attempts have been made to abstract the fundamental proprieties of structure and function of this network into a mathematical object. These efforts aim to build a Connectome: complete description of the connections between all elements of a nervous system [56].

The efforts to map human connectome started on the micro-scale (cell level). Ramon y Cajal's [57] microscopic studies of neurons (1955) theorised them to be individual units in close communication instead of a big continuous syncytial connection in between all of the cells as proposed previously by Golgi. The model of a body cell with axonal projections in close communication with similar neighbouring structures is well suited to be represented as a graph thus providing one of the earliest conceptual precedent for connectomics. Some attempts to map the macro-scale (i.e. systems of brain areas) connections between brain structure and pathology were happening in parallel. One notorious endeavour by Wernicke (1906) suggested an associative theory of brain function upon investigating the connection between brain damage and aphasia. It was proposed that

higher-order cognitive abilities (and their respective disorders) emerged from the integration (or pathological lack thereof) of anatomically distributed yet connected cortical areas [5].

The first attempts to represent a connectome were based on the neuronal studies of *Caenorhabditis elegans*, an invertebrate with 302 neurons and around 5000 synaptic connections between them. This system was completely mapped in 1986 by expert visual inspection of serial electron micrographs. Later (1998), Watts and Strogatz modelled the connectome of *C. elegans* as a binary graph, where the cell bodies of the neurons were represented as nodes and synapses as edges [58]. See figure 2.12 for a more recent illustration of the connectome of *C. elegans*. This work (among others) in the early days of the field of connectomics served as a proof of concept of the power of this technique for analyses on suitably simplified nervous systems. Circa 2001, graph theory was already being used to produce mesoscale connectomes from tract-tracing data of the macaque. Later, circa 2005, the first connectomes of humans started being drawn from MEG, EEG and MRI data [5].

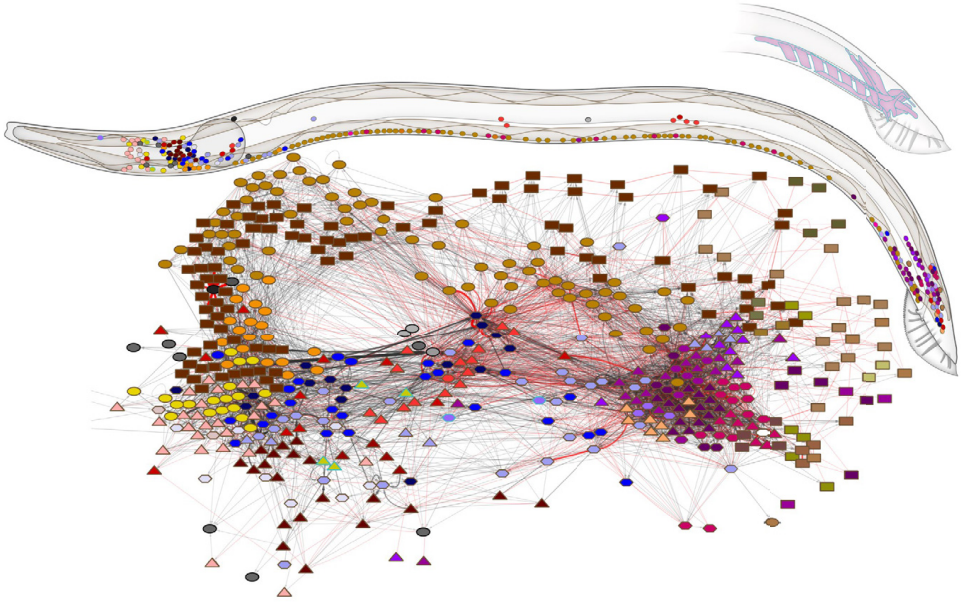

Figure 2.12 – Connectome of the adult male *Caenorhabditis elegans*. The worm diagrams show the locations of the cell nuclei represented in the connectome below. (left side and centre nuclei only, the right-side homologues of left/right pairs are not shown). Black edges represent synaptic junctions between neurons and red edges represent gap junctions. The colour and shape of the edges is related to the function/structure of the individual neurons (explanation out of scope for this thesis). Image adapted from [59].

Complex Network Analysis (CNA) is the description of the proprieties of a system through the calculation of topological measures of its network/graph representation. Simply put, a network is a set of connected mathematical objects (to which we refer to as nodes or vertices). In particular, the concept of brain connectivity finds its most natural mathematic representation as a network on which the nodes represent different brain regions and the edges the structural and/or functional connections between them [60]. Connectomes may differ in many ways. However, we will focus and define two key characteristics of a Connectome: the nature of the edges and the scale (or nature of the vertices).

Edges from networks derived from *in vivo* neuroimaging can be representative of three different, albeit deeply connected, types of connectivity: Structural connectivity, Functional connectivity and Effective connectivity. Firstly, structural connectivity is the description of a finite, albeit extensive, set of anatomical links between neural elements of the human brain organised on multiple (nested) spatial scales. [61] Secondly, functional connectivity describes undirected statistical dependencies (for example via correlation or coherence measures) between areas of the brain which may not be directly structurally connected [49]. Finally, effective connections correspond to directed causal relationships between two areas of the brain. These connectivity links describe the trade-off between invariant structural connectivity and context-sensitive functional connectivity using *a priori* knowledge [62] .

The abstraction of the connection between multiple areas of the brain as a network allows us to study many mathematical properties that have meaningful physiological interpretations. Rubinov and Sporns [3] outline several important proprieties of complex networks (viz. integration, segregation, motif, centrality, and resilience) and graph measurements that allow their quantification. Several CNA approaches to the brain's structural and functional connectivity data suggest that functional connectivity is deeply constrained by brain structure. There is evidence suggesting that the physical structure of the brain connectivity networks enables the efficient processing of information and allows the performance of complex brain functions. Furthermore, the structural organisation of the brain seems to be modular and hierarchical. In other words, the whole-brain network consists of a set of subnetworks, which themselves comprise multiple sub-modules.

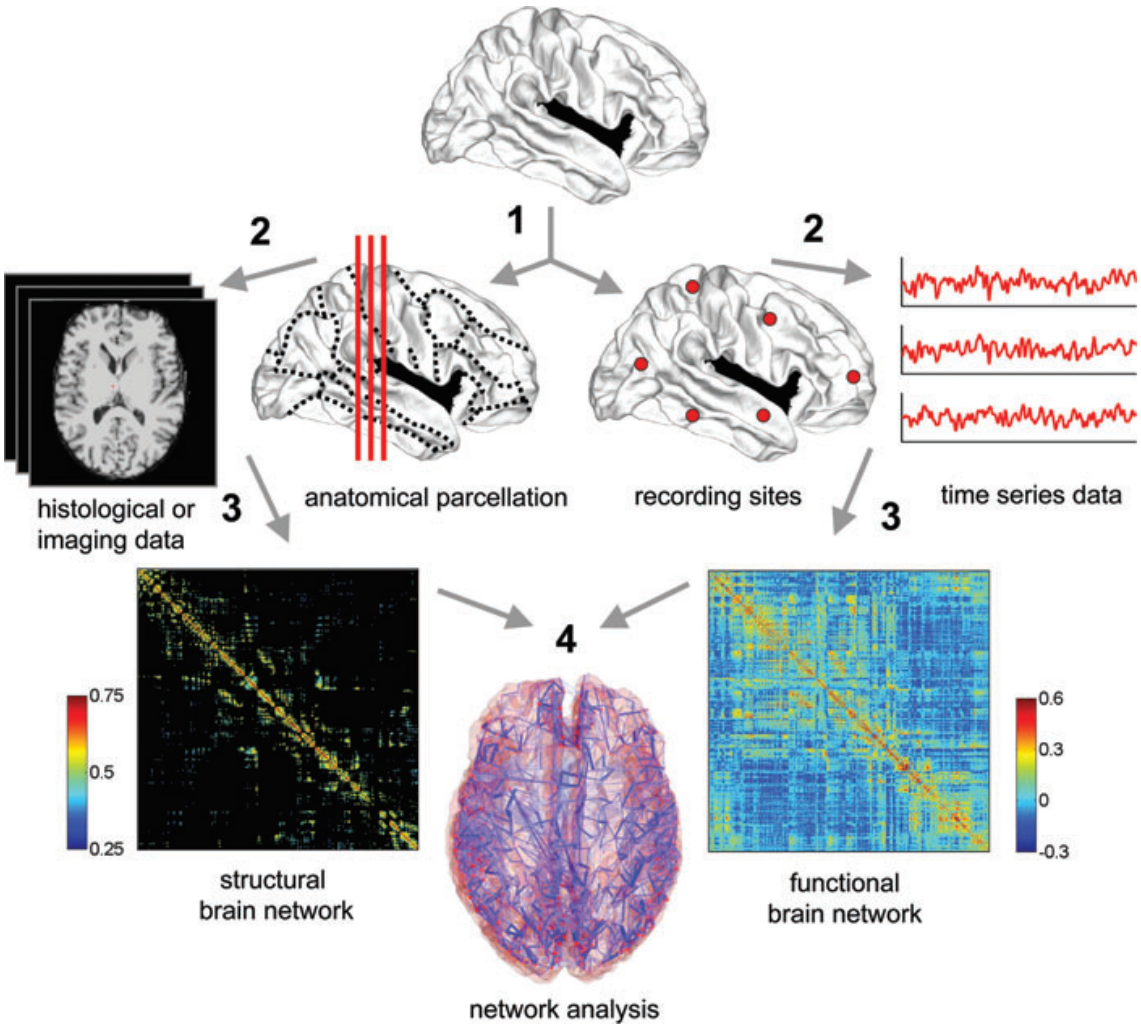

Figure 2.13 – Overview of the process of extraction of brain networks from empirical data. **1**| Parcellation of brain regions. **2**| Feature extraction based on Structural Magnetic Resonance Imaging (sMRI)/dMRI (left) or estimation functional activation based on time series data. **3**| Construction of structural (left) and functional (right) adjacency matrices based on empirical data. **4**| Network analysis. Image from [61]. Note: functional data is represented using electro/magneto-encephalography data, yielding information similar in nature to fMRI (but at a higher temporal resolution), which is the main focus of this work.

On the whole, CNA describes high cognitive functions as a global integration of information processed on a multitude local integrators [62]. Functional segregation is the ability of the brain to delegate specific tasks to specialised and densely interconnected groups of brain regions. Measures of segregation comprise clustering coefficients, local efficiency and modularity, among others. Functional integration represents the brain's ability to rapidly combine information from several different areas/networks. To quantify functional integration, characteristic path length and global efficiency are among the most

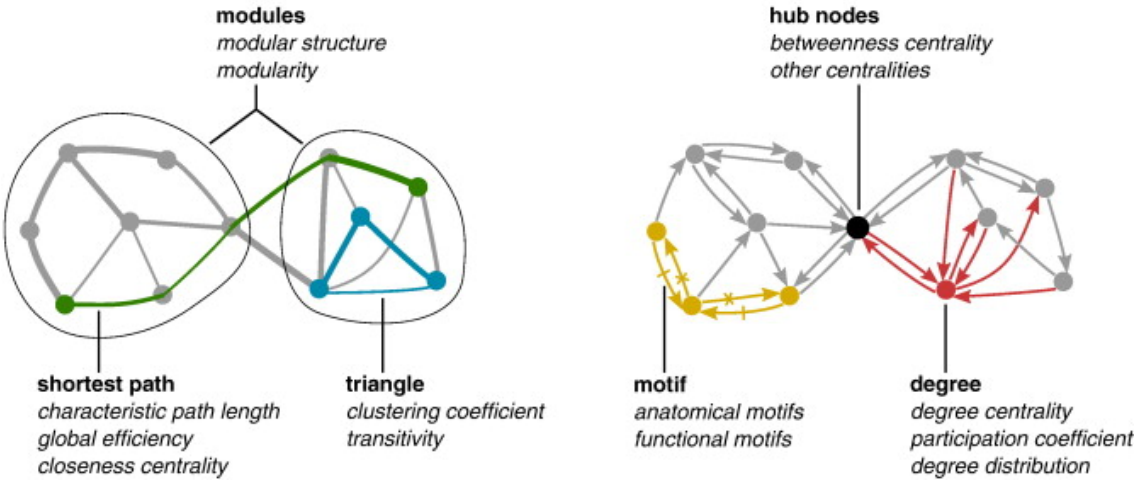

Figure 2.14 – Measures of network topology. Measures of centrality and resilience could be represented from degree-related measures (red) as well as other nodal measures (black). Segregation is deeply related with highly connected nodes (blue) while integration concerns the quantification of each node's connectivity length (green). Motifs (yellow) describe specific patterns of activation and are a key element of effective connectivity. Image from [3].

used graph measures. Centrality and resilience are also deeply connected. Measures of centrality (e.g. eigenvalue or betweenness centrality) describe a node's importance in the overall network architecture. The notion of node centrality plays a big role in the description of the network's resilience to insult. In this way, measures of resilience (e.g. degree centrality) quantify the vulnerability to insult. Motifs describe specific patters of activation, i.e. a subgraph of a structural or functional network with a particular topological configuration [3, 5, 60]. See figure 2.14 for a visual representation of some graph measurements. The complete list of all of the measures and respective definitions used in this work follows in Annex I.

2.4 Tree-Based Regression Models

Learning is essential in unknown environments. With the growing amount and complexity of available data for analysis, it may be impossible to predict all the possible outcomes of a given problem or changes over time in a reasonable timeframe. Machine Learning (ML) allows us to let a learning agent to learn information from the world of data we feed them in order to draw conclusions about situations it has not been exposed to before. If it improves its performance on future tasks after making observations about the world, we say the agent learnt. [63]

First, let us define some working nomenclature. This theoretical framework will describe the prediction of a single continuous real variable Y using a set of N continuous real variables (referred to as features in the context of ML), $X := \{\mathbf{x}_n,\ n = 1,...,N\}$. Analogous logic can be drawn for prediction of a multitude of continuous real variables, discreet variables or even categorical variables (the latter is referred to as a classification problem.)

A dataset $\mathcal{D}$ is a set such as $\mathcal{D} := \{(X, Y)\} = \{(\mathbf{x}_n, y_n),\ n = 1,...,N\}$. The hypothesis space, $\mathcal{H}$, is the set of all the possible values for a data point (i.e. the set of all hypotheses that might possibly be returned by the ML model). Furthermore, one element of Y is represented by y_i (note that it is a scalar). Likewise, $\mathbf{x}_\mathrm{i}$ refers to one row of matrix X. The boldface reiterates the fact that $\mathbf{x}_\mathrm{i}$ is not a scalar – it is a vector of features, to which we call an observation. To refer to the value of feature A in observation i we will use x_i^A.

Let $\mathcal{L}^*$ be a subset of $\mathcal{D}$ with L elements, $\mathcal{L}^* := \{(\mathbf{x}_n, y_n),\ n = 1,...,L\}$, where each y_n was generated by an unknown function, $y_\mathrm{i} \equiv f(\mathbf{x}_\mathrm{i})$. We refer to $\mathcal{L}^*$ as learning/training set. Learning is discovering a function $\hat{f}$ that approximates the true function f. The function $\hat{f}$ is called a hypothesis[13]. Learning is a search through the space of possible hypotheses, $\mathcal{H}$, for one that will perform well, even on new examples beyond the training set. To measure the accuracy of a hypothesis we give it a test set of examples that are distinct from the training set, $\mathcal{T}$. [63]

There are two main types of ML problems: classification and regression. The present thesis will only make use of the latter, thus there will be no discussion of classification problems and formalism.

Regression analysis is a set of statistical processes for estimating the relationships between a target variable and one or more independent variables (features). Regression models have four main components [64]:

- The target variable vector (Y). Y represents what we would like our model to be able to predict.
- The observation variables matrix (X). X represents the variables that will be used

[13] The symbols $\hat{f}$ and f will be used interchangeably provided that there is no ambiguity that we are not considering the unknown (and likely unknowable) real function f.

in the prediction, i.e. the observed data or the features.

- The vector of unknown parameters (β). The goal of our model is to find the β that best fits the data we have available. The nature of β is dependent on the specific model we are trying to fit. For instance, if our model is a simple linear regression,

$$y_i = f(\mathbf{x}_i, \beta) = \beta_0 + \beta_1 \cdot \mathbf{x}_i$$

- The error term (e). The error term is not observed directly from the data. It is rather a residual variable produced by the incapability of our model to explain the full variability of the data. It reflects how much the model predicts to differ from real data before empirical analysis.

In layman's terms, a regression problem assumes that the prediction variable(s), Y, can be explained by some model f that transforms the observed data X using the set of parameters β, leaving an unexplained residual e. Or, mathematically:

$$\forall(\mathbf{x}_i, y_i) \in \mathcal{D}, \quad y_i = f(\mathbf{x}_i, \beta) + e_i \tag{2.12}$$

In the following subsection, decision tree regression and random forest regression models will be described.

2.4.1 Decision Trees

Decision Trees (DT) are among the simplest, yet powerful, algorithms in machine learning. It is a fast and intelligible process, capable of working with categorical and continuous variables with little specific preprocessing. DT are also invariant to monotonous data transformations, which in turn ensure this algorithm is fairly resilient to monotonous data transformations. These characteristics result in an algorithm capable of dealing with data from different natures [65].

There are several algorithms to build a decision tree regressor but this work will only focus on the CART algorithm (see [66] for more information). A decision tree is a piecewise constant model over a set of regions $\mathbf{R}$. The set of all regions $\mathbf{R} := R_1, R_2, ..., R_m$ is called a partition scheme. In particular, for region R_m a decision tree f is defined such as:

$$f(\mathbf{x}) = c_m, \quad \forall \mathbf{x} \in R_m \tag{2.13}$$

where c_m is a constant determined for R_m. See figure 2.15 for a visual representation of equation 2.13. Fitting DT refers to finding the optimal values of c_m on the optimal partition scheme[14], $\mathbf{R}^*$, such as the Mean Square Error (MSE) (equation 2.14) of the predictions of the model ($f(\mathbf{x}_i) = \hat{y}_i$) is minimal [64, 67].

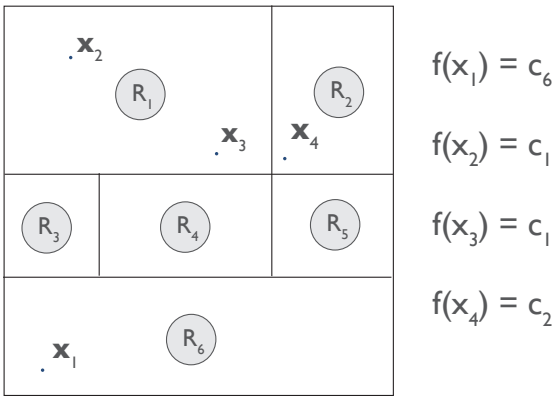

Figure 2.15 – Example of a trained DT regressor. The nomenclature is according to the main text. For visualisation purposes, this example assumes that $\mathbf{x}_n \in \mathbb{R}^2$ and that the model is previously trained such as the representation is the optimal partition schema $\mathbf{R}^*$. Note that $\mathbf{x}$ values that fall under the same region R_n have the same output c_n.

$$\mathrm{MSE}(\hat{y}_i, y_i) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{2.14}$$

If we set R_m, the optimisation of c_m follows naturally: c_m should be the average of the y values ($\langle y \rangle$) corresponding to $\mathbf{x}$ values that belong to R_m (equation 2.15).

$$\begin{aligned} c_m &= \arg\min \mathrm{MSE}(\hat{y}, y) \\ &= \arg\min \left[\sum_{(\mathbf{x},y)\in D} (f(\mathbf{x}) - y_i)^2 \right] \\ &= \arg\min \sum_{(\mathbf{x},y):\mathbf{x}\in R_m} (c_m - y_i)^2 \\ &= \langle y \rangle, \forall (\mathbf{x}, y) : \mathbf{x} \in R_m \end{aligned} \tag{2.15}$$

[14] The algorithm of the optimisation of $\mathbf{R}$ is out of the scope of this dissertation. Cf. [64–67] for more information

In spite of the ubiquitousness of this algorithm, it does not come without disadvantages. DT are highly prone to Overfitting. Overfitting is likely to occur when the tree is designed to perfectly fit all samples in the training data set. Intuitively, a perfect fitting of the data would lead to a tree with branches defined by strict rules of sparse data. In other words, a tree with nodes fitting single examples, noise and/or outliers. Overfitting affects the predictive accuracy when predicting samples that are not part of the training set [65]. Luckily, there are a set of Hyperparameters related to the structure of the tree that are used to control the extension of overfitting by limiting the complexity of the model. The discussion of the specific hyperparameters and the reasoning behind their choice will follow in the methods section of this dissertation (section 4.5.2).

