UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

Normative Model for the Diagnosis of Neuropsychiatric Disorders Using Deep Learning Methods

Duarte Filipe Oliveira Saraiva

Mestrado Integrado em Engenharia Biomédica e Biofísica
Perfil em Engenharia Clínica e Instrumentação Médica

Dissertação orientada por:
Hugo Ferreira

2021

"WHY TALK ABOUT ALL THE KNOWN AND THE UNKNOWN.
SEE HOW THE UNKNOWN MERGES INTO KNOWN."
Muhammad Rumi

Agradecimentos

A dissertação aqui apresentada é o culminar de 5 anos de trabalho árduo. Agradeço desde já ao leitor, por me acompanhar nesta jornada.

Primeiramente, agradeço ao professor Hugo Ferreira, pela orientação e apoio que me deu ao longo deste trabalho. Sem o seu contributo, não teria sido possível realizar esta dissertação. Esteve sempre presente, e ajudou-me a enfrentar os diversos obstáculos que se apresentaram ao longo deste caminho.

Agradeço também a toda a equipa do projeto Neurobioai, em especial à professora Diana Prata, à Helena, ao Vasco, à Vânia, à Carolina e ao Rui, por todas as reuniões semanais, e apoio. Sem eles, não teria sido possível manter o foco, e a motivação. Um obrigado também ao Miguel Fernandes, por me ter esclarecido algumas questões sobre o seu trabalho, no qual esta dissertação foi também inspirada.

Gostaria ainda de agradecer a todos os professores por toda a partilha de conhecimento, ao longo deste 5 anos.

A todos os amigos que fizeram parte desta minha jornada académica, o meu obrigado. Em particular, aos meus companheiros da vida académica, do NE2B2, e da equipa de Futsal da AEFCL. Foram também as atividades extracurriculares que me permitiram ser bem sucedido neste percurso académico.

Agradeço aos "Biomérdicos", por terem estado sempre a meu lado nos bons e maus momentos. Um obrigado especial ao Daniel, ao João, ao Pedro e ao Vasco pelas diversas conversas e chamadas.

Aos meus amigos de Torres, um agradecimento também especial por todas as histórias e bons momentos que hão de perdurar.

À minha família, pelo apoio incondicional. Aos meus companheiros de quatro patas, Trip, Becas, e Magali. Também aos meus avós, que embora ausentes, estarão sempre no meu coração.

Claro que tem de haver um agradecimento especial aos meus manos, Miguel e Raquel. São os meus companheiros de vida, de quem eu gosto tanto, que sempre estiveram comigo. Aos meus pais, Luís e Cristina, que são a base de tudo. Agradeço toda a educação, e amor incondicional. São a minha força, hoje e sempre.

Finalmente, à Lua. Por todo o amor, e apoio. Sem ti, teria sido impossível.

Investigação suportada pela Fundação para a Ciência e Tecnologia (FCT), através dos projetos UID/BIO/00645/2019 e DSAIPA/DS/0065/2018.

Abstract

The diagnosis of neuropsychiatric disorders (NPDs) is still exclusively dependent on the analysis of the signs and symptoms of the patients since there are no biomarkers useful for clinical practice. Considering that several signs and symptoms are shared among different NPDs, the diagnosis is sometimes incorrect. Therefore, therapeutic approaches do not always succeed, which has an impact on the quality of life of neuropsychiatric patients. Furthermore, NPDs have a global economic and demographic impact. For this reason, technological solutions, such as DL, have been researched for the optimization of diagnosis, in the non-technological field of neuropsychiatry.

However, the most promising studies on the diagnosis of NPDs with deep learning (DL) are based on binary classification, which may not be the most adequate approach to deal with the continuous spectrum of NPDs. Here, a DL-based normative model was developed to investigate functional connectivity abnormalities, that may contribute to the development of a novel diagnostic procedure. This method is here used to evaluate how patients deviate from a normal pattern learned by a group of healthy people.

To create and evaluate the normative model, resting-state functional magnetic resonance imaging (rs-fMRI) data from three different databases were used. In order to maximise the balance between the amount and the quality of the data, conditions were defined to restrict the variability of the scan parameters.

Subsequently, rs-fMRI data were trimmed to the lowest number of time points presented in the sample (150). Then, standard preprocessing steps were performed, including removal of the first 4 volumes of functional data, motion correction, spatial smoothing, and high pass filtering. Single-session independent component analysis (ICA) was run, and the FSL-FIX tool was used to clean noise and artefacts. The functional images were then registered to the T1-weighted brain extracted structural images, and finally to the Montreal Neurosciences Institute 152 standard space. Dual regression was applied using fourteen resting-state functional brain networks (FBN) previously identified in the literature. The Pearson's correlation coefficient between the extracted blood oxygen level-dependent (BOLD) time series of each FBN was calculated, and a 14x14 network connectivity matrix was generated for each subject.

The second part of the project consisted of the creation and optimization of a normative model. The normative model consisted of an autoencoder (AE) with three hidden layers. The AE was trained only in healthy subjects and was tested in both healthy subjects and neuropsychiatric patients, including schizophrenia (SCZ), bipolar disorder (BD), and attention deficit hyperactivity disorder (ADHD) patients. The hypothesis was that the model would "fail" on reconstructing data from neuropsychiatric patients.

To evaluate the model performance, graph theory metrics were applied. Besides, the mean squared error was calculated for each feature (correlation between pairs of FBN) to evaluate which regions were worse reconstructed for each group of subjects. The pipeline for NPDs was tested for a SCZ case study, with the addition of a clustering algorithm.

The results of this dissertation revealed that the proposed pipeline was able to identify patterns of functional connectivity abnormality that characterize different NPDs. Moreover, the results found for the two SCZ groups of patients were similar, which demonstrated that the normative model here presented was also generalizable. However, the group-level differences did not withstand individual-level analysis, implying that NPDs are highly heterogeneous. These findings support the idea that a precision-based medical approach, focusing on the specific functional network changes of individual patients, may be more beneficial than the traditional group-based diagnostic classification. A personalised diagnosis would allow for personalised therapy, improving the quality of life of neuropsychiatric patients.

Keywords: Normative Model, Functional Brain Networks, Neuropsychiatric Disorders, Functional Connectivity, Deep Learning

Resumo

O diagnóstico de doenças neuropsiquiátricas (DNP) é realizado exclusivamente através dos sinais e sintomas do paciente, uma vez que até ao momento, não foram encontrados biomarcadores com utilidade para a prática clínica. Considerando que vários sinais e sintomas são partilhados entre diferentes DNP, o diagnóstico destas é bastante complexo. Por esse motivo, muitos doentes são diagnosticados incorretamente ou tardiamente. Como consequência, as abordagens terapêuticas nem sempre são bem sucedidas, e a qualidade de vida dos pacientes com DNP é prejudicada. Além disso, as DNP têm um elevado impacto económico e demográfico a nível mundial, o que abre espaço para a progressão da aprendizagem profunda (AP) no ramo da neuropsiquiatria, onde a tecnologia ainda é praticamente inexistente.

No entanto, os estudos mais promissores no diagnóstico de DNP com AP baseiam-se numa classificação binária, que aparenta não ser a abordagem mais adequada para enfrentar o espectro contínuo das DNP. Neste estudo, foi criado um modelo normativo baseado em AP para investigar anomalias de conectividade funcional, capazes de contribuir para um novo procedimento de diagnóstico. Este método desenvolvido aqui utilizado avalia de que forma os doentes se desviam de um padrão característico de uma amostra de indivíduos saudáveis.

Foram utilizados dados de ressonância magnética funcional em estado de repouso (RMfr) de três bases de dados diferentes para criar e avaliar o modelo normativo. Para maximizar o equilíbrio entre a quantidade e a qualidade dos dados, foram definidas algumas condições para restringir a variabilidade dos parâmetros de ressonância magnética.

Posteriormente, os dados de RMfr foram cortados para o menor número de pontos de tempo apresentado na amostra (150 pontos). Depois, foram realizadas diversas etapas de pré-processamento tradicionais recorrendo ao uso de funcionalidades do software FSL. Nomeadamente, 1) remoção dos primeiros 4 volumes de dados funcionais, 2) correção do movimento da cabeça através da ferramenta MCFLIRT, 3) suavização espacial com um kernel Gaussiano de 5 mm de largura à meia altura, 4) filtragem de passa alto com frequência de corte de 0.01 Hz. De seguida, a análise de componentes independentes foi aplicada, e a ferramenta FSL-FIX foi utilizada para remover ruído e artefactos que estivessem presentes nos dados funcionais. Os dados foram então alinhados com o cérebro padrão *Montreal Neurosciences Institute 152* (MNI152). Para tal, os dados funcionais foram primeiramenete alinhados com as imagens estruturais ponderadas em T1 do mesmo indivíduo, por meio do método *boundary-based registration* (BBR). De seguida, uma transformação não linear foi aplicada de modo a alinhar a imagem estrutural com o cérebro padrão MNI152. Depois, *FSL's Dual Regression* foi aplicada utilizando catorze redes cerebrais funcionais (RCF) previamente identificadas na literatura. Foi calculado o coeficiente de correlação de Pearson entre as séries temporais extraídas de cada RCF, e foi gerada uma matriz de conectividade funcional 14x14 para cada indivíduo.

A segunda parte do projeto consistiu na criação e optimização de um modelo normativo. Um *autoencoder* foi a rede neuronal escolhida para arquitetura do modelo normativo. Um *autoencoder*

consiste numa rede neuronal que tenciona reconstruir os dados que lhe são introduzidos. Os dados de entrada são compactados numa camada mais reduzida, e em seguida, a saída é reconstruída através dessa representação compactada. Neste projeto, o modelo foi treinado apenas com indivíduos saudáveis e foi testado tanto em indivíduos saudáveis, como em doentes neuropsiquiátricos, em particular doentes com perturbações nos espetros da esquizofrenia, doença bipolar, e défice de atenção com hiperatividade. A hipótese colocada neste estudo foi de que o modelo deveria "falhar" mais na reconstrução dos dados dos doentes, do que dos indíviduos saudáveis.

Para avaliar o desempenho do modelo, foram aplicadas métricas de teoria de grafos. Além disso, o erro quadrático médio (EQM) foi calculado para cada correlação entre pares de RCF. Desta forma, tornou-se possível identificar os pares de regiões que foram pior reconstruídos para cada grupo de indivíduos. O modelo normativo foi testado para um caso de estudo de esquizofrenia, em conjunto com um algoritmo de *clustering*, de modo a avalizar a generalização do modelo.

Os resultados desta dissertação revelaram que o modelo proposto foi capaz de identificar padrões de conectividade funcional anormais distintos para cada uma das DNP estudadas. Além disso, os resultados encontrados para os dois grupos de pacientes com esquizofrenia foram semelhantes, o que demonstra que o modelo normativo é generalizável. Para a esquizofrenia, são de realçar as interações de RCF entre a rede dos gânglios da base e a rede *default mode* dorsal, entre a rede dos gânglios da base e a rede *higher visual*, entre a rede dos gânglios da base e a rede da ínsula posterior, e entre a rede *higher visual*, e a rede *executive control* direita, uma vez que aparentam ser úteis para caracterizar a doença. Contudo, as diferenças a nível de grupo não resistiram a uma análise a nível individual, sugerindo que as DNP são extremamente heterogéneas. Estas conclusões reforçam que uma abordagem médica personalizada, centrada nas anomalias da conectividade funcional de cada paciente poderá ser mais vantajosa do que a classificação diagnóstica mais tradicional, baseada no agrupamento de indivíduos. Realizando um diagnóstico personalizado seria possível aplicar também uma terapia personalizada, que em consequência melhoraria a qualidade vida das pessoas com DNP.

Palavras-chave: Modelo Normativo, Redes Cerebrais Funcionais, Perturbações Neuropsiquiátricas, Conectividade Funcional, Aprendizagem Profunda

Scientific Output

D. Saraiva, H. A. Ferreira. 2021. “Effect of BOLD signal data trimming on functional connectivity metrics”. Abstract submission, and poster presentation at 1st ISMRM Iberian Chapter Annual Meeting 2021, June 16-17, 2021 A.2

D. Saraiva, H. A. Ferreira. 2021. “BOLD signal data trimming results in increased functional connectivity”. Abstract submission, and poster presentation at ESMRMB 38th Annual Scientific Meeting, October 7-9, 2021 A.2

Contents

List of Figures

LIST OF FIGURES

List of Tables

Acronyms

AAE adversarial autoencoder. xiii, 21, 22, 23, 28, 29, 38

AD Alzheimer's disease. 1, 28, 29

ADHD attention deficit hyperactivity disorder. iv, xiii, xvi, 1, 24, 25, 28, 30, 31, 32, 33, 38, 41, 45, 46, 47, 50, 51, 61

AE autoencoder. iv, xii, xiv, xvi, 19, 21, 22, 28, 29, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 49, 53, 61

AI artificial intelligence. 2, 17

AN auditory network. xii, xv, 10, 11, 12, 50, 78

ASD autism spectrum disorder. 1, 28, 29

ASN anterior salience network. xii, xiv, xv, 10, 11, 12, 46, 47, 49, 50, 51, 53, 54, 78

ATP adenosine triphosphate. 9

AUC area under the receiver operating characteristic curve. 27

BBR boundary-based registration. 35

BD bipolar disorder. iv, xiii, xvi, 1, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 38, 41, 45, 47, 50, 51, 61

BGN basal ganglia network. xii, xiv, xv, 11, 12, 46, 47, 48, 49, 50, 51, 52, 53, 57, 58, 59, 61, 78

BOLD blood oxygen level-dependent. iv, xii, 6, 9, 10, 13, 15, 16, 34, 37

CAGR compound anual growth rate. 2

$CMRO_2$ cerebral metabolic rate of oxygen. 9

CNS central nervous system. 4, 5

COBRE Center for Biomedical Research Excellence. xiii, xiv, xv, xvi, 31, 32, 33, 34, 38, 41, 42, 53, 54, 56, 57, 59

CoRR Consortium for Reliability and Reproducibility. xiii, xvi, 31, 32, 38

DDMN dorsal default mode network. xii, xiv, xv, 10, 11, 12, 46, 48, 49, 50, 51, 53, 57, 58, 59, 61, 78

deoxy-Hb deoxygenated hemoglobin. 9

Acronyms

DL deep learning. iv, 2, 17, 20, 21, 27, 28, 29, 30, 61, 62

DOF degrees of freedom. 36

DSM-5 The Diagnostic and Statistical Manual of Mental Disorders 5. 1, 2, 24, 26, 59, 61

EC eyes closed. 14, 31

EO eyes open. 14, 31, 33

EO-F eyes fixated on a target. 14, 31, 33

EPI echo-planar imaging. 8, 13

FBN resting-state functional brain networks. iv, xii, xiv, xv, 2, 10, 11, 12, 13, 14, 15, 16, 24, 25, 26, 27, 36, 37, 38, 39, 40, 41, 44, 45, 46, 47, 48, 49, 50, 51, 54, 55, 56, 57, 61, 62, 78

FID free induction decay. 7

fMRI functional magnetic resonance imaging. xiii, 6, 8, 9, 10, 13, 14, 16, 24, 28, 32, 33, 34

GAN generative adversarial network. xiii, 21, 22, 23, 37, 38

HVN higher visual network. xii, xiv, xv, 10, 11, 12, 46, 47, 48, 49, 50, 51, 52, 53, 57, 58, 59, 61, 78

ICA independent component analysis. iv, xii, 13, 14, 15, 16, 27, 34, 35, 36

IR inversion recovery. 8

LECN left executive control network. xii, xiv, xv, 10, 11, 12, 47, 49, 50, 51, 53, 54, 78

LN language network. xii, xiv, xv, 10, 11, 12, 46, 47, 48, 49, 50, 51, 53, 54, 78

MCI mild cognitive impairment. 1, 28, 29

MDD major depressive disorder. 27, 28

ML machine learning. 2, 17, 27, 33, 61

MNI152 Montreal Neurosciences Institute 152. iv, xiii, 14, 35, 36

MPRAGE magnetisation prepared gradient echo imaging. 8

MRI magnetic resonance imaging. 6, 7, 8, 9, 10, 14, 31, 33

MSE$_f$ mean squared error per feature. xiv, 40, 47, 48, 49, 50, 51, 55, 56

MSE$_s$ mean squared error per subject. xiv, xvi, 39, 40, 42, 43, 44, 45, 53, 54

NN neural network. xii, 17, 18, 19, 20, 21, 22, 27

NPD neuropsychiatric disorder. iv, v, xv, 1, 2, 14, 17, 20, 23, 24, 25, 26, 27, 28, 29, 30, 38, 47, 49, 51, 58, 59, 60, 61, 62, 76

O_2 oxygen. 9

oxy-Hb oxygenated hemoglobin. 9

PN precuneus network. xii, xv, 11, 12, 47, 48, 49, 50, 54, 78

PNS peripheral nervous system. 4

PSN posterior salience network. xii, xiv, xv, 10, 11, 12, 46, 47, 49, 50, 51, 52, 53, 57, 58, 59, 61, 78

PVN primary visual network. xii, xiv, xv, 10, 11, 12, 49, 50, 51, 53, 78

RECN right executive control network. xii, xiv, xv, 10, 11, 12, 46, 47, 49, 50, 51, 52, 53, 54, 57, 58, 59, 61, 78

ReLu rectified linear unit. 18, 19, 20, 39, 43, 44

RF radiofrequency. xi, 7, 8, 13

ROI region-of-interest. 15, 36

rs-fMRI resting-state functional magnetic resonance imaging. iv, 2, 10, 12, 13, 14, 15, 16, 17, 26, 28, 30, 31, 33, 34, 61, 62

SCA seed-based correlation analysis. 14, 15

SCZ schizophrenia. iv, v, xiii, xiv, xv, xvi, 1, 3, 25, 26, 27, 28, 29, 30, 31, 32, 33, 38, 41, 45, 47, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61

SCZA schizoaffective. xiii, 31, 32, 33, 34

sMRI structural functional magnetic resonance imaging. 27, 28, 29

SN sensorimotor network. xii, xv, 10, 11, 12, 47, 48, 49, 50, 78

TE echo time. xi, xiii, 8, 9, 31, 32, 33

TR repetition time. xi, xiii, 8, 9, 13, 14, 31, 32, 33

UCLA UCLA Consortium for Neuropsychiatric Phenomics LA5c Study. xiii, xiv, xv, xvi, 31, 32, 33, 38, 41, 42, 44, 51, 52, 53, 54, 55, 56, 57, 58

VAE variational autoencoder. 21, 23, 38

VDMN ventral default mode network. xii, xv, 10, 11, 12, 46, 47, 48, 49, 50, 54, 78

VN visuospatial network. xii, xv, 10, 11, 13, 46, 50, 78

Chapter 1

Introduction

This dissertation proposes a normative model approach to investigate brain functional connectivity changes in groups of neuropsychiatric patients, enabling a new diagnostic procedure. This first chapter exposes the context and motivation for this dissertation, as well as the objectives and structure.

1.1 Context and Motivation

In a technological world, where the "invention" concept is no longer associated with a disruptive idea but with a continuous spectrum, and where most of those inventions are not innovations, it is necessary to think "out of the box" in order to do something new and different.

The diagnosis of psychiatric disorders still depends exclusively on the clinical evaluation of the patient's signs and symptoms and is mainly based on the The Diagnostic and Statistical Manual of Mental Disorders 5 (DSM-5) [1] as no biomarkers of mental disorders have been shown to be useful for clinical practice [2, 3, 4]. This is not surprising since, for the human eye, the brain images of most psychiatric patients are not sufficient to classify the presence or absence of a pathological condition. Psychiatric disorders are a fascinating and abstract world, and we still do not understand clearly what makes a psychiatric disorder. Psychiatrists have been distinguished from other medical doctors because they focus on people's thoughts, perceptions, moods, and behaviours instead of people's skin, bones, or muscles, which leads to a complex diagnostic process [5]. However, while scientists study the brain, they learn that mental disorders are associated with changes in brain structure and function and have a biological basis [6]. This is why the term "neuropsychiatric" has been used to represent a paradigm shift with regard to the responsibility of psychiatrists in the diagnosis and control of behavioural disorders with simultaneous and evident brain pathology [7]. Neuropsychiatry is the integrated neuroscience of neurology and psychiatry that seeks to investigate the psychiatric symptoms of neurological disorders and the neurobiological basis of psychiatric disorders [8]. Although this term is not consensual in the scientific community [8, 7, 5, 9, 10], we use here the definition of Shento and Turetsky [9] that the neuropsychiatric disorders (NPDs) encompass both psychiatry and behavioural neurology disorders, such as schizophrenia (SCZ), bipolar disorder (BD), attention deficit hyperactivity disorder(ADHD), autism spectrum disorder (ASD), mild cognitive impairment (MCI), and Alzheimer's disease (AD).

Considering that these disorders are often related or have coexisting symptoms, the diagnosis becomes more challenging, and consequently, the true prevalence of NPDs globally remains poorly understood. However, they can be analysed as a category that includes psychiatric and neurological disorders. In 2016, neurological disorders were the leading cause of disability and the second leading cause of death worldwide [11]. Additionally, it is estimated that around 11% of the population live with

a psychiatric disorder [12]. In terms of market size, the neurodiagnostic market size was estimated to be USD 14.77 billion in 2019. Its size is anticipated to surpass USD 27.10 billion by 2027, registering a compound anual growth rate (CAGR) of 7.8% [13]. Besides, the global mental health software market was estimated to be worth USD 1.36 billion in 2020, and it is growing at a CAGR of 13.87% to reach a value of USD 2.6 billion by 2025 [14]. These values do not consider the effects of the COVID-19 pandemics, which are expected to boost due to brain sequelae that are being observed in survivors and especially due to the disruption in societal living with lockdown periods with social distancing, and loss of jobs and loved ones [15, 16].