2.4.2 Random Forest

As mentioned in the previous section, decision trees are prone to overfitting. Moreover, this algorithm is likely to lead to suboptimal performances because it relies on a greedy procedure to partition the hypothesis space. Bagging is a commonly used method to mitigate these issues and to balance the Bias-Variance Trade-off. [64]

Bagging (short for **b**ootstrap **agg**regat**ing**) is a statistical technique designed to improve the stability and performance of ML algorithms. It is also the base building block of the random forest algorithm. Let $\mathcal{L}^*$ be a training set of size L and $\{\mathcal{L}_k^*,\ k = 1,...,B\}$ is a set of B training sets each with the same size as $\mathcal{L}^*$. The bootstrapping resampling technique makes it possible to generate B imitations of $\mathcal{L}^*$ (with known distribution). We want each $\mathcal{L}_k^*$ to approximate $\mathcal{L}^*$. The bootstrapping resampling process consists of building B training sets each by drawing with replacement L examples from $\mathcal{L}^*$ [64]. We therefore expect that an aggregation (mean) of the outcomes of the B regression models $(f_k,\ k = 1,...,B)$, trained on learning sets bootstrapped from $\mathcal{L}^*$, is a better estimate of the performance of this algorithm on unseen data than the outcome of error analysis on a single classifier trained only on $\mathcal{L}^*$ [68][15]. The process is illustrated in figure 2.16.

The random forest regressor algorithm is essentially the application of the general bagging method to tree learners. The main difference between the general process of bagging and a random forest algorithm is that (contrary to the bagging process) the random

[15] For full demonstration of this property, cf. pp. 249-245 of [64].

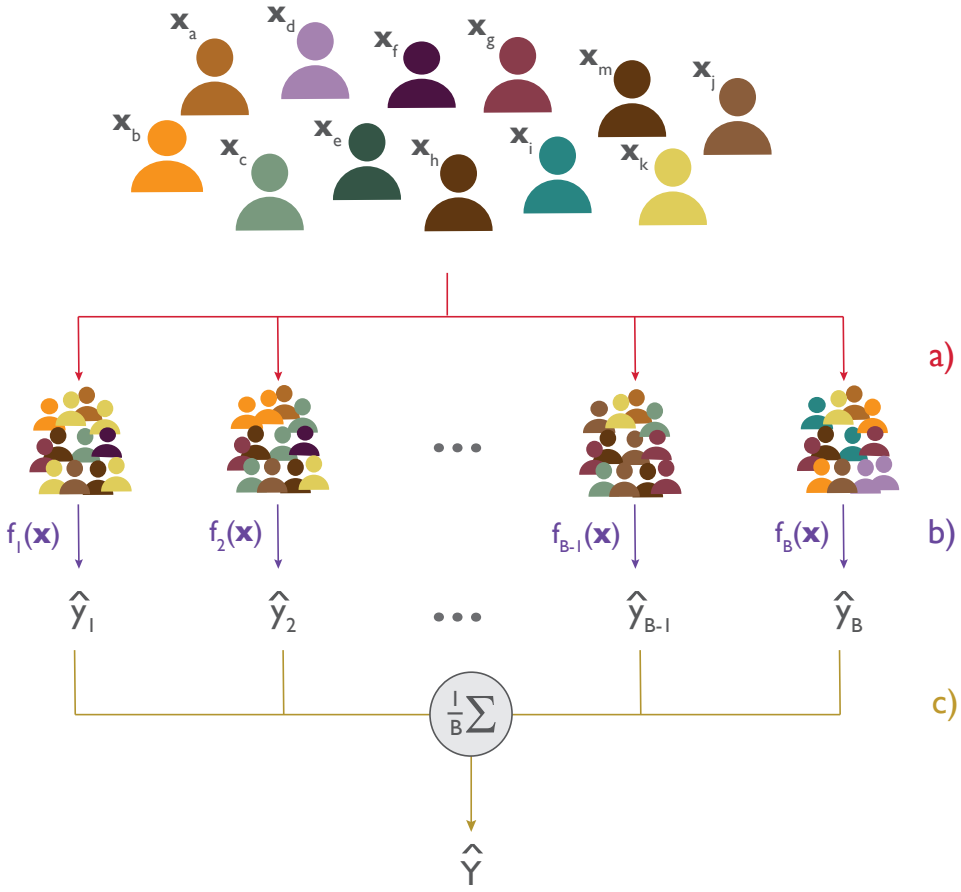

Figure 2.16 – Illustration of the bagging method. **Step a** | Draw L samples from $\mathcal{L}^*$ to form one $\mathcal{L}_k^*$. Repeat this process B times. The set of the B sets is $\{\mathcal{L}_k^*\}$ This corresponds to the bootstrapping resampling method. Note the colour repetitions: the bootstrapping method implies replacement so it is possible to draw the same sample twice. **Step b** | Next, we train a regression model $f_n(\mathbf{x})$, $n = 1, ..., B$, on each one of the bootstrapped data sets. **Step c** | Finally, the result of our ensemble method is the average of the outcomes of the B different models (Aggregation).

forest algorithm additionally only considers a random subset (without replacement) of the features per bagged dataset [64].

Chapter 3

State-of-the-Art

The motivation for the relevance of this thesis will be the focus of this chapter. It will be presented by exploring related work as follows:

1. Patients with PD (and PD-MCI) respond heterogeneously to CT (section 3.1.1).
2. There is convincing evidence that CT induces changes in connectomic measures of brain networks (section 3.1.1).
3. Connectomic measures are thought to be altered in PD (section 3.1.2).
4. ML algorithms have been successful in the prediction of treatment outcome for a multitude of psychiatric disorders using connectomic measures (section 3.2).
5. Therefore, we hypothesise that we can use ML to reliably predict the efficacy of CT based on connectomic measures of PD patients (section 3.3).

3.1 Rationale

3.1.1 Cognitive Training in Parkinson's Disease

Cognitive training is a set of activities designed to make people better at reasoning, problem solving and learning, thus targeting skills like attention, working memory or executive functioning. From a systematic point-of-view, the main goal of CT is to improve

cognition through repeated engagement of cognitive processes using one or more challenging and preferably adaptive tasks to see how the individual would adapt to untrained situations rather than if it learned the current task. [69] CT has been a promise in relieving symptoms of cognitive dysfunction in PD patients and it has been suspected that to have a neuro-protective effect [10].

The Neuroplasticity of the brain is thought to be involved in the process that mediates the relationship between CT and the improvement of cognitive functions [16]. CT is regarded as useful in preparing the brain for a better information processing [70] and computerised CT was found to be linked to improved performances on neuropsychological tests [71] and to have an impact on cerebral activation and connectivity in PD patients [37, 38]. Some studies (using MRI) have shown that CT induces brain structural and functional change in PD [10]. Others reported that PD patients who have undergone CT present significant alterations in their patterns of brain activation which seem to be related to cognitive function improvement [39]. In conclusion, as stated by a recent meta-analysis, there is convincing evidence that CT actively alters brain activation and connectivity patters and is thus thought to counteract dysfunctional activation and connectivity patterns associated with ageing and neurodegenerative diseases [10].

3.1.2 Connectomics in Parkinson's Disease

Regarding the connectomic alterations that occur due to PD pathology, *in vivo* brain CNA has shown that brain structural connectivity is compromised in disease state. PD patients have decreased whole brain global efficiency when compared to healthy elderly controls [72]. Nigro et al. [31] detected a decrease of the nodal strength, global efficiency, and global clustering coefficient in the intervention group relative to healthy controls. Moreover, locally different indices were significantly reduced: local efficiency (in the left palladium), nodal strength (in the bilateral putamen, the right amygdala and the right lingual gyrus) and clustering (in the inferior occipital gyrus) in patients. These results were replicated and extended by other authors, finding several local and global network abnormalities in PD patients' structural connectivity. [27–30].

On a sub-network level, functional connectivity has also shown to be compromised in several some sub-networks, such as the Default Mode Network, Fronto-Parietal Network, and Attention Networks, such as shown in [29]. These networks have lower scores on

the global efficiency and clustering coefficient on (medication-naive) PD patients when compared against healthy controls. The aforementioned topological differences have been interpreted as a network-level reduced capacity for integration and segregation. Moreover, these networks are thought to be deeply involved in executive functions and their disruption could pose as an hypothesis to explain the process of cognitive dysfunction [29].

Lastly, the imaging techniques discussed in sections 2.2.3 and 2.2.4 themselves also have contributed to an understanding of the mechanisms of cognitive and motor dysfunction in PD. For example, DTI has shown structural abnormalities in PD patients when compared to healthy controls, even before the patients exhibiting signs the disease. [19, 52, 73]

3.2 Related Work

There is mounting evidence suggesting that neuroimaging data and CNA can be tools for treatment outcome prediction for a multitude of psychiatric conditions.

Regarding the effectiveness of electroconvulsive therapy on relieving symptoms of schizophrenia, Li et al. [74] use a support vector machine classifier to distinguish patients from healthy controls. Using group information-guided independent-component analysis, a set of subject-specific intrinsic connectivity networks were calculated for each resting-state fMRI. These networks were then used to build support vector machine classifiers. The classification output score was found to be good predictor of electroconvulsive therapy treatment outcome.

On a study aiming to establish a relationship in treatment response to cognitive therapy and functional brain connectivity, Arnemann et al. [33] measured modularity of several brain network of a sample of 11 patients with acquired brain injuries. The sources of brain injury varied among patients, with 8 resulting from trauma and the remaining 3 resulting from stroke, tumour resection and chemotherapy. The results suggested that patients with higher baseline modularity scores were more likely to benefit from cognitive training than those with lower scores, reflecting the improvement in attention/executive functions after cognitive training.

Other two studies correlated connectomics and treatment outcome in Major Depressive Disorder (MDD). One study [34] investigates the relationship between a pharmacological treatment outcome and several graph measurements of brain networks. This study found that patients that responded to the treatment had a significantly smaller nodal degree in the right dorsolateral superior frontal gyrus than those who did not respond to treatment. On other study, Avissar et al. [35] investigated a sample of 27 MDD patients to determine whether baseline functional connectivity of distinct frontostriatal circuits predicted reliably response to a joint administration of transcranial magnetic stimulation (TMS) over the left dorsolateral prefrontal cortex (DLPFC) and medication. When using a linear model to correlate treatment outcome and functional connectivity between several areas of the brain, the results point towards higher baseline functional connectivity between the left DLPFC region and striatum translating into better treatment outcome.

Göttlich et al. [36] explored the relationship between connectivity data and clinical outcome of cognitive behavioral therapy in obsessive compulsive disorder (OCD). This investigation stems from the hypotheses that the amygdala plays an important role in the pathophysiology of OCD and that degree centrality is a predictor of OCD symptom severity. A statistically significant relationship was found between patients exhibiting a mean degree on the basolateral nuclei group of the amygdala within the range (mean ± one standard deviation) of healthy controls showed a better response to treatment than patients with a lower degree.

Schmitgen et al. [32] conducted a study to establish whether multimodal data could be used for dialectical behavioural therapy (DBT) outcomes on a population of 31 females with borderline personality disorder (BPD). Demographic and clinical, fMRI, and sMRI data were used to build Random Forrest (RFo) classifiers. This study concluded that the best prediction model could be built using fMRI and sMRI data, further proving the predictive power of neuroimgaing data. Accuracy analysis on the classifier outcomes showed that predicted responders were 3.54 times more likely respond to DBT than predicted non-responders. Conversely, predicted non-responders were 3.45 times more likely to not respond to DBT when compared to predicted responders.

3.3 Our strategy

The previous section established the value of connectomic alterations as prognostic biomarkers for the effectiveness of cognitive training on PD patients. Despite the growing body of literature, there are not to date any studies establishing a model of prediction of non-demented PD patients cognitive training response and using neuroimaging data. The COGTIPS study [26] data is a large corpus of neuroimaging data on a big sample of non-demented PD patients ($n = 140$) which will hopefully allow to extract statistically significant conclusions about the prognostic value of neuroimaging data.

As we established before, ML algorithms have been successfully used to estimate treatment outcome. Along the lines of those studies, we aim to build functional and structural connectomes using resting-state fMRI, T1-weighted MRI and DWI data. Then, extract relevant graph measures that are hypothesised to be related to cognitive dysfunction and to PD-MCI. These will be used to predict measures of the Tower of London (ToL) task (description follows in chapter 4.5.1). In this way, having graph measures from all patients, we will build a regression algorithm that reliably estimates the improvement on the mean reaction time and percentage correct trials on loads S4 and S5 of the ToL to CT based on classical regression algorithms (DT and RFo). Finally, we will proceed to an interpretation and analysis of the statistical and clinical significance of the results.

Chapter 4

Methods

4.1 Experimental design

This section is a summary of the methods paper of the COGTIPS study. For more detailed information see van Balkom et al. [26].

The COGTIPS study is a double-blinded RCT including one hundred and forty patients diagnosed with idiopathic PD that experience significant subjective cognitive complaints. The exclusion and inclusion criteria are detailed in the methods paper [26]. From the pool of included patients, 85 of them have undergone neuroimaging scans in the conditions described in section 4.2.

These patients were randomly allocated to a control or intervention group (1:1 ratio). The forty-two patients assigned to the intervention group have performed an individually online home-based intervention for eight weeks, three times a week during 45 minutes. The active control group performed a sequence of games with a total duration of 45 min that are hypothesised not to train specific cognitive functions.

The CT administered to the intervention group consists of thirteen different games that are sequentially performed. Nonetheless, the games are substantially different from the neuropsychological tasks used to access cognitive performance. The difficulty of the training is adaptive to the participants' performance. In this way, the difficulty of the

games will increase or decrease depending on individual performance thus challenging the subjects to continuously perform at their maximal ability.

The patients cognitive skills are assessed at five different time-points: pre-intervention (τ_0), post-intervention (τ_1), six-, twelve-, and twenty-four-month follow-up. The patient flow is presented on figure 4.1. For the purposes of this dissertation, τ_0 scans will be

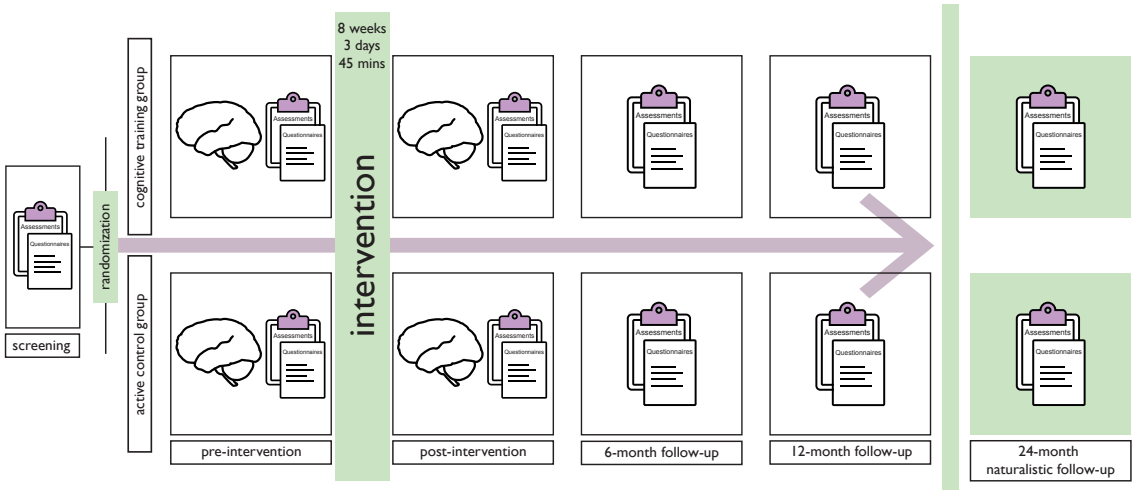

Figure 4.1 – Overview the patient flow on the COGTIPS trial. Brain scans were only acquired in the time point immediately before and after the intervention period. Cognitive performance assessment was performed on the five represented time points. Image reproduced with permission from [26].

used as well as the (change in) cognitive performance from τ_0 and τ_1. The definition of outcome measures will be detailed in section 4.5.1.

4.2 Image Acquisition

The present dissertation work did not directly comprise the MRI acquisition. However, image acquisition methods will be included in this section for completeness because MRI data is the cornerstone of this dissertation. MRI scans were acquired from 85 patients on a GE 3.0 T Discovery MR750 (General Electronics, Milwaukee, US) equipped with a 32-channel head coil at the Amsterdam UMC, location VUmc. DWI was acquired with a multi-shell single-SE echo-planar imaging sequence:

- TR = 7350 ms; TE = 81 ms;
- 2.5 mm by 2.5 mm in-plane resolution with 56 slices of 2.5 mm (no gap);

- 73 interleaved diffusion weighted directions (25 images with diffusion constant $b = 1000$ s/mm^2, 24 with $b = 2000$ s/mm^2 and 24 with $b = 3000$ s/mm^2) and 7 non-diffusion weighted volumes ($b = 0$ s/mm^2).

A 3D T_1-weighted (T_1-W) sMRI was also acquired Magnetisation-Prepared Rapid Acquisition Gradient-Echo (MPRAGE) with scan parameters according to the ADNI-3 protocol [75]:

- TR = 6.9 ms; TI = 900 ms; TE = 3.0 ms;
- Matrix size of 256 x 256;
- isotropic voxels of 1 mm^3.

Resting-state fMRI with eyes closed was acquired for 10 minutes with T_2^*-weighted echo-planar images:

- TR = 2200 ms; TE = 28 ms; Flip angle = 80°.
- 3.3 mm by 3.3 mm in plane resolution; 42 sequentially ascending slices of 3 mm with 0.3 mm gap were acquired.
- A total of 272 volumes were acquired.

For both DWI and fMRI, scans with opposite phase-encoding directions were acquired to correct for the susceptibility induced distortions. Patients followed the same protocol at both time points (τ_0 and τ_1).

4.3 Preprocessing

From an operational point-of-view, the preprocessing pipeline is the set of processes that transformed PD patients' neuroimaging data into four adjacency matrices (per patient). These matrices are four different representations of the structural and functional subject-specific connectomes. In the following sections we will go over the steps to obtain a structural connectome from DWI (section 4.3.1), a functional connectome from

resting state fMRI (section 4.3.2) and some additional preprocessing considerations on the connectomes themselves (section 4.3.3).

4.3.1 Structural Imaging Preprocessing

First, diffusion images were denoised to improve the generally low signal-to-noise ratio in DWI using the `dwidenoise` tool in `MRtrix3` [76]. Images were subsequently processed using the `EDDY` tool [77] from `FMRIB Software Library (FSL)`, version 6.0.1 [78]. We estimated the susceptibility-induced off-resonance field using `FSL topup` [79] from pairs of images with opposite phase-encoded directions and fed this field image into `FSL EDDY`. This tool corrected for susceptibility-induced distortions, eddy-current induced distortions and signal drop-out caused by movement during the diffusion encoding. Finally, within-volume and volume-to-volume motion correction was performed.

Next, Anatomically-Constrained (probabilistic) Tractography (ACT) was performed in `MRtrix3` to construct the structural connectome [76]. ACT integrates tissue information derived from a segmented structural T_1-W image to inform the propagation and termination of the streamlines during tractography (see figure 4.2 for overview of the process). The tissue response function was estimated from the preprocessed and bias-field-corrected multi-shell DWI data (`dwi2response` with `msm_5tt` algorithm). Multi-Shell Multi-Tissue Constrained spherical Deconvolution was performed to determine the fiber orientation distribution in each voxel, which was the basis for the performance of ACT by randomly seeding 100 million fibres within the white matter. Finally, the mapping of the streamlines (i.e. tractogram) was converted to a structural connectivity matrix where each edge represents the normalise count of streamlines between any two brain areas. S_{FA} was calculated by extracting the mean FA of the tracts as an alternative measure for edge weight.

We created one subject-specific parcellation of the brain for both timepoints by creating a robust template of the two T_1-W images (one at each point in time). The robust template was individually parcellated into 224 brain areas: 210 cortical and 14 subcortical areas, using `FreeSurfer` (v6.0.1) software implemented in the `fmriprep` (v1.4.0) preprocessing pipeline for fMRI [80]. The cortical brain areas were derived from registering the BNA to `FreeSurfer` space. We visually inspected the brain surfaces for any defects. The 14 subcortical areas were individually segmented by `FreeSurfer`.

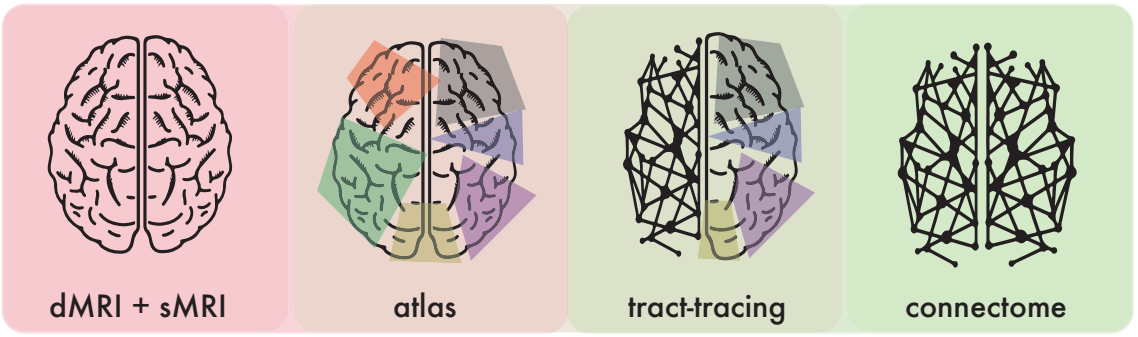

Figure 4.2 – Overview of preprocessing pipeline for extraction of a structural connectome. First, sMRI is parcellated according to the Brainnetome Atlas (BNA). Then, a tract-tracing algorithm uses the structural information to trace streamlines in between different brain areas. The number and orientation of streamlines is determined using the diffusion data. To obtain the S_S connectome, the strength of connection between two areas is a normalised count (normalised by the volume of the node) of how many streamlines connect those two regions. S_{FA} was calculated by extracting the mean FA of the tracts as an alternative measure for edge weight.