Considering that NPDs have an economic and demographic impact worldwide, the area of progression for artificial intelligence (AI) in this non-technological diagnostic field is promising. But what is preventing or delaying the use of AI in clinical practice? Several researchers believe that AI explainability is the solution for translating good research into clinical practice [17]. This may be true for some medical fields, but are we already at that stage with regard to the diagnosis of NPDs? When it comes to neuropsychiatry, before we explain these AI models, we need to improve these models and change our mindset to do something new and different. The DSM-5 authors recognized that mental disorders do not always fit completely within the boundaries of a single mental disorder and that some symptom domains may involve multiple diagnostic categories [1]. Although there is this new idea that most NPDs develop over a continuous spectrum and that multiple disorders can coexist in a patient [18], most promising studies on mental and neurological disorders diagnosis with machine learning (ML) and deep learning (DL) are based on dichotomous classification [19]. To face this continuous spectrum of NPDs, broader proposals, such as the normative model, must be applied.

1.2 Objectives

The main goal of this master dissertation is to use a normative model approach to investigate alterations in functional connectivity in NPDs based on resting-state functional magnetic resonance imaging (rs-fMRI) data, contributing to a new diagnostic procedure. Nonetheless, other secondary goals must be achieved:

1. Research on resting-state functional brain networks (FBN).
2. Review of current literature to determine the current state of the art on the diagnosis of NPDs, and their functional connectivity abnormalities.
3. Use of well-established methods for processing of rs-fMRI data.
4. Development of an optimized normative model for detecting functional connectivity abnormalities on various NPDs.
5. Application of the normative model to an independent test set.

1.3 Structure

There are five chapters in this dissertation. Chapter 1 introduces the dissertation theme, by describing the context and motivation, as well as the dissertation's objectives and structure. The background theory and state of the art, which are required to support the study, are presented in Chapter 2. Chapter 3 discusses the materials and methods used in this project, including rs-fMRI data selection

and processing, generation of functional connectivity network matrices, development of a normative model, and application to a SCZ case study. In Chapter 4 the findings are presented and discussed. The dissertation comes to a close in Chapter 5 with a summary of the main findings and conclusions, as well as recommendations for future research.

Chapter 2

Background Theory

2.1 The Human Brain

Neurosciences cover a wide range of questions about the structural and functional organization of the nervous system. Neuroscientists have gathered knowledge to answer these questions, and while there is still much to learn, we now have a better understanding of the anatomy and function of the brain. [20].

The nervous system can be divided into the sensory and the motor systems. While the sensory system is responsible for acquiring and processing information from the environment, the motor system responds to that gathered information by generating movement and other types of behaviour. In terms of structure, the nervous system can be divided into two major parts, the central nervous system (CNS), which contains the brain and the spinal cord, and the peripheral nervous system (PNS), which consists of sensory and motor components. The sensory portion incorporates sensory ganglia and nerves that link sensory receptors with CNS processing circuits. The motor portion can be divided into the somatic motor system and the autonomic motor system. The somatic motor system consists of motor axons that connect the brain and spinal cord to skeletal muscles. The autonomic motor system consists of cells and axons that innervate smooth muscles, cardiac muscle, and glands, which in turn can be differentiated into sympathetic, parasympathetic, and enteric divisions [20].

Regarding function, the CNS is responsible for detecting, analysing, and transmitting the information. In general, sensory systems receive information. The brain integrates that information and generates signals to somatic and autonomic pathways to perform specific actions [21]. All these processes depend on the two types of interconnected cells of the nervous system, neurons and glia. Nervous cells, which are made of the cell body, dendrites, axon, and axon terminals, are specialized for electrical signalling over long distances through the generation of action potentials. Neuroglial cells support functions and modulate the activity of neurons [20, 21].

In the CNS, neurons can either be arranged in nuclei, which are local accumulations of neurons that present similar connections and functions (found in the brain, and spinal cord), or in cortices, which describe sheet-like arrays of neurons (found in the cortices of the cerebral hemispheres and cerebellum). Additionally, there is a typical way to distinguish regions rich in neuronal cell bodies: grey matter (e.g., nuclei or cortices), from regions rich in axons: white matter (axon tracts and commissures, due to the presence of myelin) [20]. [see Figure 2.1]

We can divide the CNS into seven basic parts: the spinal cord, medulla, pons, midbrain, cerebellum, diencephalon, and cerebral hemispheres. Running through all of these subdivisions are fluid-filled spaces called ventricles. While the brainstem, which consists of the medulla, pons, and midbrain, surround the 4th ventricle and cerebral aqueduct, the forebrain, which contains the diencephalon and the cerebral

hemispheres encircle the 3rd and lateral ventricles. The brainstem is a channel for various pathways in the CNS, and most sensory and motor information passes through it. It also contains numerous additional nuclei involved in physiological functions, such as control of heart rate, respiration, blood pressure, and level of consciousness. The cerebellum, which appears in a more dorsal region, is responsible for the coordination and planning of movements, for learning motor tasks. The forebrain contains many subdivisions, and the cerebral right and left hemispheres are the most obvious structural anatomical regions. They are mostly characterized by multiple gyri and sulci (see Figure 2.2) that show some reasonably consistent landmarks, which help divide the hemispheres into four lobes: occipital, temporal, parietal, and frontal. Other subdivisions of the forebrain lie deeper in the cerebral hemispheres, such as the basal ganglia (involved in motor and cognitive processes), the hippocampus (involved in memory behaviour), the amygdala (involved emotional behaviour), the olfactory bulbs (involved in processing chemosensory information arising from receptor neurons in the nasal cavity), the thalamus (involved in the relaying for sensory information), and the hypothalamus (involved in the regulation of the body's homeostatic functions) [20]. [see Figure 2.1]

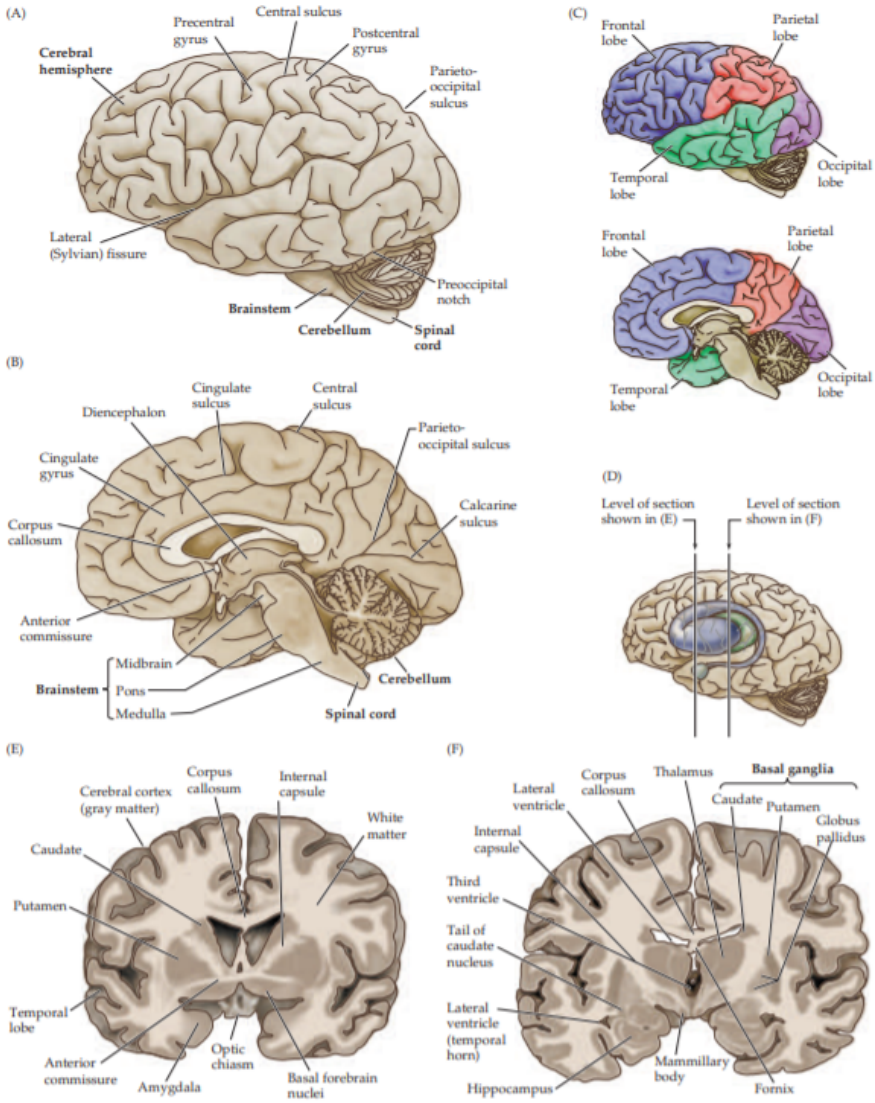


Figure 2.1: Anatomy of the brain.(from [20])

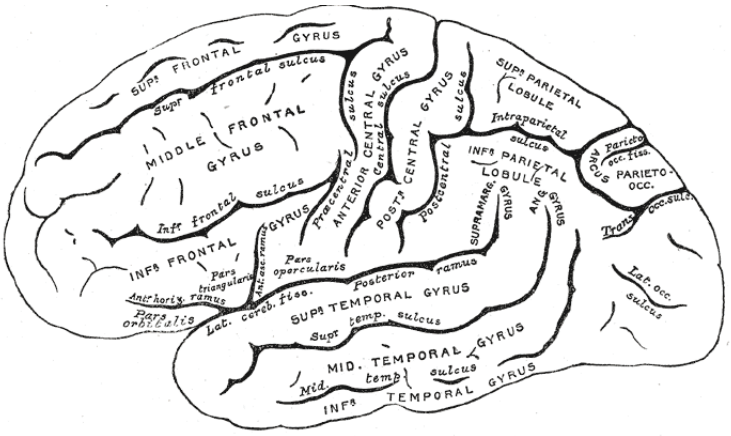


Figure 2.2: Gyri and Sulci of the brain.(from [22])

2.2 Functional Magnetic Resonance Imaging

In medicine, the extension of the physical exam started in 1895, when Wilhelm Röntgen showed the first X-ray image of a skeletal hand. Since then, a series of advances in different fields of science, such as physics, mathematics, biology, and computer science, have progressively transformed the way we look at brain anatomy and function [23].