4.3.2 Functional Imaging Preprocessing

We corrected for susceptibility induced distortions in the functional image by acquiring scans with a reversed phase-encoding direction and applying `topup` [79] from `FSL` [78]. Anatomical and functional images were subsequently preprocessed using `fmriprep` [80] (v1.4.0; the full `fmriprep` boilerplate will be described in currently under review publication from our group). The definition of the brain areas was done by creating a robust template and using `FreeSurfer` (v6.0.1) as described in the previous section. fMRI images from both time points were, skull-stripped, realigned and slice-time corrected and co-registered to the robust template.

Functional time-series were extracted. Additionally, noise-regressors were extracted per subject for further denoising. Noise-regressors included global signals within the ventricles (i.e. cerebrospinal fluid signal) and white matter, frame-wise displacement and automatically identified motion-related components based on their high-frequency content and correlation with motion parameters using automatic removal of motion artefacts using independent component analysis (`ICA-AROMA`, [81]).

We removed the first three non-steady-state volumes from the fMRI in T_1-W space and spatially smoothed the images with a 6 mm full-width half-maximum isotropic gaussian kernel. Simultaneous nuisance regression and temporal filtering ([0.009;0.13] Hz)

was performed using `Denoiser`[1]. Following the benchmark test from [82], all motion-related components identified by `ICA-AROMA` and eight tissue-averaged physiological regressors were regressed out: averaged signal in the white matter and cerebrospinal fluid, along with their temporal derivatives, squares and derivatives squared. No global signal regression was applied.

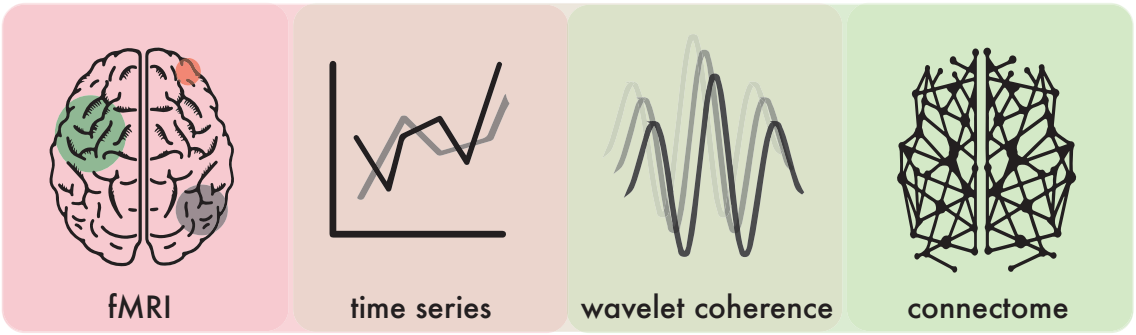

Figure 4.3 – Overview of preprocessing pipeline for extraction of function connectomes. First, resting-state fMRI are collected, parcellated according to the BNA and preprocessed. Then, time-series data is derived from the average voxel intensity of the voxels within a certain brain region. Next, wavelet coherence is calculated in between all of the time-series of different brain regions for two different bandpass frequency filters ν_1 and ν_2. Finally, the edge weight in between two nodes is defined as the wavelet coherence values for the ν_1 frequency band, originating the connectome F_1. This definition is analogous for F_2.

We define the edge strength between nodes i and j as the wavelet coherence value between time-series of the activation these nodes. Calculation of wavelet coherence was done on `MATLAB 2017b` using a modified version of the `wct` algorithm [83], version 1.0.0.0. A continuous wavelet transform using Morlet wavelets with 12 voices per octave was used. This wavelet has been widely used in the context of wavelet coherence. It's smooth and has well-developed analytical smoothing operators that effectively reduce edge effects. [84, 85].

The choice of using wavelet-based methods for definition of the connectivity between brain areas is motivated by a prospective facilitation of the examination of neurocognitive processes at different temporal scales without the edge effects in frequency space that accompany traditional bandpass filters. Wavelet-methods preserve the signal shape and are capable to denoise signal content, unlike classical correlation methods like the Pearson's coefficient [85, 86]. In turn, the calculation of the wavelet coherence between two

[1] github.com/arielletambini/denoiser

time-series is performed using Continuous Wavelet Transform (CWT) of the time series instead of (modular) discreet transforms. It is useful to think of the wavelet coherence as a localised correlation coefficient in time-frequency space. In general, CWT compares a signal with shifted and scaled ("stretched or shrunk") copies of a mother wavelet. CWT has some benefits over discreet wavelet transforms, particularly in enabling an assessment of frequency bands that are not necessarily different by powers of 2. In this way, CWT provides greater resolution while not being excessively computationally expensive. Also it has been used with success in other neuroimaging studies [85]. Discussion of analytical details of the adopted definitions regarding wavelet analysis follow in Annex II.

We derived connectomes from wavelet coherence of time-series in the following frequency bands:

- $\nu_1 = [0.06, 0.12]$ Hz
- $\nu_2 = [0.01, 0.06[$ Hz.

Let F_1 and F_2 be the connectomes derived from extraction of functional connectivity by the calculation of wavelet coherence on frequency bands ν_1 and ν_2, respectively. The ν_1 frequency band has been suggested to be a reliable and robust range that is associated with cognitive performance [86]. Achard et al. [87] demonstrated that functional networks obtained at rest displayed significant small-world structure in their data from frequency bands covering a range from 0.007 Hz to 0.45 Hz. Moreover, networks derived from higher-frequency bandpasses show greater heritability of graph statistics than those of lower-frequency bandpasses, suggesting that the latter may offer non-redundant information, which motivates our choice of ν_2 [85, 87].

4.3.3 Connectome Preprocessing

Because not every subject's DWI scan fully covered the cerebellum, the respective node was removed from all patients. Nodes 117 and 118 were excluded in all patients because there were missing streamlines in the DWI scans from these regions in over 10 subjects. Adding to those, nodes corresponding to regions with less than four voxels on the fMRI scans were excluded on the basis of being insufficient/unreliable data thus leading to the

exclusion of nodes 115 and 116 in all patients. Therefore, the CNA was performed on graphs with 220 nodes.

We constructed the two adjacency matrices from functional connectivity data, F_1 and F_2, and two adjacency matrices from structural connectivity data, S_S and S_{FA}, as outlined in sections 4.3.1 and 4.3.2, respectively.

Moreover, owing to the noisy and indirect measurement of brain structure based on water diffusion combined with the probabilistic nature of tract-tracing techniques, structural brain networks are known to contain many false-positive connections. Thresholding is a denoising technique to remove the weakest connections from a connectome as a method of mitigating spurious connections. It consists of setting all of the connections below a certain threshold to zero thus eliminating false-positive connections. [88]. There are several methods to threshold differing on the definition of the threshold value.

It is known that graph properties can rapidly change with respect to small changes in network wiring [89]. Additionally, it is not known to which extent the weaker connections are due to noise in the data, i.e. there is no ground-truth [90]. Therefore, it is not safe to say which percentage of the weaker connections could be considered false-positives. We cannot pinpoint the extent of the error that would be originated by the method of inference of the connectome [88]. So, we opted to not threshold F_1 and F_2 because removing true positives could seriously impact the calculation of the subsequent graph measures and these are our features for prediction.

Finally, to improve the accuracy of the reconstructed fibres and reduce false-positive connections, we performed Spherical-deconvolution Informed FilTering (SIFT) of tractograms (`SIFT2` in `MRtrix3`) prior to the extraction of the structural connectivity matrices S_S and S_{FA}. This thresholding method has been shown to provide a quantitative and biologically meaningful estimation of the structural connectome by applying a weight factor to individual streamlines [91] and does not warrant further thresholding of the connectivity matrix [92].

4.4 Feature Engineering

In this section we will discuss the steps necessary to extract the features to be input in the ML algorithms. The first step was the extraction of a well-defined set of graph measures from the four different modalities of connectomes (F_1, F_2, S_S and S_{FA}). Then, data was inspected for ill-defined graph measures, missing values and outliers. The measures that have an ill-defined entry for one or more patients were removed. This quality control was done by visual inspection of raincloud plots [93] for each measure. Finally, Latent Component Analysis (LCA) was applied in order to reduce the dimensionality of the data to a small number of intelligible factors.

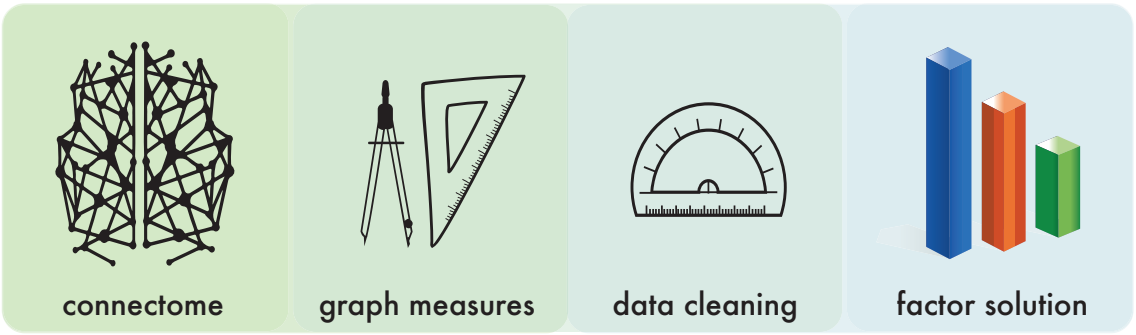

Figure 4.4 – Overview of feature engineering pipeline. First the connectomes were constructed in order to extract different graph measures. Then, the distribution of the values for the different measures and their correlations were inspected. Finally, data was processed using LCA in order to perform feature dimensionality reduction resulting in a number of factors explaining the covariance of the features.

4.4.1 Graph Measures of the Connectome

Graph measures are heavily influenced by the density of the network [89], not only in absolute value but also in their meaning/interpretation [94]. Because of our choice to use unthresholded matrices (cf. section 4.3), the functional connectivity matrices F_1 and F_2 are maximally dense; i.e. network density equals 1. Therefore, graph measures that had degenerate values (e.g. $\pm\infty$) or that had a heavy conceptual dependence (e.g. assortativity, degree centrality, eccentricity, etc.) on graph density were excluded from subsequent analyses.

The graph measures were extracted from the adjacency matrices using the `Brain`

Connectivity Toolbox[2], version 2019-03-03 [3]. Subnetworks were defined according to the 7-network parcellation of Yeo et al. [95]. The subnetworks considered were the Default Mode Network, the Fronto-Parietal Network, the SomatoMotor Network, the Dorsal Attention Network, the Ventral Attention Network, the Limbic Network, and the Visual Network.

We extracted whole-brain measures (graph measures using the entire set of 220 nodes) and subnetwork measures (graph measures on sub-graphs) on the four connectomes: S_S, S_{FA}, F_1 and F_2. The extracted whole-brain measures were: transitivity, efficiency, characteristic path length, mean global routing efficiency, modularity, small-worldness and average clustering coefficient. The subnetwork measures were: transitivity, global efficiency, characteristic path length, average clustering coefficient, average participation coefficient, average betweenness centrality, and average eigenvalue centrality. Additionally between-network connectivity was calculated for all the seven subnetworks. The definitions of the graph-derived measures follow in Annex I.

4.4.2 Latent Component Analysis

The graph features define a high-dimensional feature space where the number of features highly exceeds the number of observations. In our case, per patient we extracted: 9 whole brain measures, 9 subnetwork measures for each one of the 7 subnetworks, and $(7-1)^2$ between subnetwork measures. We performed this for 4 representations of the connectome, resulting in to 432 variables for each of the 80 patients.

Not only are Radiomics datasets often high-dimensional, they also tend to have a heavy multicollinearity burden. Many graph-derived measurements reflect similar phenomena. In our case, this redundancy is especially notorious as we extract the same measures in four connectomes from the same subject. The high-dimensionality and the linear dependence of the data result in a singular, ill-conditioned feature space leading downstream analyses to have unstable results [14]. In order to deal with this, Latent Component Analysis (LCA) was performed on the full set of graph measures. This was done using the FMradio version 1.1.1 [96][3]. FMradio was imported to Python (version

[2]Using MATLAB 2017b.
[3]We use R in version 3.6.0.

3.6.8) using the package `rpy2`[4] (v3.3.3).

`FMradio` implements a pipeline for exploratory factor analysis optimised for radiomics data. The goal is to project the radiomic feature space onto a lower-dimensional and near-orthogonal space that retains most of the information contained in the full data set and is well-behaved. This approach differs conceptually from other common methods such as Principal Component Analysis because it seeks to explain the observed **covariance** of the features through a small number of explanatory latent features instead of the observed **variance** [14].

The theoretical framework of LCA is detailed by Peeters et al. [14]. See Algorithm 1 for a pseudocode illustration of the process. The process comprises four steps:

1. Redundancy filtering for features that have Pearson's correlations above a threshold, t_c(that is a hyperparameter and will be subject to optimisation).

2. Find the optimal penalised maximum likelihood representation of the correlation matrix. In other words, we want to obtain a well-behaved (i.e. invertible and well-conditioned) regularised estimator of the filtered correlation matrix.

3. Perform factor-analytic data compression on the regular estimator. That is, the original feature-space is projected onto a (much) lower-dimensional latent feature-space (using the covariance structure model).

4. Calculate subject-specific latent factor-scores.

The first three steps fit the model, that is, given the data, find the optimal factor solutions. This construction allows us to use a previously fitted factor solution on new data using the loadings vector, Λ, and the uniqueness vector, Ξ. Keeping track of Λ and Ξ is be fundamental in order not to peak into the test set on the model validation . This process will be further explained in section 4.5.3. The last step is the data transformation step that applies the learnt factor representation to our data.

[4]Reference to rpy2 website: https://pypi.org/project/rpy2/

Algorithm 1 Latent Component Analysis

Input:
$\mathbf{X}_{\text{train}}$, a subset of the features $\mathbf{X}$.
Y_{train}, a subset of the target variable Y.
t_c, the correlation filtering threshold.
n^*, the number of factor we wish to use on our factor compression.

Output:
$\mathbf{X}_\text{T}$, the factor scores of $\mathbf{X}$
Λ, loadings vector of the factor solution.
Ξ, the uniqueness vector of the factor solution

$\text{R} \leftarrow$ `corr`$(\mathbf{X})$ ▷ correlation matrix of the features
$\text{R}_\text{T} \leftarrow$ `redundancyFilter`(R, t_c)
$\text{R}^*_\text{T} \leftarrow$ `regularCorrelation`(R_T)
$\Lambda, \Xi \leftarrow$ `factorAnalysis`(R^*_T)
$\mathbf{X}_\text{T} \leftarrow$ `factorScore`$(\mathbf{X}, \Lambda, \Xi)$
return $\mathbf{X}_\text{T}, \Lambda, \Xi$

4.5 Machine Learning Model Development

Our selected regression ML models were `DecisionTreeRegressor` and `RandomForestRegressor` from `scikit-learn` [97]. A custom estimator was designed to integrate the LCA and the ML algorithms into a single object. We will refer to the custom estimator as the `treeEstimator` if it uses DT regression, `forestEstimator` if it uses RFo regression and `treeModelEstimator` as a generic term when it is not relevant what regression type was used. This was necessary in order to not peak into the test set in the cross-validation steps. Had LCA and the algorithm been fitted separate and sequentially, the factors input to fit the ML model would have peeked into the test set. This was archived through the creation of a new class that inherits from `sklearn.base.BaseEstimator` and subsequent definition of custom fit and predict methods for our ensemble algorithms. See algorithm 2 and algorithm 3 respectively for the pseudocode of the `fit` and `predict` methods.

In this work, we evaluated the performance of four different algorithms in the prediction of four different cognitive outcomes, listed below. See figure 4.5 for an overview of the full pipeline for the methods section of this thesis. These will henceforth be referred to as prototype algorithms.

- Decision Tree Algorithm with Factors as Features (DF-F): implementation of the previously described `treeEstimator`.

- Random Forest Algorithm with Factors as Features (RF-F): implementation of the previously described `forestEstimator`.

- Decision Tree Algorithm with Graph Measures as features (DT-N): simple decision tree for regression.

- Random Forest Algorithm with Graph Measures as features (RF-N): simple random forest for regression.

Algorithm 2 Fit Method of `treeModelEstimator`

Input:
$\mathbf{X}_{\text{train}}$, a subset of the features $\mathbf{X}$.
Y_{train}, a subset of the target variable Y.
t_c, the correlation filtering threshold.
n^*, the number of factors.
θ, a configuration for the random forest algorithm.
Output:
Λ, loadings vector of the factor solution.
Ξ, the uniqueness vector of the factor solution

$\mathbf{X}_{\text{T}}, \Lambda, \Xi \leftarrow$ `latentComponentAnalysis`$(\mathbf{X}_{\text{train}}, n^*, \tau)$ ▹ see algorithm 1
treeModel $\leftarrow$ `treeModelConstructor`(θ)
treeModel`.fit`$(\mathbf{X}_{\text{T}}, \text{Y}_{\text{train}})$
return Λ, Ξ

Algorithm 3 Predict Method of `treeEstimator`

Input:
$\mathbf{x}$, an observation.
treeModel, the tree model (RFo or DT) previously fit on train data.
Λ, loadings vector of the fitted factor solution.
Ξ, the uniqueness vector of the fitted factor solution

Output:
$\hat{y}$, the model prediction.

$\mathbf{x}_{\text{T}} \leftarrow$ `factorScore`$(\mathbf{x}, \Lambda, \Xi)$
$\hat{y} \leftarrow$ treeModel`.predict`$(\mathbf{x}_{\text{T}})$
return $\hat{y}$

In the following sections, the development of the ML model will be broken down. First, the target variables will be described in section 4.5.1. This section aims to provide an overview on the methodological process to transform the (raw) clinical outcome measures into the target variables for our prediction model. Then, the pipeline for hyperparameter optimisation will be shortly discussed in section 4.5.2. Finally, in 4.5.3 the nested Cross-Validation (CV) process used to evaluate the performance of our algorithms will be detailed alongside with the statistical analyses carried out of the results.

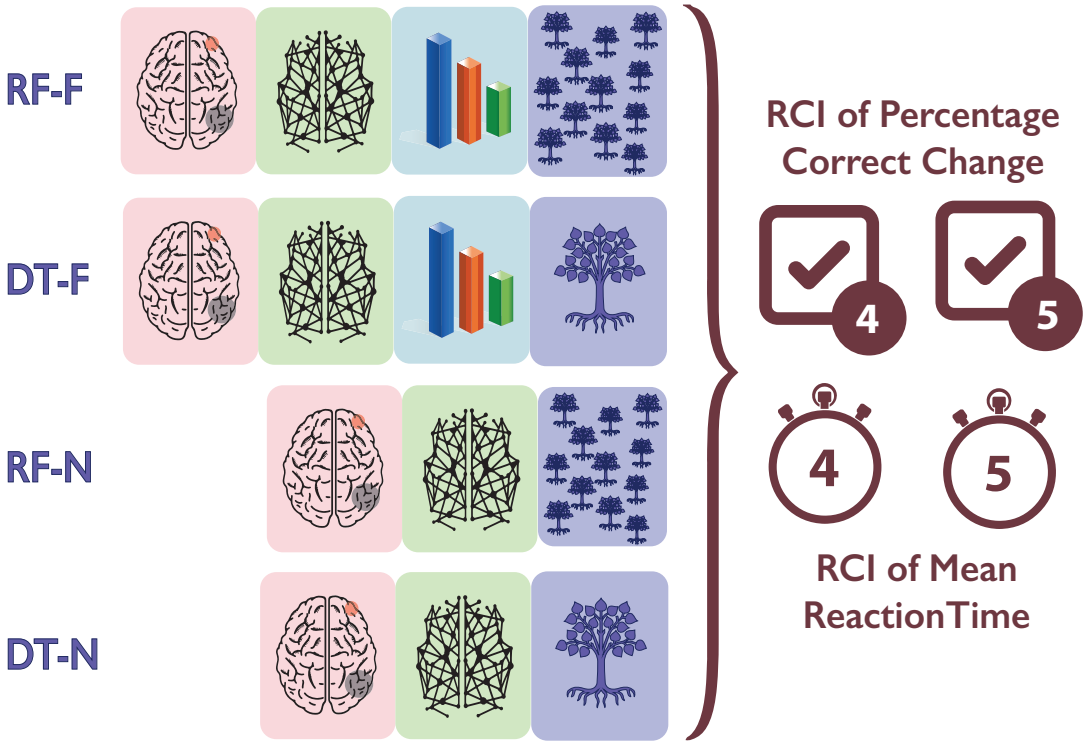

Figure 4.5 – Full Pipeline Overview for the different algorithms. Each line represents a different pipeline for a different prototype alogorithm. In pink, data acquisition process (sMRI, dMRI and fMRI). In green, connectome extraction process. In the two top lines in blue, factor compression (i.e. LCA) process. In lavender, a schematic representation of the DT (single tree) or RFo (multiple trees) regression algorithms. Finally, in brown, variables to be predicted by one of our four prototype algorithms.