As science is a continuous process, new inventions are always dependent on previously acquired knowledge. The history of functional magnetic resonance imaging (fMRI) follows that rule: while the mathematical basis of magnetic resonance imaging (MRI) is dated to the 1800s due to the great work of Jean Baptiste Joseph Fourier, and the discovery of the fundamental physics and biological properties of nuclear magnetic resonance started in the 1930s with the advances of Isidor Rabi and other scientists, the invention of MRI solely occurred in the 1970s. Although the contributions of Raymond Damadian and Richard Ernst are not underestimated, this big step was attributed to Paul Lauterbur and Peter Mansfield by rewarding them later with a Nobel Prize in Physiology or Medicine in 2003. However, it is only in 1990, that Seiji Ogawa observed the blood oxygen level-dependent (BOLD) effect, which is the mechanism that underlies fMRI [23].

2.2.1 Principles of Magnetic Resonance

fMRI scans are based on the same atomic physics principles as MRI scans. The fundamental principle of MRI is that atomic spins interact with magnetic fields like small magnets, allowing researchers to measure and manipulate their magnetic state. Considering that the human body has a very substantial water content, the magnetic properties of hydrogen protons are used as a target [24].

When a strong external magnetic field, B_0, is applied, the hydrogen protons as a consequence of having both a magnetic momentum and angular momentum, will precess around the direction of that field in such a way that the frequency of rotation is proportional to B_0. This relation is governed by the Larmor equation [24, 25, 26, 27]: [see Figure 2.3]

$$\omega_0 = \lambda B_0 \tag{2.1}$$

where ω_0 is the Larmor frequency, B_0 is the magnetic field strength, and λ is the gyromagnetic ratio for the proton (ratio between the magnetic moment and the angular momentum) [27].

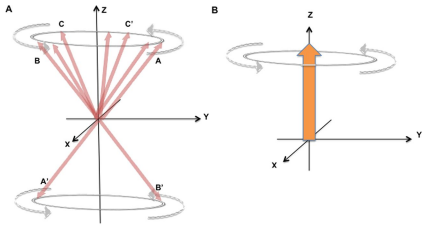


Figure 2.3: Longitudinal magnetisation: When a B_0 field is applied (orientated with the z-axis), hydrogen protons align either with (parallel) or against (antiparallel) the orientation of B_0. The parallel is the preferred state of alignment since it is the one that requires the least energy. Therefore, more protons align parallel (A,B,C,C') with B_0, than antiparallel (A', B'). The magnetic moments of protons precessing in the external magnetic field cancel each other out on the transverse direction of B_0 directions (A, A'; B, B'; C, C'). Regarding the longitudinal direction, since there is an excess of protons aligned parallel to B_0 (C, C'), the result is a sum magnetic field vector oriented with B_0 [from [28]]

This rotation is important because the magnetisation can be rotated away from its alignment along the B_0 axis by applying a radiofrequency (RF) magnetic field, B_1, perpendicular to the external magnetic field, with the Larmor frequency for a short period. This RF magnetic field has the effect of shifting the orientation of the proton's magnetisation from the B_0 field to the transverse plane. This results in a transverse component of the magnetisation, that can be measured [26, 24]. After applying the RF pulse, the nucleus will tend to reestablish its initial configuration. Therefore, the magnetisation will tend to grow back along the longitudinal direction of the B_0 field and decrease along its transverse direction. The decrease of the transverse magnetisation induces a electrical current in a receiving coil, that generates a signal, known as free induction decay(FID). FID is a time-domain signal, which can be studied in the frequency-domain by applying a Fourier transform [25, 27]. [see Figure 2.4]

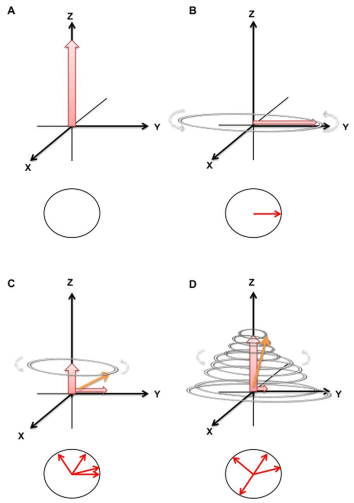


Figure 2.4: Recovery of magnetisation following a 90° RF pulse: (A) Hydrogen protons aligned with B_0 produce a sum vector with longitudinal magnetisation and no transverse magnetisation. (B) When a RF pulse is applied perpendicular to the B_0 field, the longitudinal magnetisation becomes zero and the transverse magnetisation emerges. Immediately following that RF pulse, protons precess in phase in the transverse plane (single vector in the lower circle). (C) After the 90° RF pulse, protons dephase (multiple vectors in the lower circle), the transverse magnetisation decreases and the longitudinal magnetisation starts to recover. (D) During this process, the whole system continues precessing and so the sum vector takes a spiralling motion [from [28]]

The produced MRI signal is made up of the sum of the contributions of every single hydrogen nucleus inside the scanner. Still, it is necessary to determine the location of each contribution to reconstruct the MRI image. This is possible by relating a specific frequency to a specific location, considering that the signal's frequency depends on the field strength. To do so, while acquiring the signal measurements,

magnetic fields that vary with location, called gradient fields, are added to the B_0 field. Gradients can be applied in any orthogonal direction using the three sets of gradient coils, Gx, Gy, and Gz, within the MRI system. As a result, the signals from the protons have a frequency that depends on their location. Thus, the difference in frequencies can be used to distinguish signals from different locations. The combination of the three gradients with multiple measurements allows for the three-dimensional image reconstruction [24, 27]. [see Figure 2.5]

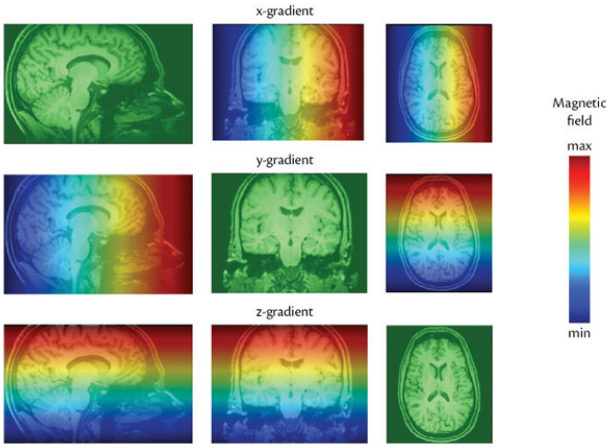


Figure 2.5: Illustration of the three gradient fields: x, y, and z. The colours represent the strength of the magnetic field at each location, which is present throughout the head and makes the signal frequency depend on the spatial location. In these examples, the gradients are applied independently, although, in practice, they are applied in combination. (from [24])

As previously mentioned the recovery of the magnetisation vector leads to a decrease along the transverse direction, and an increase along the longitudinal direction of the external magnetic field. This increment of the longitudinal component of the magnetisation vector emerges from the interaction between the spins and the atomic neighbourhood and can be characterized by a time constant T1, the longitudinal relaxation time [26]. Differently, the decrease of its transverse component magnetisation results in part due to the dephasing of spins caused by variations in the local precessional frequencies and is characterized by a decay time called T2. The rate of dephasing is speeded up by inhomogeneities in B_0, and rather than observing T2, the composite decay factor $T2^*$ is observed [26, 25]. Therefore, it is possible to use the fact that T1 and T2 constants are different along the human body to generate contrast in MRI images.

In a traditional spin-echo sequence, two RF pulses are applied, with an angle of 90º and 180º, respectively, and the time between the two pulses is half of the echo time (TE). The TE is the time between the 90º pulse and the signal. This process can be repeated several times, and the repetition time (TR) defines the time between the 90º RF pulses [25]. In practice, for TE $\simeq$ T2 and long TR (relative to T1), the image is weighted by T2, and the contrast between tissues with different T2 is enhanced. For TR $\leq$ T1, and short TE (relative to T1), the image is weighted by T1, and the contrast between tissues with different T1 is intensified. For long TR and short TE, a proton density-weighted image is generated [26]. [See Figure. 2.6]

Nevertheless, is it important to mention that other sequences can also be used to obtain a MRI image, such as inversion recovery (IR), magnetisation prepared gradient echo imaging (MPRAGE), among others [29]. In the case of fMRI, echo-planar imaging (EPI) is the most commonly used sequence [30].

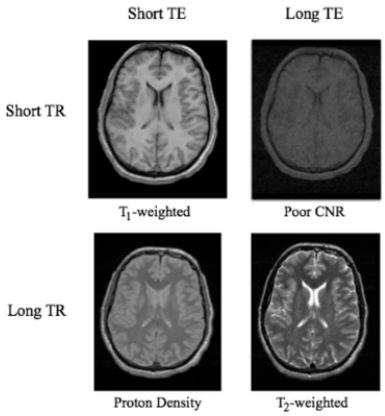


Figure 2.6: The choice of TE and TR determines the type of contrast. Notice that for T1-weighted images, cerebrospinal fluid is dark, and white matter is brighter than grey matter. For T2-weighted images, the cerebrospinal fluid appears bright, and the white matter appears darker than grey matter. (from [24])

2.2.2 BOLD effect

As previously mentioned, the basic principles of fMRI are based on the underlying mechanisms of MRI, and the distinction between the two medical imaging techniques lay on the fact that fMRI is BOLD effect-based, allowing the generation of images of metabolic activity, rather than the production of very detailed anatomical images, as is in the case with MRI. fMRI is sensitive to hemodynamics alterations, particularly to oxygenated hemoglobin (oxy-Hb) and deoxygenated hemoglobin (deoxy-Hb) states, since oxygen (O_2) presence affects hemoglobin magnetic properties, which results in local fMRI measurable magnetic field variations. While oxy-Hb is diamagnetic and intensifies the magnetic signal, deoxy-Hb has 4 unpaired electrons and is highly paramagnetic, which results in local magnetic field distortions, leading to a reduction in signal intensity [24, 31].

These hemodynamic changes occur due to alterations in neuronal firing. All neural signalling processes require energy in the form of adenosine triphosphate (ATP). When a particular brain region is activated, more energy is required, which increases the cerebral metabolic rate of oxygen ($CMRO_2$). While the local stores of O_2 are consumed, the capillaries vasodilate, and the increased blood flow acts to restore the local levels of O_2. For reasons that are not fully understood, more O_2 is delivered than needed to offset the increase in $CMRO_2$. Consequently, there is an increase of deoxy-Hb, and a decrease of oxy-Hb, fast followed by a physiological response that increases oxy-Hb and decreases deoxy-Hb [31]. This is more complex than expected since neuronal activity tends to increase O_2 usage, blood volume, and blood flow, which individually have different impacts on the amount and concentration of deoxy-Hb. O_2 consumption increases the concentration of deoxy-Hb, which leads to a weaker magnetic signal. The increment in blood volume, without changing the concentration of deoxy-Hb, promotes the same effect since it raises the total amount of paramagnetic material. Differently, increased blood flow will increase the washout of the deoxy-Hb present, which reduces the amount of paramagnetic material and increases signal intensity. Taking the three effects together, the increase of local magnetic field signal resulted from the increased blood flow, which overcomes the magnetic effects of O_2 utilization and blood volume increment. Overall, the BOLD effect is a slow mechanism that increases signal intensity on activated brain regions that appear bright in fMRI images [32].

2.2.3 Resting-state fMRI

The human brain consumes about 20% of the total energy produced by our body, even when not performing a task. Taking this into account, while fMRI is better known as a task-dependent exam (task-fMRI) in which the patient is asked to perform specific tasks during the MRI scan, there has been some interest in what the human brain is doing when it appears to be doing nothing, leading to the development of rs-fMRI [30]. Actually, a substantial interest emerged after a Biswal *et al* publication [33], where they demonstrated that the same brain regions activated during a particular task, also displayed coherent low-frequency spontaneous fMRI signal at resting-state [34].

While task-fMRI allows the association between brain localization and functionality, rs-fMRI studies are more concerned about brain functional connectivity. Functional connectivity in fMRI refers to the similarity of low-frequency fluctuations in the BOLD signal between separated brain regions, and it can be used to assess the brain's dynamics. To avoid biased results, it is more often to study functional connectivity at a resting state when no specific cognitive demands are required. Additionally, it is important to differentiate functional connectivity from effective and anatomical connectivity. Functional connectivity describes the relationship between two regions, but it does not specify the direction of that relationship (effective connectivity), or imply that the two brain regions are physically connected (anatomical connectivity) [30].

FBN consist of several brain regions functionally connected that show similar BOLD signals at resting-state. They are relevant for investigating the brain as a complex network of elements in order to assess brain dynamics in a more meaningful way that can help distinguish pathological from control situations. FBN are useful to understand the organization of cortical functions. They can also be used as biomarkers for neurological and psychiatric disorders, aiding the clinical diagnosis [35, 36].

2.2.3.1 Identified Functional Brain Networks

Several FBN have been consistently identified and named in the literature [37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 33, 41, 46, 48, 49, 50, 51]. In the context of this dissertation, the fourteen FBN that were identified in the study from Shirer *et al* [44] will be used as the main reference in terms of nomenclature, location and function [See Figure 2.7].

Nonetheless, the study from Yeo *et al* [51] should also be cited, as the seven FBN that they identified are widely cited in the literature. Thus, it is important to match the FBN from both studies [See Figure 2.8].

The anterior salience network (ASN) from Shirer *et al* [44] totally matches with parts of the ventral attention network from Yeo *et al* [51]. The auditory network (AN), and the sensorimotor network (SN) from Shirer *et al* [44] completely match with parts of the somatomotor network from Yeo *et al* [51]. The dorsal default mode network (DDMN), and the language network (LN) from Shirer *et al* [44] completely match with parts of the default mode network from Yeo *et al* [51]. The ventral default mode network (VDMN) from Shirer *et al* [44] match with both default mode and dorsal attention networks from Yeo *et al* [51]. The higher visual network (HVN), and primary visual network (PVN) from Shirer *et al* [44] totally match with parts of the visual network from Yeo *et al* [51]. The visuospatial network (VN) from Shirer *et al* [44] completely matches with several parts of the dorsal attention network from Yeo *et al* [51]. The left executive control network (LECN), and the right executive control network (RECN) from Shirer *et al* [44] match with the frontoparietal network from Yeo *et al* [51], and also present some regions that overlap with the default mode network from Yeo *et al* [51]. The posterior salience network (PSN) from Shirer *et al* [44] matches with few parts of the ventral attention network, and may also be represented by

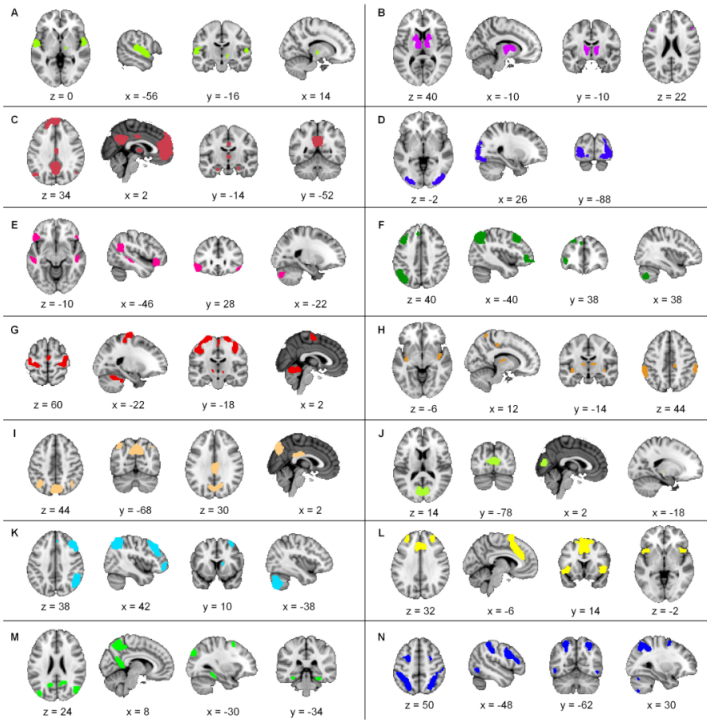


Figure 2.7: FBN from Shirer *et al* [44]. [A]- auditory network (AN); [B]- basal ganglia network (BGN); [C]- dorsal default mode network (DDMN), which is placed on the posterior cingulate cortex/medial prefrontal cortex; [D]- higher visual network (HVN); [E]- language network (LN); [F]- left executive control network (LECN), which is placed on the left dorsolateral prefrontal cortex/left parietal lobe; [G]- sensorimotor network (SN); [H]- posterior salience network (PSN), located on the posterior insula; [I]- precuneus network (PN); [J]- primary visual network (PVN); [K]- right executive control network (RECN), which is placed on the right dorsolateral prefrontal cortex/right parietal lobe; [L]- anterior salience network (ASN), located on the insula/dorsal anterior cingulate cortex; [M]- ventral default mode network (VDMN), which is placed on the retrosplenial cortex/medial temporal lobe; [N]- visuospatial network (VN), which involves the intraparietal sulcus/frontal eye fields. Additionally information about the anatomy of FBN can be found in Figure A.3

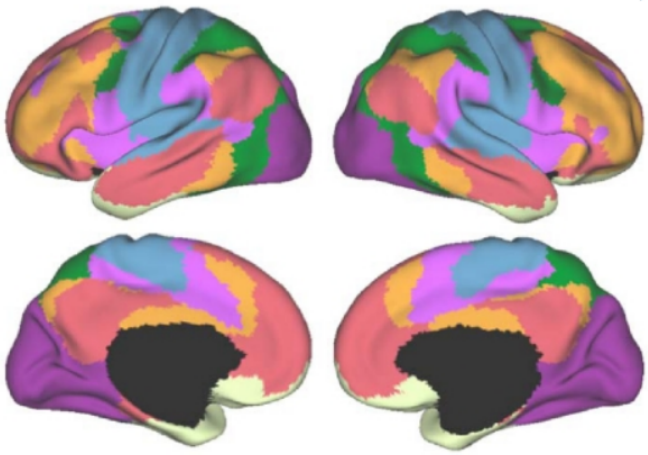


Figure 2.8: FBN from Yeo *et al* [51]. purple- visual network; blue- somatomotor network; green- dorsal attention network; violet- ventral attention network; cream- limbic network; orange- frontoparietal network; red- default mode network

parts of the frontoparietal network from Yeo *et al* [51]. The precuneus network (PN) from Shirer *et al* [44] matches with both default mode and frontoparietal networks from Yeo *et al* [51]. The basal ganglia network (BGN) from Shirer *et al* [44] is not present in the study from Yeo *et al* [51]. On the contrary, the limbic network from Yeo *et al* [51] is not present in the study form Shirer *et al* [44]. Additionally, it must be noticed that there are different FBN from Shirer *et al* [44] that may represent only one FBN

from Yeo *et al* [51]. Besides, some of those different FBN from Shirer *et al* [44] are present in different FBN from Yeo *et al* [51]. Nevertheless, in general, there are more parts of FBN from Yeo *et al* [51] that do not have a match of FBN from Shirer *et al* [44], than the contrary.

Afterwards, it is possible to associate those FBN with their anatomy and function. The ASN encompasses the dorsal anterior cingulate cortex and the anterior insula. These regions of the FBN are activated in the presence of several heterogeneous stimuli, and this FBN plays a role in the linking between cognition and emotion/interoception processes [50]. The dorsal anterior cingulate cortex structure is associated with executive control, learning, adjustment, economic choice, and self-control [52]. The anterior insula mediates interoceptive attention [53]. The AN is located in the primary auditory cortex and is related to audition, music and speech processes. The BGN is located in the basal ganglia and thalamus, and it is linked with motor, pain, and somatosensory processes, and a wide range of mental processes, including reward tasks, classical conditioning, and anxiety [50]. The basal ganglia refer to the caudate, putamen, globus pallidus, nucleus accumbens, substantia nigra and subthalamic nucleus. It is responsible for controlling voluntary movements, cognition, and emotional processes. The thalamus acts as a relay station that receives, processes, and sends sensory information to an associated cortical area [54]. The DDMN corresponds to the posterior cingulate cortex, medial prefrontal cortex, angular gyrus, and hippocampus. All regions of this FBN overlap the well known, default mode network, which is possibly the most widely studied FBN in rs-fMRI since it is activated when no task is being done. This FBN is related to internal mental-state processes, such as self-referential processing, interoception, autobiographical memory retrieval, or imagining the future. [43, 55]. The VDMN is located on the medial temporal lobe, retrosplenial cortex, posterior cingulate cortex, and regions of the cerebellum. Several regions of this FBN are also part of the default mode network. Regarding function, it is interesting to notice that while the VDMN is related to the vividness but not the valence of imagined events, the DDMN is related to the valence but not the vividness of those imagined events [56].