4.5.1 Outcome Measures

The primary outcome of the COGTIPS study is the efficacy of CT on executive functions measured through ToL task outcome measures. The ToL paradigm is based on a modified version of the Tower of Hanoi problem. In the COGTIPS study, a computerised version of this trial is used as described by Trujillo et al. [98], on which the patients execute a total

of 90 trials per session[5]. This task has been successfully used to show that PD patients exhibit a faster rate of cognitive decline with age than matched healthy subjects [99].

The ToL task requires participants to mentally manipulate a configuration of beads stacked on pegs of varying lengths. In this paradigm patients are expect to reach an end-goal configuration while respecting several constraints [98]:

1. Only one piece may be moved at a time.
2. Each piece may be moved only from peg to peg.
3. Only a specified number of pieces may be left on each peg at every point in time.

Instructions are given to plan the whole sequence of moves that must be carried out mentally, before executing the sequence. There are five planning conditions that range in difficulty, with possible solutions ranging from one to five steps [100]. The difficulty of each planning condition is referred to as task-load – the shortest solution length is S1 and the longest S5 [26]. See figure 4.6 for a simplified visual representation of the task.

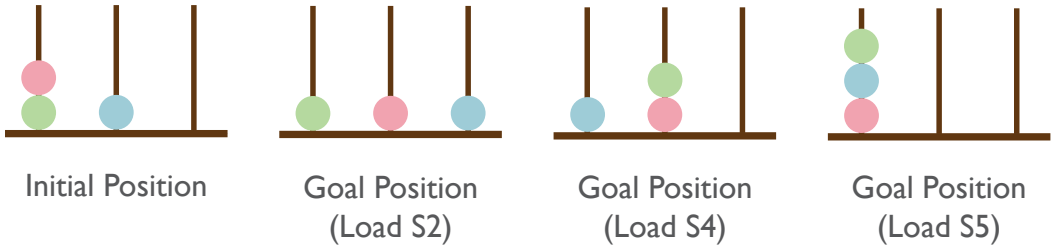

Figure 4.6 – Representation of the ToL task. The image will be described sequentially from the left. first initial position. Second: Goal position with solution minimum length of 2 moves (i.e. task-load S2) from the initial position. Third: same as second with minimum length of 4 moves from the initial position. Forth: same as previous two with solution minimum length of 5 moves.

Several cognitive perfomance measures can be extracted from the performance of this task. We will focus on the percentage of correctly answered trials and on the mean

[5]This version of the task that includes an extra level, S0, which accounts for 36 additional "counting trials" (see [98] for more information).

reaction time. The Percentage of Correctly Answered Trials (PCAT) on load Sx is defined as the proportion of correctly answered trials on load Sx; i.e. the goal-state was reached without violating any of the constraints in less than 45 s. The Mean Reaction Time (MRT) on load Sx is the mean reaction time for the correctly answered trials on load Sx. Concretely, we used the PCAT and the MRT of the ToL task-loads S4 and S5 as our main cognitive outcome measures.

For the purposes of this work, we ascribe different meanings to the terms (cognitive) outcome measures and target variables. While the former refers to the aforementioned cognitive performance outcomes of the ToL task, the latter denotes the variables to be predicted by the regression ML model. Specifically, the four different (cognitive) target variables are the RCI of our outcome measures (i.e. PCAT and MRT for task-loads S4 and S5) between time points τ_0 and τ_1. This index is said to be resilient to practice effects characteristic of the cognitive measures and accounts for the score variability both in time points τ_0 and τ_1 [101].

The RCI for each cognitive outcome for each patient is calculated using the Iverson Standard Error of the Difference $\left(\text{SED}_{\text{Iverson}}\right)$ calculation as described in [101]:

$$\text{RCI}_{\text{Iverson}} = \left(\mathcal{S}(\tau_0) - \mathcal{S}(\tau_1)\right) - \frac{\mu_1 - \mu_0}{\text{SED}_{\text{Iverson}}} \tag{4.1}$$

where $\mathcal{S}(\tau_x)$ is the score of the outcome measure on time τ_x, μ_x is the control group mean of the outcome measure on time τ_x and $\text{SED}_{\text{Iverson}}$ is defined as:

$$\text{SED}_{\text{Iverson}} = \sqrt{\left[\mathcal{S}(\tau_0) \cdot \sqrt{1 - r_{0,1}}\right]^2 + \left[\mathcal{S}(\tau_1) \cdot \sqrt{1 - r_{0,1}}\right]^2} \tag{4.2}$$

with $r_{x,y}$ being the Pearson correlation coefficient between the scores of the outcome measures on times τ_x and τ_y.

4.5.2 Hyperparameter Optimisation

The hyperparamenter optimisation process is a necessary step in order to fine-tune the performance of the regression model. ML models often have a wide range of Hyperparameters, which cannot be learned through the learning fitting process. Therefore, it is necessary to perform a search into the hypothesis space of all possible Configurations

so the model can reach peak performance on our dataset, $\mathfrak{D}$. This defines a new reinforcement learning problem: we want to minimise the error of our model by choosing an appropriate Configuration.

The hyperparameter optimisation was performed using the `Tree-structured Parzen Estimator (TPE)` approach in the `hyperopt` package [102] (version 0.2.3). `TPE` is one of the Bayesian Optimisation Methods (BOM) for hyperparameter search. This family of methods aim to build a probabilistic model of the objective function and use it to select the most promising hyperparameters to evaluate in the true objective function. In this context the objective function takes the hyperparameters as inputs and outputs the Root Mean Square Error (RMSE) of our estimator (equation 4.3).

$$\mathrm{RMSE}\left(\hat{y}_i, y_i\right) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \tag{4.3}$$

The main advantages of BOM over other common methods (such as Grid Search and Random Search) is that the former perform searches informed by the past decisions on a continuous space defined by probability density functions while the latter do a memory-less search on a discreet static space [102]. This means that not only is the efficiency of BOM superior, but also these tend to be less severely sub-optimal than blind search methods. BOM work by approximating the probability distribution function of the score (e.g. the RMSE) given the hyperparameters. In other words, BOM maximise the expected improvement from one iteration to the next as measured through the loss. Then, the set of hyperparameters predicted to have the least error (by the probabilistic model of the objective function) is input into the objective function and evaluated. Finally, the probabilistic model is updated with the results from the real function. The literature stated the lower limit for the number of iterations is 200. [102, 103]. In our research we opted to conservatively iterate over this process 500 times.

The search domain for the hyperparameter optimisation of `RF-F` is the space defined by the following probability distributions:

- `max_depth` $= \mathcal{U}\left[2, 15, 1\right]$
- `min_samples_leaf` $= \mathcal{U}\left(0.1, 0.5\right)$
- `max_samples` $= \mathcal{U}\left(0.1, 0.5\right)$

- `max_features` $= \mathcal{U}\left(0.1, 0.9\right)$
- `n_estimators` $= \mathcal{U}\left[100, 3000, 50\right]$
- `redundancy-threshold` $= \mathcal{U}\left[0.9, 0.95, 0.01\right]$
- `number_of_factors` $= \mathcal{U}\left[2, 36, 1\right]$

where $\mathcal{U}\left(a, b\right)$ is the uniform distribution of minimum value a and maximum value b and $\mathcal{U}\left[a, b, s\right]$ is the discreet uniform distribution with step s. The first 5 hyperparameters have names and definitions consistent with the original `scikit-learn` definitions [97]. The last two are hyperparameters relate to LCA (see section 4.4.2). `redundancy-threshold` refers to t_c and `number_of_factors` to n^* as described in algorithms 1 and 2. The search domain for optimisation for t_c is defined as suggested by the authors of the package [14]. The probability function of the number of factors is defined by the minimum value possible (i.e. 2 factors) until the first Guttman bound (upper bound suggested by the authors [14]). All of the remaining hyperparameters of `scikit-learn`'s `RandomForestRegressor` and `DecisionTreeRegressor` were not optimised and the default values were used. The optimisation domain for `DF-F` is the same as above with the exception of `max_samples` and `n_estimators` that are specific to the RFo algorithm and thus not included. For `RF-N` and `DT-N`, the domains are the same as for for `RF-F` and `DF-F` respectively, with exception of `redundancy-threshold` and `number_of_factors`, which are specific to the LCA.

4.5.3 Performance Evaluation

The performance of the prototype algorithms is evaluated through a k-h-nested-CV process, where $k = 10$ and $h = 5$ are the number of folds in the outer and inner loops of validation, respectively. The outer-loop refers to the optimisation of the hyperparameters (i.e. tuning). A train-test split of 80% of the data allocated to a tuning set, $\mathcal{L}$, and 20% to a test set, $\mathcal{T}$. $\mathcal{L}$ and $\mathcal{T}$ are disjoint subsets of $\mathcal{D}$. These will henceforth be referred to as surrogate sets.

$\mathcal{L}$ is further split in the inner-loop. The split is again of 80% for the learning set used to fit the candidate, $\mathcal{L}^*$, and 20% for the validation set used to the performance of the candidate, $\mathcal{T}^*$. By candidate we mean a prototype algorithm with a completely defined

configuration. However, the inner-loop performance estimation on $\mathcal{T}^*$ is inherently biased because the candidate was chosen by minimising the RMSE (as described in section 4.5.2) [4], hence the need for the outer-loop. Therefore, we estimate the generalisation error not on $\mathcal{T}^*$ but on $\mathcal{T}$, which has not been seen by the algorithm. The whole tuning process is repeated ten times with different subsets of data points allocated to $\mathcal{L}$ and $\mathcal{T}$ to quantify how reliable the error estimation is. The nested-CV process is summarised in figure 4.7.

The 10-5-nested-CV process was repeated a total of 16 times for the cognitive target variables (i.e. one time per target variable (4) per prototype algorithm (4)). All of the algorithms are tuned and tested on the same 10 surrogate $\mathcal{L}$ and $\mathcal{T}$ sets so their performances can be compared.

The process of CV implies the following assumption: the generalisation error of the final model with the optimal configuration trained on $\mathcal{D}$ tends to the average of the error of the 10-fold cross-validated models trained on their respective $\mathcal{L}$ and evaluated on an unseen hold-out set, $\mathcal{T}$ [4]. The process of choosing 10 different surrogate datasets can be seen as introducing small perturbations on $\mathcal{L}$ [4]. Taking this into account, we expect a stable fit for each one of the algorithms predicting each of the target variables. By stable fit we mean that all the 10 different iterations of the hyperparameter optimisation reached the same optimal configuration. If the fit is not stable, the model is likely highly susceptible to vibration effects: small changes in $\mathcal{L}$ lead to big changes in performance [104].

4.5.3.1 Statistical Analyses

Each one of the ten iterations of the 10-5-nested-CV results on an estimation of the RMSE on a different surrogate tuning and test sets. At the end of the process we have 10 different error estimations, which were averaged together. 95% Confidence Interval (CI) for the mean of the RMSE were calculated with one-sample t-tests. To use this parametric test we assumed normality of the distribution of the RMSE. This assumption is corroborated by the absolute values of kurtosis and skewness not being larger than 2, see Appendix A.

The algorithms must be compared to a baseline regressor to have some measure of effectiveness. From equation array 2.15, it follows that in this context the baseline regressor should be defined as a constant function. To recapitulate, the DT regression algorithm

works by iteratively attributing the average value of all examples that fall within a certain region for an optimal partition scheme. In this way, a suitable simplification of this process would be to not consider any partition scheme and just attribute the mean value of all of the previously seen examples. Therefore, we opted for a baseline regressor that always predicts new examples to be equal to the average of all of the labels (i.e. reliable change index of mean reaction time/percentage correct on loads S4 and S5) on the training set; i.e. $\mathcal{L}$.

Independent t-tests were performed on each of the 16 different runs of the 10-5-nested-CV to compare the algorithm performances with the performance of the baseline. These tests compared the 10-5-nested-CV estimations of the RMSE of the 4 different algorithms predicting the 4 different targets. As the independent t-test is fairly resilient to small variations in the group variances, the groups were all assumed to have equal variances[6] on the grounds of the ratio of the variances of the groups not being larger than 3. See Appendix A for the variance of the estimations.

4.5.3.2 *Post hoc* Analyses

We conducted *post hoc* analyses to evaluate the prototype algorithms' performances predicting age to understand the impact of the size of our sample in our main outcome prediction. Age prediction was carried out on two different sample sizes. Firstly, we run 10-5-nested-CV on the four algorithms with the intervention group data ($n = 42$, same set as for the prediction of the cognitive outcomes). Then, we repeated this process with the control and intervention group data ($n = 80$). These analyses are meant to evaluate the effect of the target variable and sample size on the performance of the four prototype algorithms. Similar statistical significance tests were conducted for algorithm performance evaluation. The uncertainty of the mean value of the RMSE is calculated as described before for the cognitive target variables. Normality and equal variances were assumed on the same grounds as before. See Appendix A for variance, skewness and kurtosis values of the error estimation for age prediction.

[6] For more information on this approximation, see [105], pp. 112, 113.

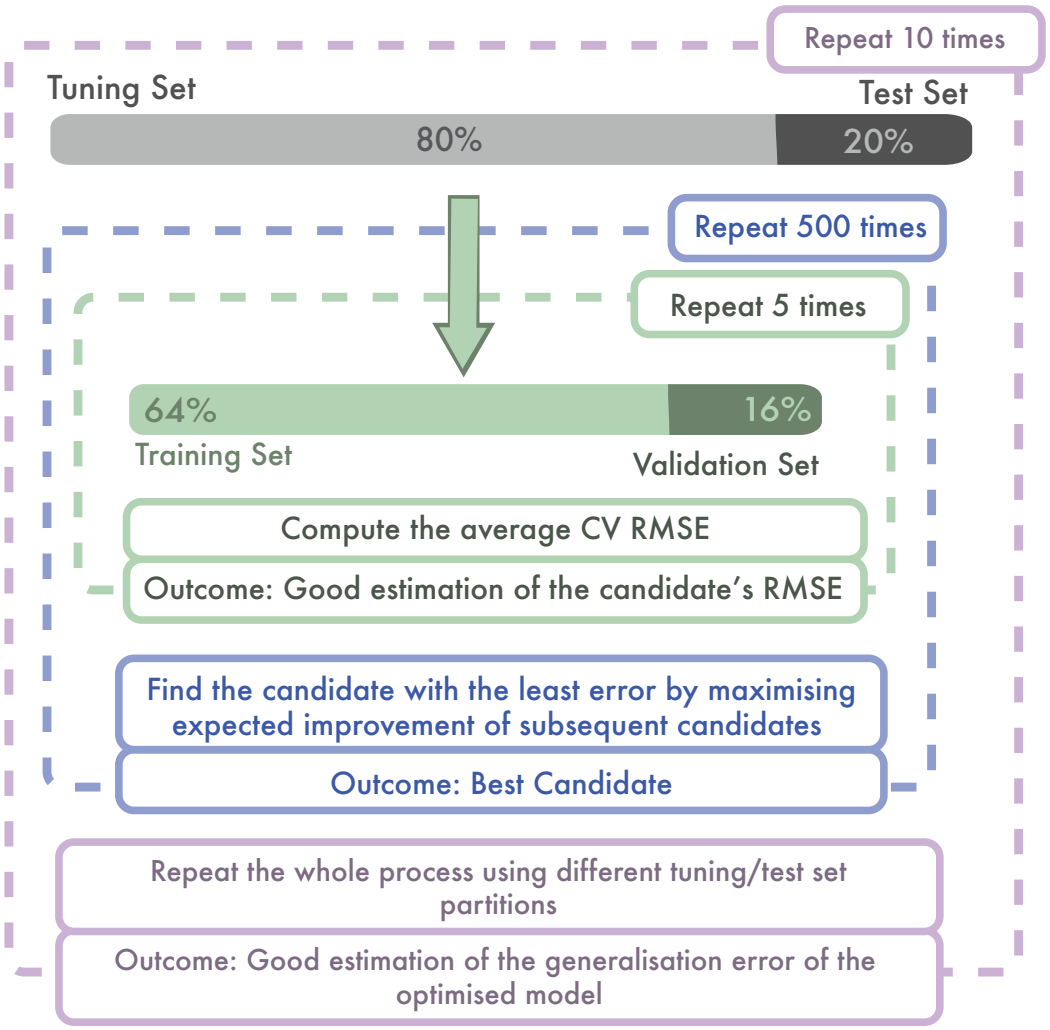

Figure 4.7 – 10-5-Nested CV. This process is repeated 500 times for all of the iterations of the TPE algorithm, represented by the blue loop. The middle (blue) loop is repeated 500 in order to find the configuration that leads to the lowest RMSE. The RMSE of any candidate is estimated using a 5-fold CV process on the tuning set, further dividing it into training set (used to fit the candidate) and validation set (used to evaluate the candidate). This process is represented by the inner (green) loop. Finally, because the choice of the best candidate is highly dependent on the initial partition of the dataset into tuning and test sets, this process is repeated 10 times. This process is represented by the outer (purple) loop. Therefore, tuning can be considered part of the learning process and the overall estimation of the generalisation error of the best candidate is a 10-fold CV calculation of the RMSE of the best candidate on a held-out (test) set.

Chapter 5

Results

5.1 Descriptive Statistics

The group characteristics are described in a paper currently under review from our research group. On figure 5.1, the distribution of the cognitive target variables is graphed. We can visually understand that all of the outcome measures have similar distributions. On figure 5.2, the distribution of the age on both the intervention ($n = 42$) and on the control+intervention group ($n = 80$) is graphed[1]. We can see both groups have very similar age distributions.

5.2 Cognitive Outcomes Prediction

We estimated the performance of four different algorithms predicting four different cognitive target variables. These target variables are the RCI of four cognitive performance outcome measures of the ToL task: MRT on ToL task-load S4, MRT on ToL task-load S5, PCAT on ToL task-load S4, and PCAT on ToL task-load S5. The performance of our prototype algorithms was assessed using the RMSE between the predictions of our algorithm and the real values. Lower values of RMSE indicate a better fit; i.e. the algorithm predicts correctly our test examples. We averaged 10 different estimations of this error on

[1] For the purposes of this section n represents the amount of elements in a given set used to carry out specific analyses.

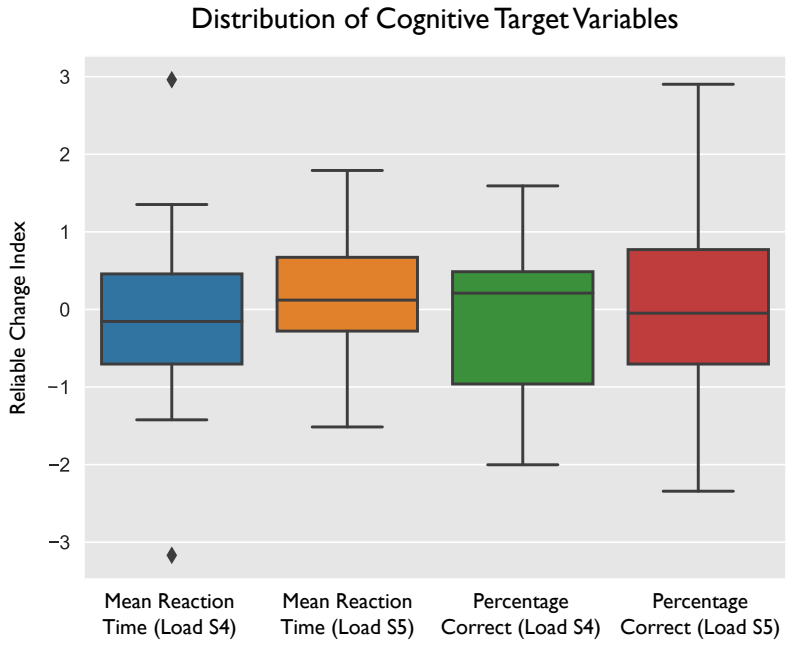

Figure 5.1 – Distribution of the Cognitive Target Variables. The diamonds represent outliers from the distribution. The target variables are represented through the Iverson reliable change index of different cognitive outcomes of the ToL task. In blue and orange, the MRT of loads S4 and S5 respectively. In green and red, the PCAT of loads S4 and S5 respectively.

different partitions of $\mathcal{D}$ in order to have a better estimation of the generalisation RMSE (i.e. the error on an unseen set when the different algorithms are trained with 100% of the available data).

Table 5.1 and figure 5.3 summarise the mean RMSE ± half of the range of the 95% CI for our 4 different prototype algorithms (i.e. `DF-F`, `DT-N`, `RF-F` and `RF-N`) predicting the four cognitive targets (i.e. MRT and PCAT on loads S4 and S5). The average error of the prototype different algorithms and the baseline is similar across the four different target variables. The prediction of the RCI MRT (Load S5) yields the least average of error for all the four algorithms. Moreover, this is the only target that accounts for an average value of the RMSE of one of the prototype algorithms (i.e. `RF-N`) lower than the baseline. However, the 95% CI of the mean values of the RMSE of `RF-N` (RMSE = 0.7 ± 0.1) and of the baseline (RMSE = 0.8 ± 0.1) overlap. There are no evident differences in the average generalisation error of any of our prototype algorithms. This seems to indicate that no algorithm performs better than the baseline in the prediction of cognitive target variables.