The HVN is located on the secondary visual cortex, which includes the left and right occipital gyri, and is related to higher-level visual processing, such as reading tasks. The PVN is located on the primary visual cortex, at the calcarine sulcus, and is related to simple visual stimuli, such as flashlights. The LN encompasses the inferior frontal gyrus (Broca's area), left middle temporal gyrus, angular gyrus, superior temporal gyrus, and supramarginal gyrus (Wernicke's Area), and is associated with speech and language processes. While Broca's area is mostly associated with the production of language, Wernicke's area is mostly related to the understanding of spoken language [20]. The LECN and RECN are located on left/right frontoparietal regions, including the left/right dorsolateral prefrontal cortex, posterior parietal, and cerebellum regions, respectively. Those FBN are essential to coordinate behaviour efficiently [57]. They correspond to several cognition paradigms, that require externally directed attention, such as working memory, relational integration, response inhibition, and task-set switching. Regions of those FBN have also been related to several creative thought processes. While the RECN appears to be involved in multiple cognitive processes, such as reasoning, attention, inhibition, and memory, the LECN is strongly related to working memory, and language tasks [58, 50]. The PSN is mostly located on the posterior insula. This region represents interoceptive information about the body's physiological status and appears to play a role in the recognition, intensity encoding, localization, learning, and memory of painful events [59]. Although the PN is mostly located on the precuneus, it also involves the midcingulate cortex and posterior cingulate cortex. The precuneus is involved in complex functions, such as integration of information relating to environment perception, mental imagery strategies, episodic memory retrieval, affective responses to pain, and sense of self [60, 61]. The SN includes somatosensory (postcentral gyrus) and motor (precentral gyrus) regions and extends to the supplementary motor area, and is related

to motor tasks [62]. The VN is located on the intraparietal sulcus and frontal eye fields and is activated when attention is oriented in space, such as when planning eye movements, grasping movements, or defensive head movements [63, 64].

Further information about the anatomy of those FBN can be found in Figure A.3.

After identifying several FBN, it is important to recognize that the generation of those FBN is also dependent on the quality of rs-fMRI data acquisition and preprocessing.

2.3 fMRI Data Analysis

2.3.1 Data Acquisition and Preprocessing

The quality of fMRI studies is dependent on data acquisition and data preprocessing, since both influence the final results. The most commonly used sequences for acquiring rs-fMRI data are based on EPI. EPI is a fast fMRI sequence that allows the acquisition of single images in only a fraction of a second. It achieves that speed by obtaining all spatial-encoding information after a single RF pulse. Besides fast imaging acquisition, EPI sequence is also less sensitive to motion artefact than conventional fMRI sequences and has the ability to image rapid physiologic processes of the human body, such as blood flow [65, 66].

Besides specifying the fMRI sequence, several scan parameters need to be defined for acquiring fMRI data.TR, and voxel size are important scan parameters that need to be set. While TR controls the temporal resolution, voxel size determines the spatial resolution of the data. By reducing the TR, temporal resolution is improved, since it increases the total number of time points. Small voxels improve spatial resolution, but the amount of signal in each voxel is reduced, which decreases the signal-to-noise ratio (SNR). In addition, there is also a trade-off between spatial and temporal resolution. Reducing the slice thickness (one of the dimensions of a voxel) implies that more slices are needed to cover the entire brain, and consequently TR needs to be higher, in order to cover the whole brain, which results in a worse temporal resolution. As a result, voxel sizes typically range from 2 mm to 4 mm, and TR usually ranges between 1 s to 3 s [30].

Inhomogeneities in the magnetic field, which cause signal dropout or distortion, are another aspect of data acquisition to be aware of. The use of fieldmaps, which must be obtained during data acquisition, is very helpful for the step of distortion correction. However, most fMRI public databases do not contain those fieldmaps.

Regarding scan duration, Van Dijk and colleagues [67] showed that 5 min were sufficient to stabilize BOLD data, and as consequence most current studies acquire 5–7 min of rs-fMRI data [67, 68]. Even though, increasing the data scan time to 10-15 min should be considered, since the increase in data points results in less noisy functional connectivity estimates [30]. A study from Birn *et al* [68] showed that the reliability of measuring individual differences in the strength of connectivity continues to increase for scan durations longer than 6 min [68].

Regarding TR, Demetriou *et al* [69] demonstrated that even though rs-fMRI can substantially benefit from higher temporal resolution, substantial reductions in TR do not produce strong benefits in statistical reliability in a straightforward manner. However, there is a highly consistent effect of increasing TR on a dual regression approach, showing a particular benefit of higher temporal resolutions (eg. TR$<$1s) when using ICA-based approaches [69]. The advantages of short TR enabled by multi-slice imaging were investigated by Jahanian *et al* [70]. They concluded that whole-brain rs-fMRI studies may benefit from increased temporal resolution. However, they did point out that most rs-fMRI studies collect data

for 5–12 min with a long TR (TR>2s) and have been shown to be well replicated at the group-level in NPDs populations [70]. Recently, a study from Huotari *et al* [71] demonstrated that the conventional resting-state metrics, such as ICA, were not markedly affected by different TRs (0.1–3s). In their study, ICA detected FBN in all TR settings, but there was some variability between the results. In fact, short TRs appear to benefit rs-fMRI measures, particularly ICA [71].

The scanner that is used also influences the results of fMRI studies. Zhao *et al* [72] showed that inter-scanner reliability is not as good as the intra-scanner reliability. Nonetheless, multi-site studies, which are characterized by multiple subjects and multiple scanners, are gaining great attention, as the combination of different databases provides more data, and increases the statistical confidence of the neuroimaging studies [73].

Another important decision when performing fMRI studies is about the eyes condition during scan. In rs-fMRI studies, eyes closed (EC), eyes open (EO), and eyes fixated on a target (EO-F) are the most common eyes condition. Yan *et al* [74] suggested that cautious should be taken into account when choosing the type of eyes condition during scan. Still, the difference between EO and EO-F is not as significant as the difference between EO and EC conditions [67, 74, 72].

Several parameters influence fMRI data. Actually, one major drawbacks of rs-fMRI is that there is still no standard defined acquisition method for rs-fMRI studies, regardless of whether they are conducted on 1.5T, 3T, or 7T MRI scanners. As previously mentioned, the duration of the resting-state examination, which typically vary between a few minutes up to 12 min, influences the results. In addition, TR used on rs-fMRI studies is typically between 1s to 3s. As periodic signals such as heart rate or respiration rate may affect outcomes differently, different TRs give varied results. The eyes condition is another variable that varies between different studies and also influences the results. The MRI scanner, on the other hand, is always dependent on where the study is conducted, and it is not as easy to control that parameter as the other manageable parameters. [75].

Overall, the selection of scan parameters should be optimized to maximise the amount of data, reduce noise caused by the influence of scan parameters, and maintain the heterogeneity that leads to robust studies.

The acquired rs-fMRI data must be preprocessed by applying conventional and specific methods for rs-fMRI. Firstly, some conventional fMRI data preprocessing methods must be used. The scientific community considers that motion correction and distortion, temporal filtering, and registration are always necessary preprocessing steps. Still, slice-time correction, spatial normalisation, and spatial smoothing may also improve data quality. It is also important to mention that in the registration stage, the additional use of an individual brain extracted T1-weighted structural image for registering the functional image to the structural image, and register the structural image to a standard image, like Montreal Neurosciences Institute 152 image (MNI152), is quite relevant.

Secondly, some additional preprocessing specific steps for rs-fMRI may be useful, such as nuisance regression or an ICA-based clean-up method, which can remove noise from physiological and movement sources [30].

2.3.2 Functional Connectivity Analysis Methods

Following the preprocessing stage, a variety of methods can be used to measure functional connectivity. ICA and seed-based correlation analysis (SCA) are the primarily chosen methods to study functional connectivity [76, 77, 78]. Both can be applied to one subject or can be used on a group-level approach. For group-level analysis, both show excellent reliability at reproducing previously founded

FBN [37]. Although ICA and SCA results are similar, they are conceptually different [77].

2.3.2.1 Seed-based Correlation Analysis

In this approach, a brain region-of-interest (ROI) is defined, and it intends to create a map showing the functional connectivity strength between each voxel in the brain and the seed region, ROI. After defining the seed region, the BOLD time series is extracted or calculated (if the ROI contains more than one voxel). Therefore, a correlation coefficient between the time series of each voxel in the brain and the time series of the seed region is calculated, and the result is entered into the connectivity map at the location of each respective voxel. To allow group-level analysis, Fisher's r-to-Z transformation is often applied to each subject correlation map [30]. Although its simplicity enables the direct study of voxel regions to which the seed region is connected, SCA requires prior knowledge of ROIs, which is not always straightforward [78, 79, 80]. Besides that, SCA does not consider the connectivity patterns on a whole-brain scale, and the results are limited and biased by the defined seed region [78, 30]. In fact, with SCA, the connectivity between two co-activating brain regions is computed to be high, regardless of whether the two regions belong or not to the same intrinsically connected network. Hence the activation of one FBN cannot be distinguished from a co-activation of multiple FBN [77].

2.3.2.2 Independent Component Analysis

ICA can consistently estimate multiple FBN that make up the resting-state signal [77]. This fully data-driven method efficiently examines general connectivity patterns from the whole brain by estimating maximally independent components using independent statistical measures [81, 78]. This blind source separation algorithm transforms brain signals into their latent sources by selecting which components explain most of the variance within the dataset [35]. By assuming that the rs-fMRI data is composed of a mixture of unobserved but separable underlying components, ICA tries to decompose this complex data into simpler and readable components. In fact, a time course is associated with each voxel of the brain, and ICA tries to separate this spatiotemporal brain data into a spatial and a temporal portion. Then, it assumes that the generated components will either be spatial or temporal independents. The most common approach is to look for spatial independent components that are maximally non-Gaussian, which consequently allow the identification, for each component, of a spatial map and a timecourse [30].

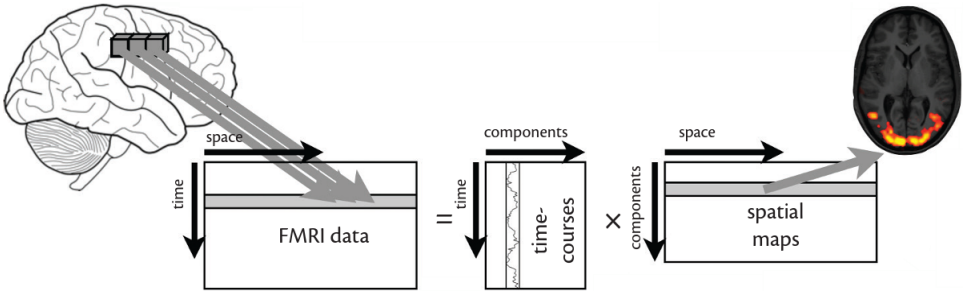


Figure 2.9: Matrix representation of ICA method. On the left, there is a representation of the BOLD data, where each row represents data from a 3D volume (all voxels) at one time point and each column represents data from all time points at one voxel, per time point, that represents the entire 3D brain. After running ICA, this data is unmixed into a set of components that are each described by a timecourse and a spatial map. One timecourse is extracted for each component, and for each of the timecourses, there is a spatial map. Therefore, the number of rows in the timecourse matrix (number of time points) is equal to the number of rows of the input matrix, and the number of columns in the spatial map matrix (number of voxels) is equal to the number of columns of the input matrix. As a consequence, the number of columns in the timecourses matrix is the same as the number of rows in the spatial map matrix and represents the number of ICA components [30]

Besides applying ICA to a single subject, it is also common to perform group-ICA, to identify FBN for a larger population. Group-ICA is typically performed by spatially registering all subjects to a common standard space and then temporally concatenating the fMRI data from all subjects. This means that, for each voxel, the data from subject 2 is pasted after the data from subject 1 (and so forth). The result is a two-dimensional matrix where the columns contain the long, concatenated timeseries of all subjects, and the rows contain information from all voxels in the brain. After performing group-ICA, the output still contains a set of spatial maps (one per component, representing the group map), and a set of timeseries (a very long concatenated timeseries of all subjects, for each component) [30].