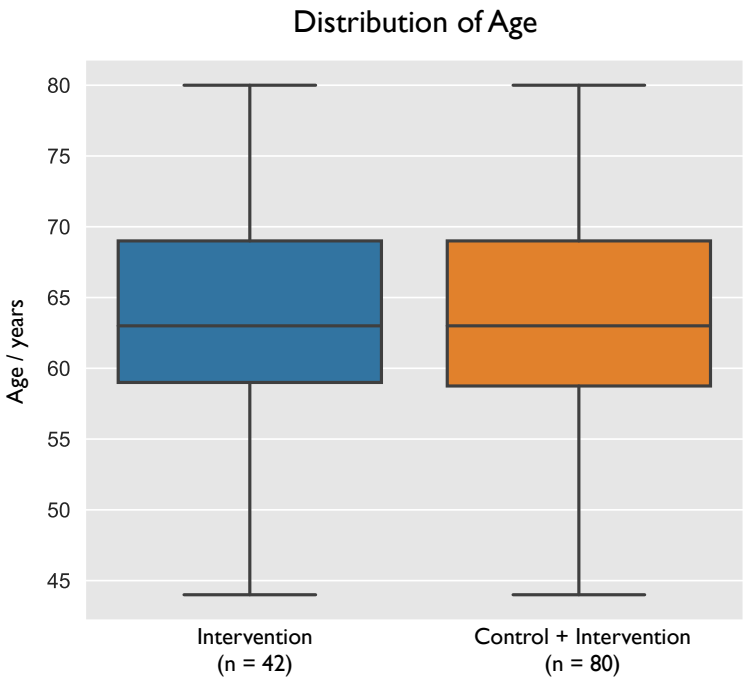

Figure 5.2 – Distribution of the Age on Control+Intervention ($n = 80$) and Intervention ($n = 42$) Groups. The Oy axis is measured in years. In blue, the distribution of ages on the intervention group. In orange, the distribution of ages on the intervention+control groups.

Two-tailed independent t-tests were conducted to ensure whether the performance of our algorithms differs significantly from the baseline algorithm. In layman's terms, the baseline performance is roughly equivalent to "the simplest regression model possible". We found no statistically significant difference between the baseline performance and the performances of any of our four algorithms. This is true for all four different algorithms predicting all four different cognitive target variables. The t-student statistic values and respective p-values (for $\alpha = 0.05$) of the performance of the four prototype algorithms are summarised in table 5.2. These results further stress that our four prototype algorithms algorithms (trained on different subsets of $\mathcal{D}$) are not better than the baseline at predicting our cognitive target variables. Therefore, we are compelled to conclude that the added complexity of our models does not imply a higher predictive accuracy of the targets with respect to the baseline.

Table 5.1 – Mean 10-5-nested-CV RMSE of the different algorithms predicting the RCI of different Cognitive Target Variables. Each column represents a different outcome. The values are presented as Mean ± half of the range of the 95% CI (for the 10 estimations of the RMSE). All values are adimensional.

	RMSE of RCI of MRT (Load S4)	RMSE of RCI of MRT (Load S5)	RMSE of RCI of PCAT (Load S4)	RMSE of RCI of PCAT (Load S5)
DF-F	1.1 ± 0.1	0.8 ± 0.1	1.0 ± 0.1	1.4 ± 0.3
DT-N	1.0 ± 0.2	0.8 ± 0.1	0.90 ± 0.08	1.3 ± 0.3
RF-F	1.0 ± 0.2	0.8 ± 0.1	0.9 ± 0.1	1.2 ± 0.3
RF-N	1.0 ± 0.2	0.7 ± 0.1	0.9 ± 0.1	1.2 ± 0.2
Baseline	1.0 ± 0.2	0.8 ± 0.1	0.9 ± 0.1	1.2 ± 0.3

Table 5.2 – Two-tailed Independent t-tests to compare the performance of the algorithms with the baseline performance to predict cognitive targets (one per column). Results in the form t(degrees of freedom) = t-statistic, p = p-value. Significance level, $\alpha = 0.05$.

	RCI of MRT (Load S4)	RCI of MRT (Load S5)	RCI of PCAT (Load S4)	RCI of PCAT (Load S5)
DF-F	t(18) = 0.869, p = .397	t(18) = 1.793, p = .092	t(18) = 1.127, p = .276	t(18) = 0.899, p = .382
DT-N	t(18) = 0.154, p = .880	t(18) = 1.127, p = .276	t(18) = -0.501, p = .623	t(18) = 0.383, p = .707
RF-F	t(18) = -0.043, p = .966	t(18) = 0.662, p = .518	t(18) = 0.027, p = .978	t(18) = 0.097, p = .924
RF-N	t(18) = 0.016, p = .988	t(18) = 0.209, p = .837	t(18) = 0.128, p = .900	t(18) = 0.118, p = .907

5.3 Age Prediction

We predicted the average 10-5-nested-CV RMSE on the intervention group ($n = 42$) and on a set comprising both the intervention ($n = 42$) and the control ($n = 38$) groups to compare the results of the prediction of this variable versus the results of the cognitive outcomes. The four algorithms and the baseline have an average RMSE of 9 ± 1 years for $n = 42$. For $n = 80$, all of the prototype algorithms have a lower average RMSE than the baseline. Concretely, RF-N has the best performance (RMSE = 6.1 ± 0.6 years) with respect to the baseline (RMSE = 7.1 ± 0.8 years). Notwithstanding the lower average RMSE of the prototype algorithms, all of their 95% CI overlap with the baseline 95% CI. The mean performances of the algorithms (± half of the range of the 95% CI) are represented on table 5.3 and figure 5.4.

Furthermore, two-tailed independent t-tests were conducted to investigate whether the performance of any of our algorithms is significantly different from the baseline. The performance of RF-N showed a significant difference from the baseline $\big(t(18) = -2.222, p < 0.05\big)$. This shows that the mean value of the 10-5-nested-CV RMSE of the RF-N differs significantly from the mean value of the 10-5-nested-CV RMSE of the baseline

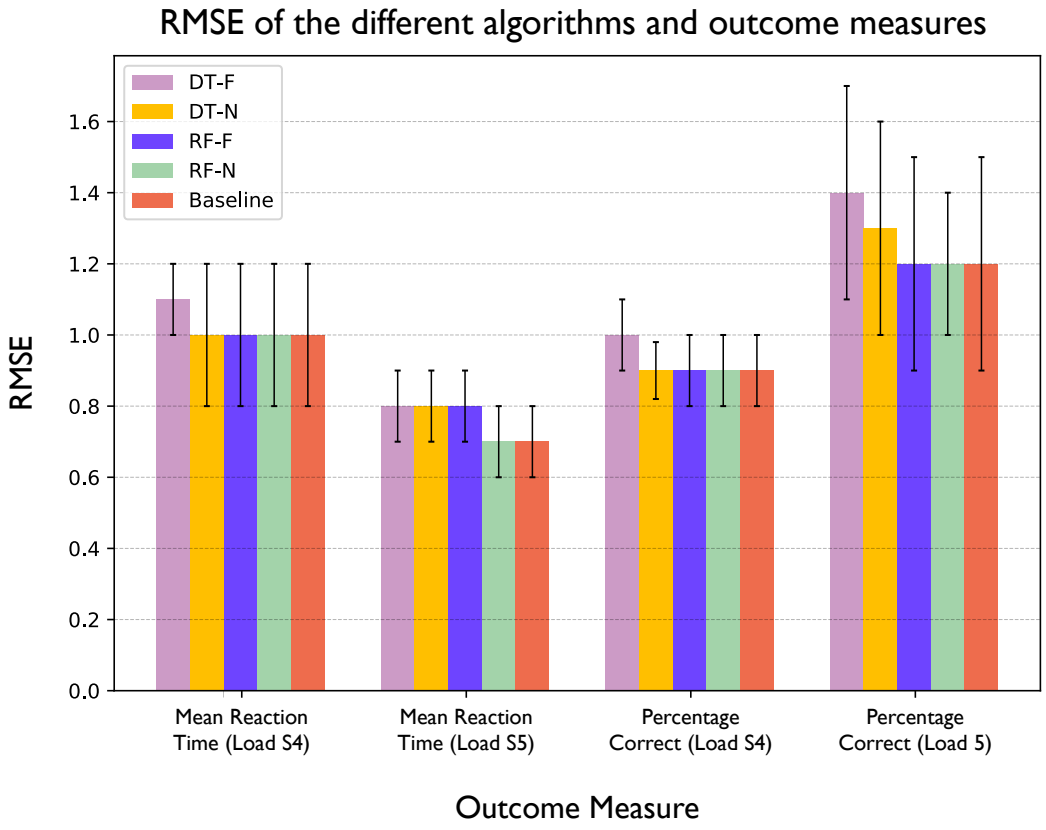

Figure 5.3 – Mean 10-5-nested-CV RMSE of the different algorithms predicting different cognitive target variables. Different set of bars represent outcomes. Each bar is the average 10-5-nested-CV performance of the algorithm on the subscribed measure. The error bars are the 95% CI of the mean value. The baseline performance is in orange.

algorithm. This result seems to indicate that RF−N on average performs better than the baseline on the task of age prediction on the dataset of 80 PD patients. We found no other significant differences between the performances of the baseline algorithm and our prototype algorithms on the 10 surrogate datasets (i.e. 10-5-nested-CV RMSE). The t-student statistic values and respective p-values (for $\alpha = 0.05$) of the independent t-tests are summarised in table 5.4.

In light of our results, we are compelled to conclude that (in this prediction task) different sample sizes yield differences in performance. The average RMSE of the prototype algorithms is lower for the eighty-samples set than for the subset with 42 samples. Moreover, the independent t-tests indicate that the performance of the RF−N is significantly better than the baseline performance for the larger sample. Therefore, the results seem to show that it is possible to predict age with RF−N for the group with eighty patients.

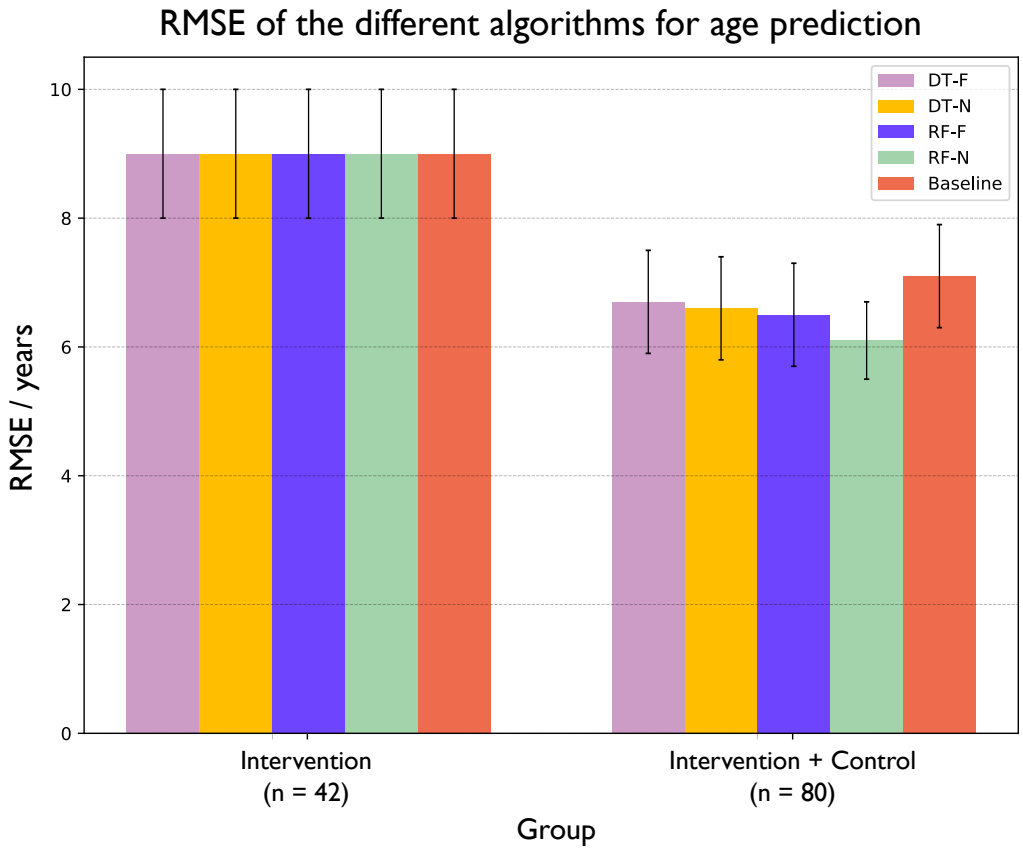

Figure 5.4 – Mean 10-5-nested-CV RMSE of different algorithms for age prediction. Different set of bars represent training sets with different sizes (i.e. $n = \{40, 80\}$). Each bar is the average 10-5-nested-CV performance of the algorithm on the subscribed measure. The error bars are the 95% CI of the mean value. The baseline performance is in orange.

Table 5.3 – Mean 10-5-nested-CV RMSE of the different algorithms predicting age. Each column represents a different outcome. The values are presented as Mean ± half of the range of the 95% CI (for the 10 estimations of the RMSE). RMSE units is years.

	RMSE of Age ($n = 42$)	RMSE of Age ($n = 80$)
DF-F	9 ± 1 years	6.7 ± 0.8 years
DT-N	9 ± 1 years	6.6 ± 0.8 years
RF-F	9 ± 1 years	6.5 ± 0.8 years
RF-N	9 ± 1 years	6.1 ± 0.6 years
Baseline	9 ± 1 years	7.1 ± 0.8 years

Table 5.4 – Two-tailed Independent t-tests to compare the performance of the algorithms with the baseline performance to predict age. Each column represents a different sample size. Results in the form t(degrees of freedom) = t-statistic, p = p-value. Significance level, $\alpha = 0.05$.

	Age ($n = 42$)	Age ($n = 80$)
DF–F	t(18) = -0.869, $p = .398$	t(18) = -0.759, $p = .457$
DT–N	t(18) = -0.868, $p = .398$	t(18) = -1.058, $p = .303$
RF–F	t(18) = -0.626, $p = .548$	t(18) = -1.268, $p = .218$
RF–N	t(18) = -0.937, $p = .362$	**t(18) = -2.222, $p = .037$**

Chapter 6

Discussion

In this dissertation work, we investigated whether it is possible to predict the outcome of CT in PD patients using solely connectomic measures derived from structural and functional MRI as predictive features. Four different regression models were used. We compared the performance of simple random forest and decision tree regression algorithms with (RF−F/DF−F) and without (RF−N/DT−N) prior latent component analysis. Four hundred and thirty-two graph measures derived from functional and structural connectomes were used as features and cognitive performance indices were used as target variables for prediction. The target variables are four different Iverson Reliable Change Indices of ToL outcomes: MRT and PCAT on ToL task-loads S4 and S5. None of the performances of our four prototype algorithms significantly differed from the baseline performance in the prediction of any of the four main outcomes.

We conducted additional *post hoc* analyses to further investigate the effect of the sample size on predictive accuracy. For this reason, age prediction was carried out in two groups with different sizes: 42 and 80 samples. The results on the group with 80 patients showed a lower average RMSE. In fact, the difference in the mean performances of RF−N and of the baseline reached statistical significance on the age prediction task on the data set with the biggest cardinality. Therefore, we hypothesise the lack of statistical significance in our main results as a product of three distinct study design features synergistically interacting - a **small sample size** and a **small effect size** and a **conservative**

error estimation method.

The sample/dataset size is of paramount importance to the performance of ML algorithms. Probabilistically, a small sample is unlikely to represent accurately the entire spectrum of characteristics of a population thus not meeting the fundamental Independent and Identically Distributed (i.i.d.) assumption.[1] This may lead to a potential limitation of the generalisability of the predicted results to other independent sample sets (e.g. the test set). In a 2018 study Cui and Gong [106] have shown the dramatic effect of sample size on individualised behavioural prediction in studies that used functional-connectivity-derived features. They found that the most common ML regression algorithms reach acceptable performances on the aforementioned task given a sufficiently large[2] data set. This need is in part motivated by the increased likelihood of violation of the i.i.d. assumption for smaller sample sizes. This work dealt with a total of 42 PD patients to carry out the main analyses, which is far removed from the benchmarks set by Cui and Gong [106]. Notwithstanding such remarks, our reduced sample size is not uncommon as most neuroimaging studies have samples of typically less of 100 subjects and are strife with confounding effects[3] [107].

Furthermore, the hyperparameter optimisation process was unstable. See Appendix B for the several optimal configurations with different partitions of the training and test sets for the RF−F algorithm. This instability seems to be an *ad hoc* manifestation of the susceptibility of our pipelines to **vibration effects**. Vibration effects describe the drastic impact some methodological choices (e.g. the hyperparameters) have a on the error of some unstable pipelines [104]. In this way, we interpret the instability of the hyperparameter optimisation process in two different but compatible ways:

1. The surrogate sets are not representative of the whole dataset; i.e. they are not i.i.d. subsets of $\mathcal{D}$. This can either be due to a small sample size, the intrinsically large

[1] In ML theory, examples are often considered to be i.i.d.. This implies that all samples stem from the same generative process and that the generative process is assumed to have no memory of past generated samples. Furthermore, i.i.d. datasets allows us to make the stationary assumption: there is a probability distribution over examples that remains stationary over time. These two assumptions together connect the past (learnt examples) to the future (new unseen examples) thus allowing predictions to be possible [63].

[2] Even though it is not directly stated in the paper, it is visually identifiable that by large sample sizes it is meant a sample larger than 200 different subjects. See figure 2 of [106] for more detailed information.

[3] For instance, in our case, we can point out that the temporal structure of the trial (i.e. a longitudinal study) diminishes power by any possible carry-over effects; i.e. the cognitive outcome measures in τ_0 and τ_1 are not strictly independent.

sampling noise in neuroimaging studies (for instance, due to highly heterogeneous groups), or because the features are not meaningfully related to the outcomes.

2. The prediction model does not adequately fit the data. This can be due to an inappropriate model choice (cf. Bias-Variance Trade-off) or for the same reasons as the previous remark.

Adding to the above, the results of the *post hoc* age prediction analyses further corroborate the hypothesis of lack of data. Functional connectivity (graph) measures are thought to be significantly related to age [108]. In particular, Khosla et al. [109] have archived a minimal RMSE of 2.1 years predicting age with state-of-the-art connectome-based ensemble learning. The prediction of age has a well-established proof-of-concept and should therefore be possible. In our work, not only has the RF−N algorithm reached significance levels with the bigger dataset ($n = 80$) but also the average RMSE of all algorithms is lower when compared to the smaller dataset ($n = 42$).[4] A direct comparison to our results (average RMSE of RF−N of 6.1± 0.6 years) is not possible taking into account that Khosla et al. had a wildly bigger sample (polling data collected from over 1000 individuals) and used state-of-the-art dedicated pipelines for functional data collection (ensemble of tailored convolutional neural networks). In this way, the significance of results of RF−N age prediction on PD patients leads us to infer that the methods presented on this paper hold promise, even if these result in errors that are far removed from the peak-performance found in literature.

Sample size has also a subtle effect on predictive modelling approaches when compared to standard statistical methods. Varoquaux [107] analyses the influence of some common methodological choices on the generalisation error estimation of neuroimaging studies in a 2018 paper. One important finding of this study reports that the effect of having a small sample size is especially dismal in the specific case of testing predictive models. Therefore, studies based on predictive modelling require larger sample sizes

[4] From a theoretical standpoint, the *No Free Lunch Theorem* should not be forgotten in these considerations. This theorem postulates that "if an algorithm performs well on a certain class of problems then it necessarily pays for that with degraded performance on the set of all remaining problems" [110]. We tried to mitigate this effect by not considering a single estimation RMSE but rather averaging the RMSE over a group of 10 different but closely related training sets. However, this assumption is still constrained directly by our sample and indirectly by the universe defined by the generalisation of our sample. This epistemological uncertainty works both ways. On the one hand, it supports our motivation for a negative result not implying an incorrect methodology. On the other, it introduces some epistemic ambiguity on our significant outcome.

than standard statistical approaches. From the seven treatment outcome prediction studies in section 3.2, four [33–36] of them use standard statistical approaches (e.g. Pierson correlations, among others) which can partially motivate their success with relatively smaller sample sizes.

As made clear in section 3.2, there is mounting evidence suggesting that neuroimaging data and CNA can be used to predict treatment outcome in a multitude of psychiatric conditions. Nonetheless, regression ML models using graph measures as features are not (yet) a common approach to the problem of prediction of treatment outcome in neuropsychiatry. We described several related works which were exclusively classification problems thus hindering any direct comparison with the findings of this dissertation. A 2016 systematic review on the use of ML on Human Connectome Data from MRI [111] analysed seventy-seven different studies on this topic. From those, only four [112–115] included regression analysis, none of which was conducted using graph-derived features. To the best of our knowledge, the lack of publications on regression-based predictive models in neuropsychiatry is still egregious to the present day. Moreover, these studies that report positive results have bigger sample sizes than the work carried out on this dissertation, ranging from 60 [112] to 115 subjects [113, 114].

Despite the sparsity of publications concerning regression-based predictive models, a recently published study by Sun et al. [116] has shown that it is possible discern between remitters and non-remitters of electroconvulsive therapy for major depressive disorder using a regression model based on whole-brain functional connectivity. However, not only did they have a bigger sample size (122 patients), there is also compelling evidence that electroconvulsive therapy has a larger effect size than CT. A 2019 meta-analysis [117] has described the effect size of CT on several cognitive in terms of the Hedges' g. These were described as small to moderate positive treatment effects (compared with control interventions) in four domains (global cognitive function, memory, working memory and executive function). Conversely, the effects of electroconvulsive therapy for Major Depressive Disorder are reported as to have large and very large effect sizes in the mental health component score (also characterised through the Hedges' g) [118]. The disparity of effect sizes seems to indicate that CT may have a smaller effect size than electroconvulsive therapy in the domain of their respective relevant outcome measures. Attention must be drawn to the fact that small sample sizes and small effect sizes combined have

an especially destructive effect on statistical power in neuroscience. Not only does the small sample size reduce the chance of detecting a true positive effect, it also reduces the likelihood that a statistically significant result reflects a true positive effect [119]. Therefore, we may not have enough power to make predictions as sample and effect sizes are two key elements for statistical power.

Additionally, none of the previously mentioned ML-regression-based use k-h-nested-CV (or equivalent techniques such as unbiased bootstrapping based methods [4]) to estimate the generalisation error of their respective models. We believe this has a non-trivial effect on the statistical significance of our results because it leads to a conservative estimation of the average error and the respective uncertainty of our prototype algorithms [107]. Conservative error estimations reduce the power of our study thus further stressing the need for bigger sample sizes.

Finally, the choice of the amount of structural and functional measures we used as features may also be a factor leading to negative results. This data-driven approach created a high-dimension problem, where the number of features (432) is vastly larger than the number of observations (42). Overfitting is the hallmark of high-dimension problems. In order to bypass it, highly regularised approaches often become the methods of choice [64]. In our case, we used Latent Component Analysis as described in section 4.4.2 and 4.5. The LCA did not bring any significant improvement to the versions of the prototype algorithms that did not use it. We hypothesise that the lack of effectiveness of LCA is due to the regularisation strategies inherent to the RFo algorithm (i.e. bagging of features and of samples). To sum up, the fact that the RFo uses different subsets of features and samples iteratively in the training process might have undermined the effect of LCA, which is used for roughly the same end.

6.1 Future Work and Recommendations

The previous discussion points provide grounds for recommendations with respect to future work on regression analysis using ML in neuropsychiatry. What constitutes a fair interpretation of new discoveries is unavoidably subjective. However, critical discussion of limitations, caveats, and a reserved stance against one's findings is fundamental for scientific progress.

Firstly, future work is needed with respect to the understanding of LCA of multimodal graph measures. A data-driven approach was adopted for the purposes of this work. In other words, all the four hundred and thirty-two features were input into an unsupervised dimensionality reduction method and then into a ML algorithm, without any further manual feature selection. However, as mentioned before, radiomics data are often high-dimensional and strife with a significant multicolinarity burden. To sum up, comprehensive analysis of the features that are eliminated in redundancy filtering may shed some light onto the nature of connectomic measures. Moreover, a more informed choice of features might ultimately improve the performance of the prediction algorithms by not burdening the process with a lot of meaningless variables for the specific problem.

Regarding the performance estimation of ML models, a general recommendation should be issued for the mainstream use of more unbiased validation processes such as *h*-*k*-nested-CV (or equivalent). There is a growing need for accurate error analysis in neuropsychiatry as the corpus of studies using ML in this field grows. Tsamardinos et al. [4] described several common processes in to estimate the generalisation error of ML algorithms. It was described that unbiased techniques for error estimation – such as nested-CV or bootstrap bias-corrected CV – often lead to less outstanding and potentially non-significant results. Thereby, an accurate and unbiased estimation of the generalisation error is of the utmost importance as more and more ML-based softwares are deployed in clinical practice. If deployment of such algorithms is the end goal, more work is also needed in making systems resilient to dataset shifts and adversarial attacks, as described by Kelly et al. [120]. Nonetheless, the publication of studies showing non-significant results is fundamental for the construction of scientific knowledge. Underpowered studies are a staple of the scientific method and even these may lead to significant results down the line in aggregative studies such as prospective meta-analyses [104].

Lastly, the potential impact of the reduced size of our sample was extensively discussed in this chapter. Therefore, suffice to say that the sample size should exceed the hundred, should it be possible.

6.2 Conclusion

In this work we aimed to predict several cognitive outcome measures using data derived from functional and structural MRI. We used custom models based on the Random Forest and Decision Trees regression algorithms to reach our goal. The performance of our algorithms did not have a significant level of improvement relatively to the baseline. Notwithstanding the failure to predict our main outcomes, we hypothesise that the main setback of this work was the small size sample of our sample.

References

[1] William R. Hendee, E. Russell Ritenour, and Kenneth R. Hoffmann. *Medical Imaging Physics*. 4th. 2003. ISBN: 0471382264. DOI: 10.1118/1.1563664.

[2] Marc Claesen and Bart De Moor. «Hyperparameter Search in Machine Learning.» In: (2015), pp. 10–14. arXiv: 1502.02127. URL: http://arxiv.org/abs/1502.02127.

[3] Mikail Rubinov and Olaf Sporns. «Complex network measures of brain connectivity: Uses and interpretations.» In: *NeuroImage* 52.3 (2010), pp. 1059–1069. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2009.10.003.

[4] Ioannis Tsamardinos, Elissavet Greasidou, and Giorgos Borboudakis. «Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation.» In: *Machine Learning* 107.12 (2018), pp. 1895–1922. ISSN: 15730565. DOI: 10.1007/s10994-018-5714-4. arXiv: 1708.07180. URL: https://doi.org/10.1007/s10994-018-5714-4.

[5] Alex Fornito, Andrew Zalesky, and Edward T Bullmore. *Fundamentals of Brain Network Analysis*. 1st ed. Amsterdam, Boston: Elsevier/Academic Press, 2016. ISBN: 9780124079083.

[6] Lauren Walker, Leonidas Stefanis, and Johannes Attems. «Dementia and Cognitive Impairment: Epidemiology, Diagnosis, and Treatment.» In: *Journal of Neurochemistry* 150.5 (2019), pp. 467–474. ISSN: 14714159. DOI: 10.1111/jnc.14698.

[7] B D Cullity and C D Graham. «Diamagnetism and Paramagnetism.» In: *Introduction to Magnetic Materials*. IEEE, 2009, pp. 87–114. ISBN: 9780470386316. DOI: 10.1002/9780470386323.ch3. URL: http://ieeexplore.ieee.org/document/5488888.

[8] W. R. G. Gibb and A. J. Lees. «The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson ' s disease.» In: *Journal of Neurology, Neurosurgery, and Psychiatry* 51.6 (1988), pp. 745–752.

[9] Henry W. Mahncke, Amy Bronstone, and Michael M. Merzenich. «Brain plasticity and functional losses in the aged: scientific bases for a novel intervention.» In: *Progress in Brain Research* 157.06 (2006), pp. 81–109. ISSN: 00796123. DOI: 10.1016/S0079-6123(06)57006-2.

[10] Tim D. van Balkom et al. «The Effects of Cognitive Training on Brain Network Activity and Connectivity in Aging and Neurodegenerative Diseases: a Systematic Review.» In: *Neuropsychology Review* 30.2 (2020), pp. 267–286. ISSN: 15736660. DOI: 10.1007/s11065-020-09440-w.

[11] Geoffrey I. Webb. «Overfitting.» In: *Encyclopedia of Machine Learning*. Ed. by Claude Sammut and Geoffrey I Webb. Boston, MA: Springer US, 2010, p. 744. ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_623. URL: https://doi.org/10.1007/978-0-387-30164-8%7B%5C_%7D623.

[12] Dennis W. Dickson. «Neuropathology of Parkinson Disease Dennis.» In: *Parkinsonism Related Disorders* 46.1 (2018), S30–S33. DOI: 10.1016/j.physbeh.2017.03.040.

[13] Donald W. McRobbie et al. *MRI from picture to proton*. 2nd ed. Cambridge: Cambridge University Press, 2006, pp. 325–345. ISBN: 9781107706958. DOI: 10.1017/9781107706958.

[14] Carel F. W. Peeters et al. «Stable prediction with radiomics data.» In: (2019). arXiv: 1903.11696.

[15] J. Jankovic. «Parkinson's disease: Clinical features and diagnosis.» In: *Journal of Neurology, Neurosurgery and Psychiatry* 79.4 (2008), pp. 368–376. ISSN: 1468330X. DOI: 10.1136/jnnp.2007.131045.

[16] Sjors C.F. van de Weijer et al. «The Parkin'Play study: Protocol of a phase II randomized controlled trial to assess the effects of a health game on cognition in Parkinson's disease.» In: *BMC Neurology* 16.1 (2016), p. 209. ISSN: 14712377. DOI: 10.1186/s12883-016-0731-z.

[17] Gebhard Sammer et al. «Training of executive functions in Parkinson's disease.» In: *Journal of the Neurological Sciences* 248.1-2 (2006), pp. 115–119. ISSN: 0022510X. DOI: 10.1016/j.jns.2006.05.028.

[18] Irene Litvan et al. «Diagnostic Criteria for Mild Cognitive Impairment in Parkinson's Disease: Movement Disorder Society Task Force Guidelines.» In: *Movement Disorders* 27.3 (2012), pp. 349–356. DOI: 10.1002/mds.24893.Diagnostic.

[19] Rimona S. Weil, Alyssa A. Costantini, and Anette E. Schrag. «Mild Cognitive Impairment in Parkinson's Disease—What Is It?» In: *Current Neurology and Neuroscience Reports* 18.4 (2018). ISSN: 15346293. DOI: 10.1007/s11910-018-0823-9.

[20] Jennifer G. Goldman and Irene Litvan. «Mild Cognitive Impairment in Parkinson's Disease.» In: *Minerva medica* 102.6 (2011), pp. 441–459. ISSN: 15378276. DOI: 10.1038/jid.2014.371. arXiv: NIHMS150003. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf.

[21] D. Aarsland et al. «Mild cognitive impairment in Parkinson disease: A multicenter pooled analysis.» In: *Neurology* 75.12 (2010), pp. 1062–1069. ISSN: 1526632X. DOI: 10.1212/WNL.0b013e3181f39d0e.

[22] D. Aarsland et al. «Risk of dementia in Parkinson's disease: A community-based, prospective study.» In: *Neurology* 56.6 (2001), pp. 730–736. ISSN: 0028-3878.

[23] Courtney C. Walton et al. «Cognitive Training in Parkinson's Disease: A Theoretical Perspective.» In: *Neurorehabilitation and Neural Repair* 31.3 (2017), pp. 207–216. ISSN: 15526844. DOI: 10.1177/1545968316680489.

[24] D. I. Sitzer, E. W. Twamley, and D. V. Jeste. «Cognitive training in Alzheimer's disease: A meta-analysis of the literature.» In: *Acta Psychiatrica Scandinavica* 114.2 (2006), pp. 75–90. ISSN: 0001690X. DOI: 10.1111/j.1600-0447.2006.00789.x.

[25] Isabella H.K. Leung et al. «Cognitive training in Parkinson disease: A systematic review and meta-analysis.» In: *Neurology* 85.21 (2015), pp. 1843–1851. ISSN: 1526632X. DOI: 10.1212/WNL.0000000000002145.

[26] Tim D. van Balkom et al. «COGTIPS: a double-blind randomized active controlled trial protocol to study the effect of home-based, online cognitive training on cognition and brain networks in Parkinson's disease.» In: *BMC Neurology* 19.1 (2019), p. 179. DOI: 10.1186/s12883-019-1403-6.

[27] Koji Kamagata et al. «Connectome analysis with diffusion MRI in idiopathic Parkinson's disease: Evaluation using multi-shell, multi-tissue, constrained spherical deconvolution.» In: *NeuroImage: Clinical* 17.June 2017 (2018), pp. 518–529. ISSN: 22131582. DOI: 10.1016/j.nicl.2017.11.007.

[28] Changhong Li et al. «Impaired topological architecture of brain structural networks in idiopathic Parkinson's disease: a DTI study.» In: *Brain Imaging and Behavior* 11.1 (2017), pp. 113–128. ISSN: 19317565. DOI: 10.1007/s11682-015-9501-6.

[29] Chris Vriend et al. «Global and Subnetwork Changes of the Structural Connectome in de novo Parkinson's Disease.» In: *Neuroscience* 386 (2018), pp. 295–308. ISSN: 18737544. DOI: 10.1016/j.neuroscience.2018.06.050.

[30] Ming Ching Wen et al. «Structural connectome alterations in prodromal and de novo Parkinson's disease patients.» In: *Parkinsonism and Related Disorders* 45 (2017), pp. 21–27. ISSN: 18735126. DOI: 10.1016/j.parkreldis.2017.09.019.

[31] S. Nigro et al. «Characterizing structural neural networks in de novo Parkinson disease patients using diffusion tensor imaging.» In: *Human Brain Mapping* 37.12 (2016), pp. 4500–4510. ISSN: 10970193. DOI: 10.1002/hbm.23324.

[32] Mike M. Schmitgen et al. «Individualized treatment response prediction of dialectical behavior therapy for borderline personality disorder using multimodal magnetic resonance imaging.» In: *Brain and Behavior* 9.9 (2019). ISSN: 2162-3279. DOI: 10.1002/brb3.1384.

[33] Katelyn L. Arnemann et al. «Functional brain network modularity predicts response to cognitive training after brain injury.» In: *Neurology* 84.15 (2015), pp. 1568–1574. ISSN: 1526632X. DOI: 10.1212/WNL.0000000000001476.

[34] Zhenghua Hou et al. «Divergent topological architecture of the default mode network as a pretreatment predictor of early antidepressant response in major depressive disorder.» In: *Scientific Reports* 6.December (2016), pp. 1–9. ISSN: 20452322. DOI: 10.1038/srep39243.

[35] Michael Avissar et al. «Functional connectivity of the left DLPFC to striatum predicts treatment response of depression to TMS.» In: *Brain Stimulation* 10.5

(2017), pp. 919–925. ISSN: 18764754. DOI: 10.1016/j.brs.2017.07.002. URL: http://dx.doi.org/10.1016/j.brs.2017.07.002.

[36] Krista M. Lisdahl et al. «Neural mechanisms of brain plasticity with complex cognitive training in healthy seniors.» In: *Cerebral Cortex* 25.1 (2017). Ed. by Wim Van Hecke, Louise Emsell, and Stefan Sunaert, pp. 115–119. ISSN: 14602199. DOI: 10.1093/cercor/bht234. arXiv: 9809069v1 [arXiv:gr-qc]. URL: http://dx.doi.org/10.1186/s13063-016-1253-0%20http://dx.doi.org/10.1007/s11065-016-9330-4%20https://doi.org/10.1186/s13063-018-3143-0%20http://dx.doi.org/10.1016/j.neubiorev.2015.09.007%20http://dx.doi.org/10.1016/S0140-6736(09)61257-5%20http://dx.doi.org/10.1038/n.

[37] María Díez-Cirarda et al. «Increased brain connectivity and activation after cognitive rehabilitation in Parkinson's disease: a randomized controlled trial.» In: *Brain Imaging and Behavior* 11.6 (2017), pp. 1640–1651. ISSN: 19317565. DOI: 10.1007/s11682-016-9639-x.

[38] Antonio Cerasa et al. «Neurofunctional correlates of attention rehabilitation in Parkinson's disease: An explorative study.» In: *Neurological Sciences* 35.8 (2014), pp. 1173–1180. ISSN: 15903478. DOI: 10.1007/s10072-014-1666-z.

[39] Cristina Nombela et al. «Cognitive rehabilitation in Parkinson's disease: Evidence from neuroimaging.» In: *Frontiers in Neurology* 2 (2011), p. 82. ISSN: 1664-2295. DOI: 10.3389/fneur.2011.00082.

[40] Michael J. Aminoff, David A. Greenberg, and Roger P. Simon. *Clinical neurology*. 9th. Vol. 113, 317-3. McGraw-Hill Education, 2015. ISBN: 9780071841436. DOI: 10.1201/b17272.

[41] Wenya Yang et al. «Current and projected future economic burden of Parkinson's disease in the U.S.» In: *npj Parkinson's Disease* 6.1 (2020), pp. 1–9. ISSN: 23738057. DOI: 10.1038/s41531-020-0117-1. URL: http://dx.doi.org/10.1038/s41531-020-0117-1.

[42] Werner Poewe et al. «Parkinson disease.» In: *Nature Reviews Disease Primers* 3 (2017), pp. 1–21. ISSN: 2056676X. DOI: 10.1038/nrdp.2017.13.

[43] Carol P. Weingarten et al. «Neuroimaging of Parkinson's Disease: Expanding views.» In: *Neuroscience and Biobehavioral Reviews* 59.1 (2015), pp. 16–52. DOI: 10.1016/j.physbeh.2017.03.040.

[44] E. Ray Dorsey et al. «Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016.» In: *The Lancet Neurology* 17.11 (2018), pp. 939–953. ISSN: 14744465. DOI: 10.1016/S1474-4422(18)30295-3.

[45] V. L. Feigin et al. «Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015.» In: *The Lancet Neurology* 16.11 (2017), pp. 877–897. ISSN: 14744465. DOI: 10.1016/S1474-4422(17)30299-5.

[46] Walter A. Rocca. «The burden of Parkinson's disease: a worldwide perspective.» In: *The Lancet Neurology* 17.11 (2018), pp. 928–929. ISSN: 14744465. DOI: 10.1016/S1474-4422(18)30355-7. URL: http://dx.doi.org/10.1016/S1474-4422(18)30355-7.

[47] European's Parkinson's Association Disease. *The European Parkinson's Disease Standards of Care Consensus Statement*. London, 2011. URL: https://www.epda.eu.com/latest/resources/the-european-parkinsons-disease-standards-of-care-consensus-statement/.

[48] Saul Martinez-Horta and Jaime Kulisevsky. «Mild cognitive impairment in Parkinson's disease.» In: *Journal of Neural Transmission* 126.7 (2019), pp. 897–904. ISSN: 14351463. DOI: 10.1007/s00702-019-02003-1. URL: https://doi.org/10.1007/s00702-019-02003-1.

[49] Manlio De Domenico. «Multilayer modeling and analysis of human brain networks.» In: *GigaScience* 6.5 (2017), pp. 1–8. ISSN: 2047217X. DOI: 10.1093/gigascience/gix004.

[50] Charles Jean; Bransden, Brian Harold; Joachain. *Physics of Atoms and Molecules*. 1st. London: Longman Publishing Group, 1982, p. 207.

[51] Matthew Rowe et al. «Concepts of Diffusion in MRI.» In: *Diffusion Tensor Imaging*. Ed. by Wim Van Hecke, Louise Emsell, and Stefan Sunaert. New York: Springer,

2016. Chap. 3, pp. 23–36. ISBN: 978-1-4939-3117-0. DOI: 10.1007/978-1-4939-3118-7.

[52] Koene R.A. Van Dijk et al. «Intrinsic functional connectivity as a tool for human connectomics: Theory, properties, and optimization.» In: *Journal of Neurophysiology* 103.1 (2010), pp. 297–321. ISSN: 00223077. DOI: 10.1152/jn.00783.2009.

[53] Louise Emsel, Wim Van Hecke, and Jacques-Donald Tournier. «Introduction to Diffusion Tensor Imaging.» In: *Diffusion Tensor Imaging: A Practical Handbook*. Ed. by Wim Van Hecke, Louise Emsell, and Stefan Sunaert. 1st ed. New York: Springer - New York, 2016. Chap. 2, pp. 7–19. ISBN: 9781493931187. DOI: 10.1007/978-1-4939-3118-7.

[54] Henrietta Tomán, Róbert Tornai, and Marianna Zichar. «Complex fiber visualization.» In: *Annales Mathematicae et Informaticae* 34.November (2007), pp. 103–109. ISSN: 17876117.

[55] Thijs Dhollander. «From Diffusion to the Diffusion Tensor.» In: *Diffusion Tensor Imaging*. Ed. by Wim Van Hecke, Louise Emsell, and Stefan Sunaert. New York: Springer, 2016. Chap. 4, pp. 37–65. ISBN: 978-1-4939-3117-0. DOI: 10.1007/978-1-4939-3118-7.

[56] Olaf Sporns, Giulio Tononi, and Rolf Kötter. «The human connectome: A structural description of the human brain.» In: *PLoS Computational Biology* 1.4 (2005), pp. 0245–0251. ISSN: 15537358. DOI: 10.1371/journal.pcbi.0010042.

[57] Santiago Ramón y Cajal. *Texture of the Nervous System of Man and the Vertebrates*. 1st ed. Wien: Springer-Verlag, 1999. ISBN: 9783709173237. DOI: 10.16309/j.cnki.issn.1007-1776.2003.03.004.