As highlighted before, ICA is also useful for cleaning rs-fMRI data, when applied to each subject individually. Even though some generated components explain the brain signal and are useful for functional connectivity analysis, other noise components must be removed. Each independent component is described by a spatial map, a time series, and a power spectrum. By analysing those, it is possible to perform hand label classification and remove all noise components, and clean rs-fMRI data [82]. However, this method may be time-consuming for a lot of data. Some researchers created automatic tools that label and remove noise components, such as ICA-AROMA [83] and FIX [84, 85], which can be used within the FSL software.

2.3.3 Functional Connectivity Metrics

Functional connectivity can be evaluated with both intra-network and inter-network metrics. Those assess connectivity within and between several FBN, respectively [86]. Graph-based connectivity analyses allow the investigation of connections, edges, between the defined functional components, nodes [30]. Functional connectivity is measured by evaluating temporal synchrony and correlations between distinct activated nodes [9]. Although those nodes can be the identified FBN, it is also common to break them down into smaller nodes. The reason for that is that studies of brain function suggest that subregions in these FBN do act distinctly. However, the optimal characteristics of brain parcellation are still unknown [30].

For rs-fMRI studies, node definition can be divided into data driven-based and functional atlas-based approaches. ICA, which is a common data-driven method for defining the nodes, was already investigated in the previous section. Naturally, the nodes would be the set of components obtained after performing group-ICA. This data-driven approach has the main advantage of providing the best representation of the structured components that are present in the studied dataset. On the contrary, the functional atlas approach uses templates that can be acquired from previous studies available in the literature [51, 87, 44]. This approach has the advantage of ensuring that the defined nodes are consistent with the literature, which allows easy comparison between studies [30].

After defining the nodes, a common method is to extract the timeseries that represent each node. It can be performed by either averaging the BOLD data over all voxels that make up the node or by applying dual regression. Dual regression is usually preferable since it reduces the impact of noise. This method uses a group template map, which can either be a functional atlas or a group-ICA template and applies two subsequent regression analyses on each subject's preprocessed and cleaned rs-fMRI data to generate spatial and temporal subject-specific maps. Then, in order to define the edges, connectivity metrics between nodes must be calculated. Correlation metrics, such as Pearson's correlation coefficient which is sensitive to direct and indirect functional connections in the brain, are commonly used and allow the estimation of the similarity between two BOLD timeseries. This process leads to a generation of a functional network connectivity matrix, that estimates correlations between each predefined node. This

matrix, which depicts the functional connectivity between pairs of brain regions, can be visualized as a graph, which is a diagram representation of nodes and edges [30].

Graph theory metrics are used to extract higher-level measures that describe aspects of network functionality. Generally, the obtained correlation matrix is thresholded and binarized, to only retain the strongest connections between nodes. Afterwards, some graph metrics can be calculated to study the network properties. The network degree is one of the most common metrics. It defines the number of nodes that are connected to a particular node. If the degree of a node is larger than the average degree in the graph, that node is named a hub [30].

After preparing the rs-fMRI data of all subjects, it is common to perform statistical analysis, in order to evaluate differences among individuals or groups of individuals. However, due to the complexity of neuroimaging data, traditional analytical methods are neither computationally effective nor clinically relevant. The creation of DL algorithms for extracting relevant features from complex data sets suggests that they could be used for the diagnosis of NPDs [88].

2.4 Deep Learning

The early success of AI took place on concrete and formal tasks, where knowledge about the world was not mandatory. However, AI systems based on hard-code knowledge faced several difficulties in solving abstract problems. A different vision from scientists emerged, standing that those AI systems needed to learn from experience, giving them the ability to acquire their knowledge by extracting features from raw data. This qualification is known as ML. However, its performance is highly influenced by data representation. A feature consists of each piece of information that is included in data representation. ML models outcomes are influenced by those features, and sometimes it is not straightforward to know which ones should be extracted from data. A solution to this problem is representation learning, which consists of the application of ML to discover a good set of features that represent the data. However, a major source of difficulty in real-world applications is that each feature is influenced by several factors of variance. As an example, the image of a red car might be very close to black at night, and its shape silhouette depends on the viewing angle. In fact, it is challenging to extract such high-level and abstract features from raw data, and it becomes as difficult to obtain a good representation of data as solving the original problem. After all, DL can solve this major problem by introducing representations that are expressed by simpler representations. It creates complex concepts from simpler ones by employing the hierarchy of concepts. DL can represent the concept of an image of a person by combining different edges, like corners and contours [89].

DL is a type of ML that allows computers to learn and improve from experience and data [89, 90]. In the big data era, DL is receiving much attention. It describes a family of learning algorithms used to learn complex prediction models, usually multi-layer neural networks (NN)s with many hidden units [91].

2.4.1 Neural Networks

2.4.1.1 Basics of Neural Networks

NNs are the basis of DL. A typical NN has a number of layers (depth) and several neurons (width). Actually, a deep NN usually consists of an input layer, an output layer, and at least two hidden layers [92, 91]. To understand the behaviour of each neuron, the basic unit of NN, it is valuable to learn the basics of the perceptron model, since it consists of a NN with one single layer and one neuron [see

Figure 2.10]. Each neuron computes the weighted (w) sum of inputs (x) added with a term bias (b), and the predicted output of the neuron ($\hat{y}$) is calculated by feeding the computed value into an activation function (ϕ):

$$\hat{y} = \phi(z) = \phi(w^T x + b), \tag{2.2}$$

for the example presented in Figure 2.10, w^T is the weights transpose matrix with shape (1,3), x is the input matrix with shape (3,1), b is the bias matrix with shape (1,1), $\hat{y}$ is the predicted output matrix with shape (1,1), and ϕ is the activation function of the neuron.

The learning process relies on varying the weights values of the active inputs until the output of the NN corresponds to the expected value [91].

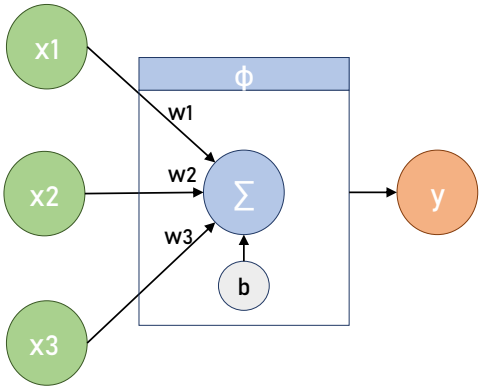


Figure 2.10: Perceptron model, which consists of a simple NN with one neuron. Equation 2.2 can be decomposed on: $\hat{y} = \phi(x1w1 + x2w2 + x3w3 + b)$, where x1, x2, and x3 are the inputs of the model (elements of the x matrix), w1, w2, and w3 represent the weights (elements of the w matrix), b is the bias term (element of the b matrix), ϕ is the activation function, and $\hat{y}$ is the predicted output of the model (the element of the $\hat{y}$ matrix).

For deep NN, each neuron works similarly to the single neuron of the perceptron model. Figure 2.11 presents the architecture of a deep NN with 2 hidden layers, which can be mathematically represented as:

$$\hat{y} = \phi^{[2]}(w^{[2]T}\phi^{[1]}(w^{[1]T}x + b^{[1]}) + b^{[2]}), \tag{2.3}$$

where x is the input matrix with shape (2,1), $w^{[1]T}$ is the weights transpose matrix of the first hidden layer with shape (3,2), $b^{[1]}$ is the bias matrix of the first hidden layer with shape (3,1), $\phi^{[1]}$ is the activation function of the neurons of the first hidden layer, $w^{[2]T}$ is the weights transpose matrix of the second hidden layer with shape (1,3), $b^{[1]}$ is the bias matrix of the second hidden layer with shape (1,1), $\phi^{[2]}$ is the activation function of the neuron of the second hidden layer, and $\hat{y}$ is the predicted output matrix with shape (1,1), which is equal to the activated value of the second hidden layer.

The activation function is responsible for performing non-linear transformations of $z = w^T x + b$. NN can learn powerful non-linear transformations, and by transforming data non-linearly into a new space, a classification problem that was not solvable by a linear classifier can become separable. Several activation functions can be used. The sigmoid activation function is a natural choice for the output layer of the NN for binary classification problems. The tangent activation function has the advantage for hidden layers of centring the data. However, both sigmoid and tangent activation functions have the derivative close to zero when the value of z is very large or very small, which slows gradient descent. As a consequence rectified linear unit (ReLu) is the most common choice for the activation function of hidden layers. A disadvantage of ReLu is that for negative values of z, the derivative is close to 0. Although

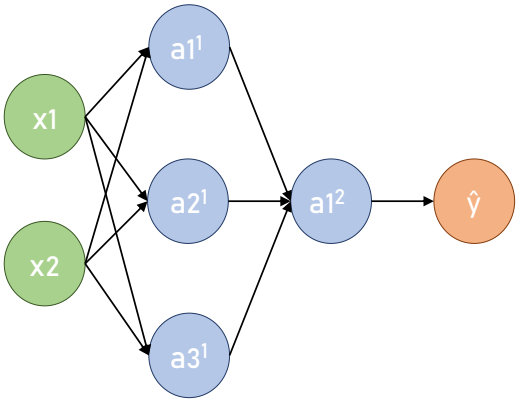


Figure 2.11: Model of a deep NN with 2 hidden layers.

in practice this works fine, another version of ReLu which is the leaky rectified linear unit (leaky ReLu) solves that problem and usually works better. A linear activation function can also be used, when we want $\hat{y} = z$, which is commonly applied for the output layer of AEs [89].