[58] Scott W Emmons. «Commentary on the Mind of a Worm.» In: *Phylosophical Transactions of the Royal Society B* 370 (2015), p. 20140309.

[59] Steven J. Cook et al. «Whole-animal connectomes of both Caenorhabditis elegans sexes.» In: *Nature* 571.7763 (2019), pp. 63–71. ISSN: 14764687. DOI: 10.1038/s41586-019-1352-7. URL: http://dx.doi.org/10.1038/s41586-019-1352-7.

[60] Ed Bullmore and Olaf Sporns. «Complex brain networks: Graph theoretical analysis of structural and functional systems.» In: *Nature Reviews Neuroscience* 10.3 (2009), pp. 186–198. ISSN: 1471003X. DOI: 10.1038/nrn2575.

[61] Olaf Sporns. «The human connectome: A complex network.» In: *Annals of the New York Academy of Sciences* 1224.1 (2011), pp. 109–125. ISSN: 17496632. DOI: 10.1111/j.1749-6632.2010.05888.x.

[62] Hae Jeong Park and Karl Friston. «Structural and functional brain networks: From connections to cognition.» In: *Science* 342.6158 (2013). ISSN: 10959203. DOI: 10.1126/science.1238411.

[63] Stuart J Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 3rd ed. Pearson Education, 2003, p. 1151. ISBN: 9780136042594. DOI: 10.1017/S0269888900007724. arXiv: 9809069v1 [arXiv:gr-qc].

[64] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. 2nd ed. Stanford: Springer, 2008, pp. 101–110, 295–330. DOI: 10.1007/978-1-4419-9863-7_941.

[65] Matt Gates. *Machine Learning for beginners*. Ed. by Audrey Swans. 1st ed. Auva Press, 2017, pp. 15–80. ISBN: 9781547039043.

[66] L. Breiman et al. *Classification and Regression Trees*. Wadsworth: Taylor & Francis, 1984. ISBN: 0412048418. DOI: 10.1895/wormbook.1.67.2.

[67] Martin Krzywinski and Naomi Altman. «Classification and regression trees.» In: *Nature Methods* 14.8 (2017), pp. 757–758. ISSN: 15487105. DOI: 10.1038/nmeth.4370.

[68] Leo Breiman. «Bagging Predictors.» In: *Machine Learning* 24.421 (1996), pp. 123–140. ISSN: 0885-6125. DOI: 10.1007/BF00058655.

[69] Benjamin Katz, Priti Shah, and David E. Meyer. «How to play 20 questions with nature and lose: Reflections on 100 years of brain-training research.» In: *Proceedings of the National Academy of Sciences of the United States of America* 115.40 (2018), pp. 9897–9904. ISSN: 10916490. DOI: 10.1073/pnas.1617102114.

[70] Sandra B. Chapman et al. «Neural mechanisms of brain plasticity with complex cognitive training in healthy seniors.» In: *Cerebral Cortex* 25.2 (2015), pp. 396–405. ISSN: 14602199. DOI: 10.1093/cercor/bht234.

[71] Philip D. Harvey et al. «Controversies in Computerized Cognitive Training.» In: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 3.11 (2018), pp. 907–915. ISSN: 24519030. DOI: 10.1016/j.bpsc.2018.06.008.

[72] Nooshin Abbasi et al. «Predicting severity and prognosis in Parkinson's disease from brain microstructure and connectivity.» In: *NeuroImage: Clinical* 25 (2020), pp. 102–111. ISSN: 2213-1582. DOI: 10.1016/j.nicl.2019.102111.

[73] Carol P. Weingarten et al. «Neuroimaging of Parkinson's disease: Expanding views.» In: *Neuroscience and Biobehavioral Reviews* 59 (2015), pp. 16–52. ISSN: 18737528. DOI: 10.1016/j.neubiorev.2015.09.007. URL: http://dx.doi.org/10.1016/j.neubiorev.2015.09.007.

[74] Peng Li et al. «Electroconvulsive therapy-induced brain functional connectivity predicts therapeutic efficacy in patients with schizophrenia: A multivariate pattern recognition study.» In: *npj Schizophrenia* 3 (2017). ISSN: 2334265X. DOI: 10.1038/s41537-017-0023-7.

[75] Michael W. Weiner et al. «The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement.» In: *Alzheimer's and Dementia* 13.5 (2017), pp. 561–571. ISSN: 15525279. DOI: 10.1016/j.jalz.2016.10.006. URL: http://dx.doi.org/10.1016/j.jalz.2016.10.006.

[76] J. Donald Tournier et al. «MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation.» In: *NeuroImage* 202.August (2019), p. 116137. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2019.116137. URL: https://doi.org/10.1016/j.neuroimage.2019.116137.

[77] Jesper L.R. Andersson and Stamatios N. Sotiropoulos. «An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging.» In: *NeuroImage* 125 (2016), pp. 1063–1078. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2015.10.019. URL: http://dx.doi.org/10.1016/j.neuroimage.2015.10.019.

[78] Stephen M. Smith et al. «Advances in functional and structural MR image analysis and implementation as FSL.» In: *NeuroImage* 23.SUPPL. 1 (2004), pp. 208–219. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2004.07.051.

[79] Jesper L.R. Andersson, Stefan Skare, and John Ashburner. «How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging.» In: *NeuroImage* 20.2 (2003), pp. 870–888. ISSN: 10538119. DOI: 10.1016/S1053-8119(03)00336-7.

[80] Oscar Esteban et al. «fMRIPrep: a robust preprocessing pipeline for functional MRI.» In: *Nature Methods* 16.1 (2019), pp. 111–116. ISSN: 15487105. DOI: 10.1038/s41592-018-0235-4.

[81] Raimon H.R. Pruim et al. «ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data.» In: *NeuroImage* 112 (2015), pp. 267–277. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2015.02.064. URL: http://dx.doi.org/10.1016/j.neuroimage.2015.02.064.

[82] Linden Parkes et al. «An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI.» In: *NeuroImage* 171.July 2017 (2018), pp. 415–436. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2017.12.073. URL: https://doi.org/10.1016/j.neuroimage.2017.12.073.

[83] Aslak Grinsted. *Cross wavelet and wavelet coherence*. 2020. URL: https://github.com/grinsted/wavelet-coherence (visited on 04/10/2019).

[84] A. Grinsted, J. C. Moore, and S. Jevrejeva. «Application of the cross wavelet transform and wavelet coherence to geophysical time series.» In: *Nonlinear Processes in Geophysics* 11.5/6 (2004), pp. 561–566. ISSN: 1023-5809. DOI: 10.5194/npg-11-561-2004.

[85] Zitong Zhang et al. «Choosing wavelet methods, filters, and lengths for functional brain network construction.» In: *PLoS ONE* 11.6 (2016), pp. 1–24. ISSN: 19326203. DOI: 10.1371/journal.pone.0157243. arXiv: 1512.07596. URL: http://dx.doi.org/10.1371/journal.pone.0157243.

[86] Chris Vriend et al. «Resting-state network topology and planning ability in healthy adults.» In: *Brain Structure and Function* 225.1 (2020), pp. 365–374. ISSN: 18632661. DOI: 10.1007/s00429-019-02004-6. URL: https://doi.org/10.1007/s00429-019-02004-6.

[87] Sophie Achard et al. «A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs.» In: *Journal of Neuroscience* 26.1 (2006), pp. 63–72. ISSN: 02706474. DOI: 10.1523/JNEUROSCI.3874-05.2006.

[88] Colin R. Buchanan et al. «The effect of network thresholding and weighting on structural brain networks in the UK Biobank.» In: *NeuroImage* 211 (2020), p. 116443. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2019.116443. URL: https://doi.org/10.1016/j.neuroimage.2019.116443.

[89] Martijn P. van den Heuvel et al. «Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations.» In: *NeuroImage* 152.February (2017), pp. 437–449. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2017.02.005.

[90] Bernadette C.M. van Wijk, Cornelis J. Stam, and Andreas Daffertshofer. «Comparing brain networks of different size and connectivity density using graph theory.» In: *PLoS ONE* 5.10 (2010). ISSN: 19326203. DOI: 10.1371/journal.pone.0013701.

[91] Robert E. Smith et al. «SIFT: Spherical-deconvolution informed filtering of tractograms.» In: *NeuroImage* 67 (2013), pp. 298–312. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2012.11.049.

[92] Peter McColgan et al. «Stability and sensitivity of structural connectomes: effect of thresholding and filtering and demonstration in neurodegeneration.» In: *bioRxiv* September (2018), p. 416826. DOI: 10.1101/416826.

[93] Micah Allen et al. «Raincloud plots: A multi-platform tool for robust data visualization.» In: *Wellcome Open Research* 4 (2019), pp. 1–40. ISSN: 2398502X. DOI: 10.12688/wellcomeopenres.15191.1.

[94] Alexander Strang et al. «Generalized relationships between characteristic path length, efficiency, clustering coefficients, and density.» In: *Social Network Analysis and Mining* 8.1 (2018), pp. 1–8. ISSN: 18695469. DOI: 10.1007/s13278-018-0492-3. arXiv: arXiv:1702.02621v2.

[95] B. T.Thomas Yeo et al. «The organization of the human cerebral cortex estimated by intrinsic functional connectivity.» In: *Journal of Neurophysiology* 106.3 (2011), pp. 1125–1165. ISSN: 00223077. DOI: 10.1152/jn.00338.2011.

[96] C.F.W. Peeters. *FMradio: Factor modeling for radiomic data*. 2019. URL: https://cran.r-project.org/web/packages/FMradio/index.html.

[97] Fabian Pedregosa et al. «Scikit-learn: Machine Learning in Python.» In: *Journal of Machine Learning Research* 12.85 (2011), pp. 2825–2830. URL: http://jmlr.org/papers/v12/pedregosa11a.html.

[98] James P Trujillo et al. «Impaired Planning in Parkinson's Disease is Reflected by Reduced Brain Activation and Connectivity.» In: (2015). DOI: 10.1002/hbm.22873.

[99] Dino Muslimović et al. «Cognitive decline in Parkinson's disease: A prospective longitudinal study.» In: *Journal of the International Neuropsychological Society* 15.3 (2009), pp. 426–437. ISSN: 14697661. DOI: 10.1017/S1355617709090614.

[100] Muriel Deutsch Lezak et al. *Neuropsychological Assessment*. 5th. New York: Oxford University Press, 2012, pp. 637–640.

[101] Kevin Duff. «Current topics in science and practice evidence-based indicators of neuropsychological change in the individual patient: Relevant concepts and methods.» In: *Archives of Clinical Neuropsychology* 27.3 (2012), pp. 248–261. ISSN: 18735843. DOI: 10.1093/arclin/acr120.

[102] James Bergstra et al. «Algorithms for hyper-parameter optimization.» In: *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011* (2011), pp. 1–9.

[103] Barrett Williams. *A Comparison of Bayesian Packages for Hyperparameter Optimization*. 2020. DOI: 10.1007/978-1-4684-9388-7. URL: https://sigopt.com/blog/comparison-bayesian-packages-hyperparameter-optimization/ (visited on 09/16/2020).

[104] John P.A. Ioannidis. «Why most discovered true associations are inflated.» In: *Epidemiology* 19.5 (2008), pp. 640–648. ISSN: 10443983. DOI: 10.1097/EDE.0b013e31818131e7.

[105] Angela Dean and Daniel Voss. *Design and Analysis of Experiments*. New York: Springer-Verlag New York, 1999, pp. 112–113. DOI: 10.1080/16089677.2016.1228745.

[106] Zaixu Cui and Gaolang Gong. «The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features.» In: *NeuroImage* 178 (2018), pp. 622–637. ISSN: 1053-8119.

DOI: https://doi.org/10.1016/j.neuroimage.2018.06.001. URL: http://www.sciencedirect.com/science/article/pii/S1053811918305081.

[107] Gaël Varoquaux. «Cross-validation failure: Small sample sizes lead to large error bars.» In: *NeuroImage* 180 (2018), pp. 68–77. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2017.06.061. arXiv: 1706.07581.

[108] Chao Zhang et al. «Sex and Age Effects of Functional Connectivity in Early Adulthood.» In: *Brain Connectivity* 6.9 (Aug. 2016), pp. 700–713. ISSN: 2158-0014. DOI: 10.1089/brain.2016.0429. URL: https://doi.org/10.1089/brain.2016.0429.

[109] Meenakshi Khosla et al. «Ensemble learning with 3D convolutional neural networks for functional connectome-based prediction.» In: *Neuroimage* 199 (2019), pp. 651–662. DOI: 10.1016/j.neuroimage.2019.06.012.Ensemble.

[110] D H Wolpert and W G Macready. «No Free Lunch Theorems for Optimization.» In: *Trans. Evol. Comp* 1.1 (Apr. 1997), pp. 67–82. ISSN: 1089-778X. DOI: 10.1109/4235.585893. URL: https://doi.org/10.1109/4235.585893.

[111] Colin J Brown and Ghassan Hamarneh. «Machine Learning on Human Connectome Data from MRI.» In: *CoRR* abs/1611.0 (2016). arXiv: 1611.08699. URL: http://arxiv.org/abs/1611.08699.

[112] Bo Yong Park et al. «Structural and functional brain connectivity of people with obesity and prediction of body mass index using connectivity.» In: *PLoS ONE* 10.11 (2015), pp. 4–6. ISSN: 19326203. DOI: 10.1371/journal.pone.0141376.

[113] Colin J Brown et al. «Predictive Subnetwork Extraction with Structural Priors for Infant Connectomes BT - Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016.» In: ed. by Sebastien Ourselin et al. Cham: Springer International Publishing, 2016, pp. 175–183. ISBN: 978-3-319-46720-7.

[114] Jeremy Kawahara et al. «BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment.» In: *NeuroImage* 146 (2017), pp. 1038–1049. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2016.09.046. URL: http://dx.doi.org/10.1016/j.neuroimage.2016.09.046.

[115] G. Ball et al. «Machine-learning to characterise neonatal functional connectivity in the preterm brain.» In: *NeuroImage* 124 (Jan. 2016), pp. 267–275. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2015.08.055.

[116] Hailun Sun et al. «Preliminary prediction of individual response to electroconvulsive therapy using whole-brain functional magnetic resonance imaging data.» In: *NeuroImage: Clinical* 26.May (2020). ISSN: 22131582. DOI: 10.1016/j.nicl.2019.102080.

[117] Haifeng Zhang et al. «Effect of computerised cognitive training on cognitive outcomes in mild cognitive impairment: A systematic review and meta-analysis.» In: *BMJ Open* 9.8 (2019). ISSN: 20446055. DOI: 10.1136/bmjopen-2018-027062.

[118] Peter Giacobbe et al. «Improvements in Health-Related Quality of Life with Electroconvulsive Therapy: A Meta-Analysis.» In: *Journal of ECT* 34.2 (2018), pp. 87–94. ISSN: 15334112. DOI: 10.1097/YCT.0000000000000486.

[119] Katherine S. Button et al. «Power failure: Why small sample size undermines the reliability of neuroscience.» In: *Nature Reviews Neuroscience* 14.5 (2013), pp. 365–376. ISSN: 1471003X. DOI: 10.1038/nrn3475.

[120] Christopher J. Kelly et al. «Key challenges for delivering clinical impact with artificial intelligence.» In: *BMC Medicine* 17.1 (2019), pp. 1–9. ISSN: 17417015. DOI: 10.1186/s12916-019-1426-2.

[121] Stephane Mallat. *A Wavelet Tour of Signal Processing - The Sparse Way*. 2nd. Elsevier, 2009. DOI: 10.1016/B978-0-12-374370-1.X0001-8.

[122] MathWorks. *Continuous 1-D wavelet transform - MATLAB cwt - MathWorks Benelux*. URL: https://nl.mathworks.com/help/wavelet/ref/cwt.html (visited on 09/28/2020).

Appendix A

Normality

In this section the kurtosis, skewness and variance values for the set of 10 optimal RMSE estimated via the 10-5-nested-CV process will be presented.

Table A.1 – DF−F Performance Estimation Descriptive Statistics

	Kurtosis	Skewness	Variance
Age ($n = 42$)	0.29	-0.45	3.21
Age ($n = 80$)	0.37	0.69	1.05
Mean Reaction Time Load S4	-1.95	-0.45	0.04
Mean Reaction Time Load S5	1.21	-0.25	0.02
Percentage Correct Load S4	-1.23	0.44	0.02
Percentage Correct Load S5	-0.96	-0.43	0.14

Table A.2 – DT−N Performance Estimation Descriptive Statistics

	Kurtosis	Skewness	Variance
Age ($n = 42$)	1.74	-1.08	2.23
Age ($n = 80$)	-0.87	-0.27	1.68
Mean Reaction Time Load S4	0.56	-1.08	0.05
Mean Reaction Time Load S5	0.36	0.44	0.02
Percentage Correct Load S4	1.56	-0.45	0.02
Percentage Correct Load S5	0.00	-0.97	0.12

Table A.3 – RF−F Performance Estimation Descriptive Statistics

	Kurtosis	Skewness	Variance
Age ($n = 42$)	-1.11	0.14	1.72
Age ($n = 80$)	-0.74	-0.27	1.00
Mean Reaction Time Load S4	-1.97	0.14	0.05
Mean Reaction Time Load S5	-0.48	0.12	0.02
Percentage Correct Load S4	0.28	0.13	0.02
Percentage Correct Load S5	-0.21	-0.73	0.12

Table A.4 – RF−N Performance Estimation Descriptive Statistics

	Kurtosis	Skewness	Variance
Age ($n = 42$)	0.18	0.88	2.34
Age ($n = 80$)	-0.10	-0.27	0.78
Mean Reaction Time Load S4	-1.20	0.88	0.06
Mean Reaction Time Load S5	-1.46	0.13	0.02
Percentage Correct Load S4	1.17	0.97	0.02
Percentage Correct Load S5	-0.20	-1.07	0.10

Table A.5 – `Baseline`Performance Estimation Descriptive Statistics

	Kurtosis	Skewness	Variance
Age ($n = 42$)	-0.68	0.44	1.87
Age ($n = 80$)	0.21	-0.58	1.61
Mean Reaction Time Load S4	-1.99	0.44	0.06
Mean Reaction Time Load S5	-0.71	0.10	0.02
Percentage Correct Load S4	1.07	0.72	0.02
Percentage Correct Load S5	0.60	-1.30	0.13

Appendix B

Optimal Training Logs for RF−F

In this section the optimal configurations from the 10 different iterations of the TPE for the RF−F will be presented. Each line on the tables below represent an optimal configuration determined by different run of the TPE algorithm in the 10-5-nested-CV process.

The main goal of this section is to illustrate the instability of the tuning process (i.e. vibration effects). The training logs from RF−F were chosen because this is the algorithm that needs optimisation over the biggest number of hyperparameters to be fully closed. The 10 optimal configurations determined by the 10 separate runs of the tuning process for the other three algorithms are equally unstable in their respective search domains.

As we can see from the table below, no two iterations of the tuning process of RF−F reached the exact same optimal configuration either in the prediction of the cognitive target variables or in age prediction.

Table B.1 – Optimal parameter at the end of optimisation for prediction of RCI of the Mean Reaction Time on Load S4

max_depth	max_features	max_samples	min_samples_leaf	n_estimators	number_of_factors	redundancy_threshold
2	0.19	0.11	0.39	150	23	0.90
3	0.89	0.49	0.11	1200	24	0.93
4	0.46	0.18	0.29	150	28	0.91
8	0.83	0.48	0.12	500	31	0.92
2	0.38	0.11	0.39	150	20	0.90
11	0.75	0.46	0.12	2400	10	0.91
8	0.75	0.47	0.15	750	17	0.93
14	0.78	0.48	0.10	2700	4	0.92
2	0.41	0.46	0.11	2100	20	0.93
4	0.42	0.43	0.11	850	12	0.95

Table B.2 – Optimal parameter at the end of optimisation for prediction of the RCI of the Mean Reaction Time on Load S5

max_depth	max_features	max_samples	min_samples_leaf	n_estimators	number_of_factors	redundancy_threshold
12	0.737	0.491	0.107	1800	9	0.95
13	0.446	0.484	0.145	1100	17	0.91
9	0.621	0.479	0.105	800	2	0.91
13	0.588	0.458	0.101	2100	13	0.95
3	0.887	0.492	0.106	1100	24	0.93
11	0.562	0.399	0.100	1350	36	0.90
2	0.753	0.495	0.112	1400	12	0.90
10	0.765	0.464	0.112	2450	17	0.94
4	0.525	0.466	0.100	1450	17	0.94
2	0.680	0.426	0.100	750	12	0.92

Table B.3 – Optimal parameter at the end of optimisation for prediction of the RCI of the Percentage of Correctly Answered Trials on Load S4

max_depth	max_features	max_samples	min_samples_leaf	n_estimators	number_of_factors	redundancy_threshold
12	0.664	0.489	0.103	2700	13	0.94
2	0.799	0.479	0.132	1100	18	0.93
13	0.852	0.448	0.104	200	31	0.94
4	0.633	0.420	0.100	400	17	0.92
3	0.827	0.448	0.102	200	26	0.94
11	0.720	0.463	0.115	1100	26	0.91
8	0.814	0.476	0.122	2250	9	0.93
3	0.664	0.500	0.109	600	27	0.92
2	0.179	0.104	0.397	100	9	0.90
4	0.709	0.500	0.136	2400	20	0.90

Table B.4 – Optimal parameter at the end of optimisation for prediction of the RCI of the Percentage of Correctly Answered Trials on Load S5

max_depth	max_features	max_samples	min_samples_leaf	n_estimators	number_of_factors	redundancy_threshold
12	0.714	0.492	0.100	2100	10	0.90
13	0.532	0.487	0.108	950	36	0.92
12	0.328	0.144	0.369	100	14	0.92
7	0.842	0.386	0.114	2150	18	0.93
13	0.300	0.482	0.107	2000	30	0.94
10	0.369	0.411	0.100	400	15	0.92
12	0.315	0.161	0.438	100	27	0.95
11	0.856	0.489	0.111	1900	16	0.95
6	0.468	0.126	0.390	100	19	0.95
8	0.205	0.127	0.374	100	33	0.92

Table B.5 – Optimal parameter at the end of optimisation for prediction of Age on Intervention+Control Groups ($n = 80$)

max_depth	max_features	max_samples	min_samples_leaf	n_estimators	number_of_factors	redundancy_threshold
7	0.705	0.453	0.149	900	13	0.93
2	0.825	0.478	0.100	1550	21	0.93
12	0.432	0.464	0.102	1450	19	0.93
6	0.716	0.485	0.196	350	18	0.92
3	0.824	0.472	0.103	2000	22	0.94
11	0.862	0.472	0.100	400	25	0.95
14	0.630	0.461	0.176	2050	21	0.95
8	0.843	0.484	0.156	1450	23	0.95
9	0.842	0.329	0.111	650	9	0.92
8	0.701	0.438	0.168	1050	35	0.91

Table B.6 – Optimal parameter at the end of optimisation for prediction of Age on Intervention Group ($n = 42$)

max_depth	max_features	max_samples	min_samples_leaf	n_estimators	number_of_factors	redundancy_threshold
14	0.352	0.418	0.144	2650	34	0.92
12	0.492	0.482	0.151	1650	8	0.91
9	0.552	0.396	0.147	1850	5	0.94
9	0.693	0.403	0.110	1950	31	0.95
12	0.288	0.492	0.127	100	25	0.95
6	0.566	0.485	0.119	2700	27	0.93
4	0.721	0.437	0.115	2750	5	0.92
5	0.434	0.448	0.100	850	28	0.91
15	0.724	0.438	0.147	2150	22	0.91
14	0.352	0.418	0.144	2650	34	0.92

ANNEX I

GRAPH MEASURES

All of the definitions below only report to measures on weighted undirected graphs. For the analogous definitions on binary and/or directed graphs, cf. [3]. All of these measures are defined according to [5] with small changes in notation.

For the purposes of this section, N refers to the set of nodes of a given network and n to the number of nodes of a network, i.e. the cardinality of N. Moreover, w_{ij} refers to the weight of the edge between nodes i and j.

I.1 Base Measures

Degree

The **binary** degree of node i $\left(k_i\right)$ quantifies how many nodes are directly connected to node i. To calculate the binary degree we need to first transform the weighted graph into a binary graph. For illustration purposes let's consider the ideal thresholding function, f, that maps weighted edges w_{ij} into binary edges a_{ij}:

$$a_{ij} = f(w_{ij}) = \begin{cases} 1, & \text{if } w_{ij} \neq 0 \\ 0, & \text{if } w_{ij} = 0 \end{cases} \tag{I.1}$$

Thus, upon the mapping of network edges (equation I.1), the degree of a node can be

calculated by:

$$k_i = \sum_{j \in N} a_{ij} \tag{I.2}$$

The **weighted** degree of a node $\left(s_i\right)$ is the sum of the weights of the edges connected to that node. This measure is referred to as the strength of the node.

$$s_i = \sum_{j \in N} w_{ij} \tag{I.3}$$

Triangles

Another important measure of a network is in how many triangles does node i partake (t_i). For a undirected weighted definition, we have to account for edge weights. Therefore, we use the geometric mean of the weights. Note that if $w_{ix} = 0$, there is no triangle between node i and node x.

$$t_i = \frac{1}{2} \sum_{j,k \in N} \sqrt[3]{w_{ij} \cdot w_{ik} \cdot w_{jk}} \tag{I.4}$$

I.2 Measures of segregation

(Average) Clustering coefficient

In informal language, the clustering coefficient of node i $\left(C_i\right)$ is the the proportion of the nodes that are connected to node i (neighbours) that are connected to each other. See figure I.1 for an illustration of this principle. Equivalently, it refers to the number of triangles that contain node i as a vertex. Or, simply put, how many of the neighbours of node i are neighbours of each other. Therefore, the clustering coefficient provides us important information related to segregation of the network. The clustering coefficient of a network $\left(C\right)$ is the average of the clustering coefficients of all its nodes .

$$C = \frac{1}{n} \sum_{i \in N} C_i = \frac{1}{n} \sum_{i \in N} \frac{2t_i}{k_i \left(k_i - 1\right)} \tag{I.5}$$

where k_i and t_i have the meaning of equations I.2 and I.4, respectively

Transitivity

Transitivity is related to clustering in a nuanced way. Transitivity estimates the probability that any two nodes connected to a third are also connected to each other, whereas

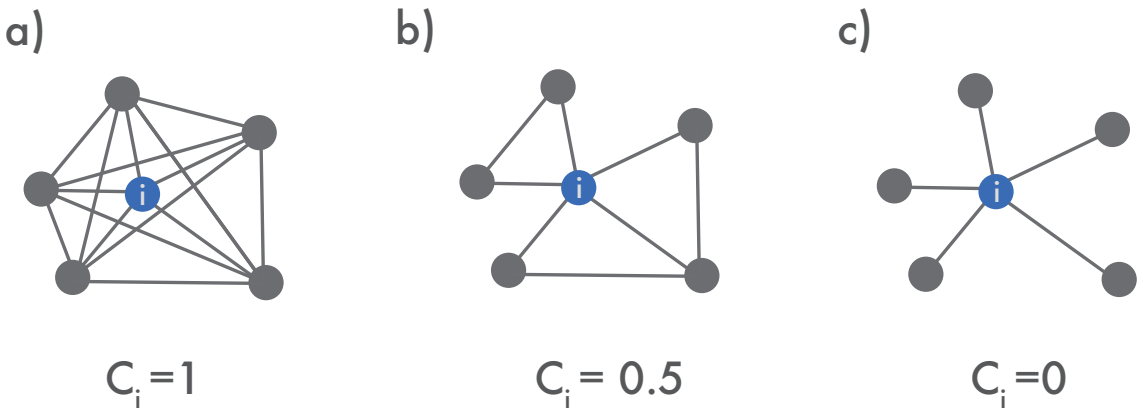

Figure I.1 – Illustration of the clustering coefficient a small network. *a* | All of the neighbours of node *i* are interconnected. In this way node *i* has the highest clustering coefficient possible. *b* | Intermediate situation. *c* | None of the neighbour of node *i* is connected. Node *i* has the lowest clustering coefficient possible.

the clustering coefficient estimates the proportion of closed triangles attached to a node, separately for each node.

$$T = \frac{\sum_{i \in N} 2t_i}{\sum_{i \in N} k_i \, (k_i - 1)} \tag{I.6}$$

I.3 Measures of integration

Characteristic Path Length

A **walk** is a set of edges in an undirected graph that is ordered to form a sequence in which any pair of successive edges shares a common node. A **path** is a walk where in which all edges and nodes are unique. The path length refers to the sum of the edge weights in the path. The **shortest path length** $\left(l_{ij}\right)$ between nodes *i* and *j* refers to an ordered set of edges linking two nodes in a network for which the sum of the weights of its constituent edges is minimal (found through Dijkstra's Algorithm). The **characteristic path length** $\left(\lambda\right)$ of a network is the average shortest path length between all possible pairs of nodes in a network (equation I.7).

$$\Lambda = \frac{1}{n(n-1)} \sum_{i \neq j} l_{ij} \tag{I.7}$$

Global efficiency

Efficiency is closely related to the shortest path. The characteristic path length is an ill-defined concept in fragmented networks, i.e. if there is no path between node *i* and

node j. The concept of global efficiency is dependent on the efficiency of information transfer on a system where the information would from from any node to any other node in parallel only via the shortest paths. It's thus related to the reciprocal of the shortest path length.

$$E = \frac{1}{n(n-1)} \sum_{\substack{i,j \in N \\ i \neq j}} \frac{1}{l_{ij}} \tag{I.8}$$

Modularity

A module in a network can be defined a set of nodes that are strongly connected with each other and their degree is higher than what would be expected in a random graph. The formalisation of these properties into a measure is the so-called modularity index Q. See figure 2.14 for a more detailed illustration of this measure.

$$Q = \frac{1}{2W} \sum_{i,j \in N} \left(w_{ij} - e_{ij} \right) \delta(i,j) \tag{I.9}$$

In the equation above, $W = \sum_{i,j \in N} w_{ij}$, is the total weight of unique edges of the network, $e_{ij} = \frac{s_i s_j}{2W}$ is the total connectivity between nodes i and j that is expected by chance (s_i and s_j following the definition of equation I.3) and $\delta(i,j)$ is the Kronecker's Delta.

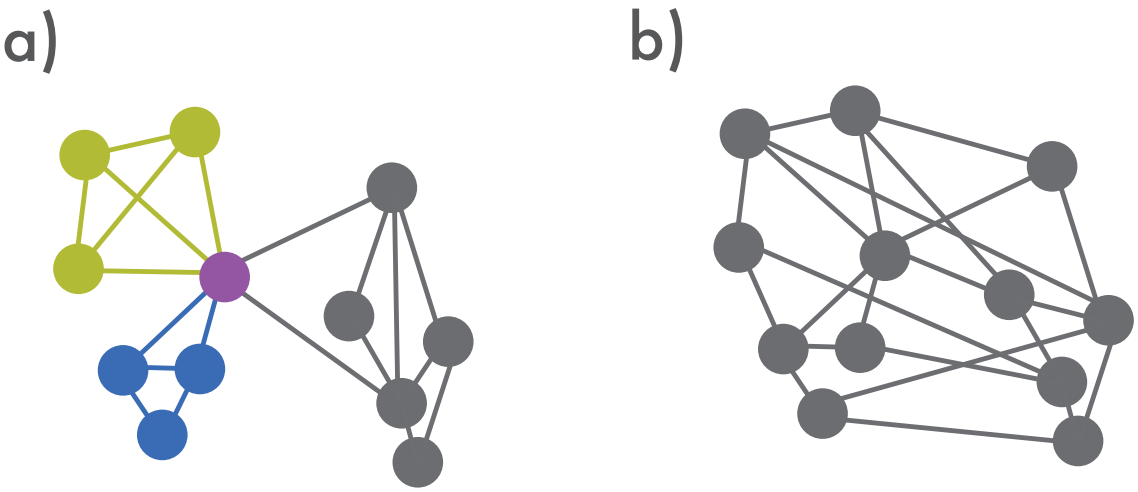

Figure I.2 – Visual comparison of modularity. **a**| An example of a network with relatively high modularity. In yellow green and gray we see three different modules: sub-graphs densely connected with higher degree than what would be expected by chance. The purple node has a higher participation coefficient than any other node in the network because it belongs to all three different modules. **b**| An example network with low modularity. All of the nodes are equally connected to each other, there is no optimal partition.

I.4 Small-worldness

Small-world is a class of networks defined by having properties that are neither completely random or regular. These networks show high clustering coefficients (like lattices) and relatively short shortest path lengths like random graphs. See figure I.3 for illustration.

The definition of small-worldness is relative to a random network. The clustering coefficient has to be relatively higher and the average path length relatively shorter than in a than in a random network. In order to compare these measures so we define the normalised clustering coefficient, γ, and the normalised characteristic path lengths, λ, as the ratio between the measures in our network and in an artificially generated random network. A new random graph with preserved weight, degree and strength distributions should be generated so that the coefficients are comparable.

In this way, the small-worldness, σ, of a network is the ratio between γ and λ, such as:

$$\sigma = \frac{\gamma}{\lambda} = \frac{\Lambda_{\text{rand}} \cdot C}{C_{\text{rand}} \cdot \Lambda} \tag{I.10}$$

where C and Λ have the meaning of equations I.5 and I.7, respectively, and the subscript "rand" refers to the measures in the null-model random network.

The interpretation of the measure follows naturally from its definition. As we discussed, a network shows small-world properties if $\lambda \approx 1$ and $\gamma > 1$. As such, values of σ that are greater than one are often used as a simple indicator of small-world organisation.

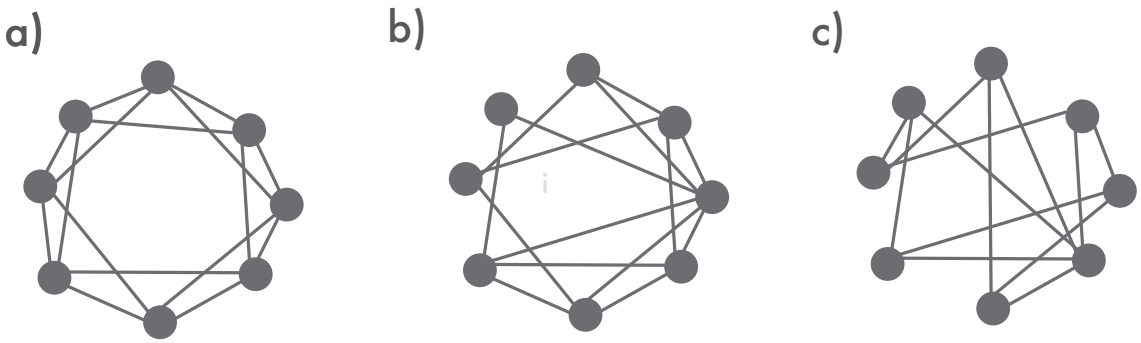

Figure I.3 – Visual comparison between lattice, small-world and random networks. **a**| Regular lattice. **b**| Small-world network. **c**| Random network.

I.5 Subnetwork Measures

Between-network Connectivity

Our subnetworks are disjoint sets of nodes. That is to say, given two different networks α and β, there is no node belonging to α that also belongs to β. The between network connectivity of sub-networks α (composed by n nodes) and β (comprising m nodes) was calculated using the following expression:

$$\mathrm{BN}_{\alpha,\beta} = \frac{1}{n \cdot m} \sum_{\mathrm{i}\in\alpha} \sum_{\mathrm{j}\in\beta} w_{\mathrm{ij}} \tag{I.11}$$

I.5.1 Measures of centrality

The extracted centrality measures are nodal, meaning that there is a uniquely defined value for each node taking into account its neighbourhood. However, they were extracted from subnetworks. We defined the betweenness/eigenvalue centrality of a subnetwork as the average centrality of all its nodes. Likewise for the participation coefficient.

Betweenness Centrality

Betweenness centrality is a measure of how much "a node is in between other two nodes". Alternatively it can be seen as the probability of a node being in a shortest path in between any other two nodes in the network. It measures the proportion of shortest paths between all node pairs in the network that pass through a given index node. The normalised betweenness centrality of node i is

$$C_{\mathrm{B}}(i) = \frac{1}{(N-1)(N-2)} \sum_{\mathrm{i}\neq\mathrm{j}\neq\mathrm{k}} \frac{\rho_{\mathrm{jk}}(i)}{\rho_{\mathrm{jk}}} \tag{I.12}$$

where $\rho_{\mathrm{jk}}(i)$ is the number of shortest paths between nodes j and k and $\rho_{\mathrm{jk}}(i)$ is the number of shortest paths between j and k that include node i.

Participation Coefficient

The participation coefficient is a measure of how much a node's connections are distributed among different modules. For node i,

$$P(i) = 1 - \sum_{\mathrm{m}=1}^{M} \left(\frac{s_{\mathrm{i}}(m)}{s_{\mathrm{i}}} \right)^2 \tag{I.13}$$

where M is the number of modules in the subnetwork, s_i is the strength of node i and $s_i(m)$ is the strength of node i considering a subnetwork defined by the nodes that belong to module m.

Eigenvalue Centrality

Eigenvalue centrality is a nodal measure that accounts for the quantity and quality of connections a specific node in a network. It considers the degree of the node and the degree of its neighbours. This estimation is done using the eigenvectors and eigenvalues of the adjacency matrix of the network.

$$C_E(i) = \frac{1}{\lambda_1} \sum_{j=1}^{N} A_{ij} x_j \tag{I.14}$$

where A is the adjacency matrix of the graph, λ_1 is the largest eigenvector of A and x is its corresponding eigenvalue. x_j represents the j-th entry of eigenvector x.

ANNEX II

Concepts of Wavelet Analysis

This appendix is based on the 2004 paper by Grinsted et al [84], and on the work of Mallat [121].

II.1 Wavelets and Wavelet transform

Although the time-frequency resolution trade-off is an unavoidable fact of physics and exist regardless of the transform used, it is possible to analyse any signal by using an alternative approach called the multiresolution analysis. Even though these methods do not essentially resolved the Heisenberg uncertainty principle, they provide more efficient tools to work around it by extracting maximal information from the signal at different scales. One method of multi-resolution analysis is the wavelet transform.

First, let us define a wavelet. A wavelet is a a square integrable function $\psi \in \mathbf{L}^2(\mathbb{R})$ with zero mean:

$$\int_{-\infty}^{+\infty} \psi(t)\, dt = 0 \tag{II.1}$$

and norm one:

$$\int_{-\infty}^{+\infty} |\psi(t)|^2\, dt = 1 \tag{II.2}$$

To $\psi(t)$ we call the mother wavelet. The mother wavelet can be scaled by $a \in \mathbb{R}^+$ and

shifted by $b \in \mathbb{R}$ to originate the children wavelets:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\,\psi\left(\frac{t-b}{a}\right) \tag{II.3}$$

We use Morlet wavelets in this work. The Morlet wavelet is a complex exponential with a gaussian envelop:

$$\psi_M(t) = e^{i\omega_0 t} e^{-\frac{t^2}{2\sigma^2}} \tag{II.4}$$

where i is the imaginary unit, σ is a measure of support (roughly equivalent spread of the gaussian curve envelop) and ω_0 is the frequency. The Morlet family is thus defined by:

$$\psi_M(t)\big|_{a,b} = \exp\left[i\omega_0\left(\frac{t-b}{a}\right)\right]\exp\left[\frac{t-b}{\sqrt{2}a\sigma}\right]^2 \tag{II.5}$$

The formal definition of the continuous wavelet transform of $f(t)$ is:

$$\Psi_f(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} f(t)\;\psi^*\left(\frac{t-b}{a}\right)dt \tag{II.6}$$

However, in the context of this work, we want to calculate the CWT of a time-series (discrete signal). In order to do so, we must discretise the wavelet respecting the Nyquist Theorem (to be able to reconstruct the signal). More accurately, we discretise the scale parameter according to the Nyquist theorem and the time parameter is then discretised with respect to the scale parameter, i.e. a different sampling rate is used for every scale. The main difference in between discrete and continuous wavelet transforms of time-series is how the scale parameter is discretised. In CWT, we are not forced to discretise the scale parameter as series of powers of two. It is thus common to define a parameter ν, the number of voices per octave or, in other words, the number of intermediate values of the scale parameter in between every doubling of the dilation of the scale. The expression of the CWT becomes defined by the discretised version of the scale parameter. Therefore we can only determine the wavelet coefficients at scales $s = 2^{-j/\nu}$, where $j \in \mathbb{N}$ and ν is the number of voices per octave.[1]

[1] The minimum and maximum scales are defined by the energy spread of the wavelet. Check [122] for more information.

II.2 Wavelet Coherence

Wavelet coherence is a measure of the correlation in between two signals. The wavelet coherence in between time-series x and y is:

$$R^2 = \frac{\left|S\left(\Psi_x(a,b)\Psi_y^*(a,b)\right)\right|^2}{S\left(|\Psi_x(a,b)|^2\right)S\left(\left|\Psi_y(a,b)\right|^2\right)} \tag{II.7}$$

where the superscript $*$ is the complex conjugate and S is a smoothing operator in time and scale for the specific wavelet we are considering [2] For real-valued time series, the wavelet coherence is real-valued if you use a real-valued analysing wavelet, and complex-valued if you use a complex-valued analysing wavelet.

[2] This operator is analytically well-defined for the Morlet wavelet. Cf. [84]

Bernardo de Azevedo Pinto Castro Maciel

Bachelor of Science

Predicting the Outcome of Cognitive Training in Parkinson's Disease using Magnetic Resonance Imaging

Dissertation submitted in partial fulfillment
of the requirements for the degree of

Master of Science in
Biomedical Engineering

February, 2021

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Bernardo de Azevedo Pinto Castro Maciel

Bachelor of Science

Predicting the Outcome of Cognitive Training in Parkinson's Disease using Magnetic Resonance Imaging

Dissertation submitted in partial fulfillment
of the requirements for the degree of

Master of Science in
Biomedical Engineering

February, 2021

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Bernardo Maciel 2021

Predicting the Outcome of Cognitive Training in Parkinson's Disease using Magnetic Resonanc 2021