2.4.1.2 Training Neural Networks

Training a NN can be understood as a two-step process. In the forward propagation step, the network is activated by computing the inputs of the first layer, which then go through the hidden layers along with the weighted connections until it activates the final layer and generates the prediction results. A cost function, such as mean squared error, is defined to measure the performance of the algorithm by comparing the predicted outputs to the real values. In the backward propagation step, the weights of connections are tuned by minimising the cost function. This optimization is typically gradient-based. In order to find the direction in which the cost function decreases the fastest, the gradient of the cost function is computed with respect to the weights and biases of each neuron. When the entire training set is passed forward and backwards through the NN once, which refers to 1 epoch, the model improves the predictions. The number of epochs is traditionally large, to allow the learning algorithm to minimise the cost function. However, optimization is not a straightforward process, and the selection of hyperparameters is really important. Learning rate, which controls the size of the step the parameters take in the direction of the gradient, is one of the most difficult hyperparameters to set because it has a significant impact on model performance. To overcome this issue of the most used optimization algorithm, stochastic gradient descent, some other methods have emerged [93, 89]. Adam optimizer [94] has been considered one of the most powerful optimization algorithms. It is an adaptive learning rate algorithm that combines properties of RMSprop [95] and AdaGrad [96] algorithms. The method computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients. Adam is relatively easy to configure and is computationally efficient. The optimization occurs by gradually updating the parameters θ (weights and biases) of the cost function J in the direction of the gradient until the cost function is minimised [89, 94]:

$$\theta_t = \theta_{t-1} + \alpha * \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}, \tag{2.4}$$

where θ are the parameters of the cost function, α is the stepsize (or learning rate), ε is a small constant for numerical stability, $\hat{m}_t$ is the bias-corrected first moment estimate $\hat{m}_t = m_t/(1-{\beta_1}^t)$, which

is calculated with the biased first moment estimate $m_t = \beta_1 * m_{t-1} + (1-\beta_1) * g_t$, and $\hat{v}_t$ is the bias-corrected second raw moment estimate $\hat{v}_t = v_t/(1-\beta_2{}^t)$, which is calculated with the biased second raw moment estimate $v_t = \beta_2 * v_{t-1} + (1-\beta_2) * g_t{}^2$. Note that β_1, and β_2 are exponential decay rates for the moment estimates, and g_t is the gradient of the cost function $J(\theta)$ [94]

The described process is then repeated for each example of the input data and the weights and biases values are gradually updated until good predictions are obtained.

The initialization of the weights and biases is also important. Bias should be initialized to zero or small positive values, and the weights may be initialized to small random values. Nevertheless, more refined approaches to weights initialization have been developed, which are commonly dependent on the kind of activation function that is being used. While for tangent activation function the Xavier initialization [97] seems to be more appropriated, for ReLu activation function the He initialization [98] is commonly used [89].

Overall, the goal of the learning process is to generate a model that performs well not just on the training data but also on new inputs (test set). This ability to perform well on new data is called generalization. The central challenge in DL is to reduce both the training error, and the test error, which represents the computed error measure of the training set, and of the test set, respectively. Underfitting occurs when the training error is not sufficiently low. Overfitting is present when the gap between the training error and the test error is too large. Overfitting can be reduced by increasing the data amount, whereas underfitting can be tackled by increasing the depth of the NN. In addition, several strategies, known as regularization techniques, are explicitly designed to deal with overfitting, by reducing the test error, which usually comes with the cost of increasing the training error. Most classical regularization approaches are based on adding a penalty parameter to the cost function, which limits the learning capacity of the model:

$$\overline{J} = J + \alpha\Omega, \tag{2.5}$$

where α is a hyperparameter that weights the relative contribution of the norm penalty term, Ω, J is the cost function, and $\overline{J}$ is the regularized cost function.

L1 and L2 regularization techniques are both based on that approach, and the key difference between them is the penalty parameter. While L2 regularization, which is the most common form of weight decay, adds the squared value of the weights $\Omega = \frac{1}{2}||w||^2$ to the cost function, L1 regularization penalty term is the absolute value of the weights $\Omega = ||w||$. For both approaches, the hyperparameter α usually ranges between 0 and 0.1 [89].

Dropout is another regularization method. It consists of randomly drop neurons along with their connections from the NN during training, which prevents the NN from depending too much on specific neurons. This means that during each epoch different neurons are eliminated from the NN. The dropout effect can be controllable by the parameter p, which stands for the probability of retaining a neural unit. When p = 1, no dropout is applied. The typical values of p are in the range of 0.5 to 0.8.

2.4.2 Normative Model

Advances in medical technologies led to an increase in physiological data, including medical images. DL-based algorithms applied to that data are considered promising for developing better diagnostic tools [93]. To improve the diagnosis of NPDs, researchers use to apply the common dichotomous classification approach to their studies. However, the normative model approach seems to be more adequate to face the continuous spectrum of NPDs.

"Normative model is an emerging method for quantifying and describing how individuals deviate

from the expected pattern learned from a population" [99]. The expected pattern is commonly learned from a population composed of healthy subjects since it can then be applied to all kinds of disorders. The generation of a normative model encompasses two steps. First, statistical models are estimated to characterize the typical brain data that explain the normative population. After that, the normative model is applied to patients to quantify their deviation from the expected learned pattern. Those variations can be analysed to compare different individuals and different pathological groups [99]. Several statistical models can be used to generate the normative model, being Gaussian process regression the most popular [100]. Nevertheless, particular attention to DL models, such as autoencoders (AEs), should be taken into account. The authors from [101] and [99] used AEs on their studies and were able to detect differences between groups of patients and healthy subjects. The underlying concept is that the AE learns to copy the input data to its output. The AE learns to encode the healthy patterns from the input data to the latent representation (which contains few neurons), and then, only using the information presented in the latent representation, it tries to reconstruct the input data as close as possible to the original inputted data. In a test stage, the AE would fail to reconstruct data from patients, and the difference between the original input data and the reconstructed output could provide a measure of alteration in each subject. [101]. The advantage of using DL models is that it enables the model to learn multiple levels of latent representations from data since it allows the extraction of deeply abstract features. Considering that, a normative model based on DL algorithms is able to learn deeply representations of the normative population, and identify the most important characteristics of the healthy brain [102, 99].

Besides AEs, other DL algorithms can also be used to create a normative model. The generative adversarial networks (GANs) are capable of learning patterns from inputted data [103], making them a viable option for generating the normative model of healthy subjects. An image could be presented to the trained GAN model of healthy subjects. The goal would be to find the most similar image generated by the GAN model to the presented image. The degree to which the presented and generated images are similar is determined by how closely the presented image follows the distribution of the data used to train the GAN model. As a result, differences between the presented and generated images could be used to detect anomalies in the presented image. [104, 105]. The use of adversarial autoencoders (AAEs) for that purpose is also interesting since they combine AEs with GANs properties [99].

2.4.2.1 Autoencoders

An AE is an unsupervised learning algorithm that learns a latent representation of the data by trying to copy its input to its output. The symmetric architecture of this NN is characterized by an input that is encoded into a latent space and then decoded and reconstructed to the output. A common property of AEs is that the size of the input layer and the output layer are the same. During the decoding process, the encoded representation, which commonly has a lower dimension than the input layer, is used to reconstruct the input data [89, 91]. This is possible because the encoded representation layer is designed to contain the key features that characterize the input data [See Figure 2.12].

The AE is a deterministic model since, at test time when the AE is fed with a particular input vector, the AE will always output the same result. A variational autoencoder (VAE) is a probabilistic version of an AE, in which the decoder's output consists of the parameters of a probability distribution of the input rather than the values of the input. This is useful for generating new data. A VAE can be characterized as a regularized AE which avoids overfitting and ensures that the latent representation layer contains good properties that allow the generative process [107].

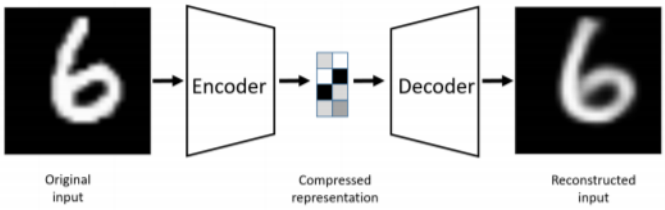


Figure 2.12: Structure of an AE. The original input is encoded, and the data is compressed into the latent space. The decoding from the latent space, which contains the important features, allows the reconstruction of the original image. (from [106])

2.4.2.2 Generative Adversarial Networks

GANs are NN that use unsupervised learning. They can generate samples very similar to the real data distribution by learning the existing patterns of the real data [108]. GANs provide a way to learn a deep representation of data by training a pair of networks in competition with each other, the generator and the discriminator [109]. To generate fake data that is identical to real data, the generative model learns the statistical distribution of real data. The goal of the generator is to deceive the discriminator. The discriminative model estimates the probability that a sample comes from the real data rather than from the generator. The discriminator is trained to maximise the probability of assigning the correct label to both training examples and samples from the generator, while the generator is trained to confuse the discriminator. The two networks are continually updating their knowledge and getting feedback on their successful changes. During optimization, each network improves its capabilities by learning from its weaknesses and taking advantage of the other weaknesses. The networks become stronger until establishing the Nash equilibrium when the generator and discriminator reach a state where one cannot progress without changing the other [See Figure 2.13] [103, 108].

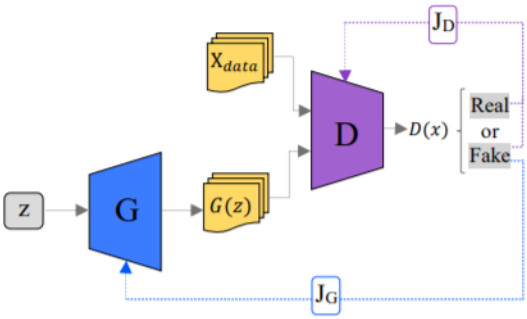


Figure 2.13: The architecture of GANs. X_{data} are real samples in the training dataset, and G(z) are fake samples synthesized by the generator G. Discriminator D judges the probability that the input data is real or fake. First, the generator takes the noise vector z (the random vector with uniform distribution or Gaussian distribution) of a fixed-length as input. Then, the generator synthesizes new data G(z) from the vector of random values. After training, the vector results in a compressed representation of the data distribution. This vector space is known as a latent space. The discriminator is trained to differentiate fake G(z) samples (attributes a value close to 0) from real X_{data} samples (attributes a value close to 1). Therefore, the discriminator requires the loss function JD and the generator requires the loss function JG to update the networks. While the generator updates its parameters only through the backpropagation signals of the fake output, the discriminator receives more information and updates its weights using fake and real output. (from [108])

2.4.2.3 Adversarial Autoencoders

The scientific community has been exploring and improving both methods and trying to fuse them in some cases. The AAE is a probabilistic AE that uses GANs approach to perform variational inference. Adversarial training may be applied between the latent representation space and a desired

prior distribution, resulting in a combined loss function that reflects both the reconstruction error and a measure of how different the predefined distribution is from that produced by a candidate encoding network [110]. In general, this AAE uses adversarial training to force the distribution of the latent code to look similar to a predefined desired distribution. This is achieved by incorporating a discriminator network into its structure. Similar to GANs model, the discriminator receives two types of inputs and tries to predict where its input data came from. It either receives random numbers sampled from the desired prior distribution, which is considered true data for the discriminator or receives the latent code, which is considered fake data for the discriminator. During training, the discriminator will make predictions regarding whether its input data was sampled from the prior distribution or the latent code. In summary, the adversarial training forces the encoder to produce a latent code able to mislead the discriminator into predicting that the inputted encoded samples are just another sample from the desired distribution [See Figure 2.14] [99].

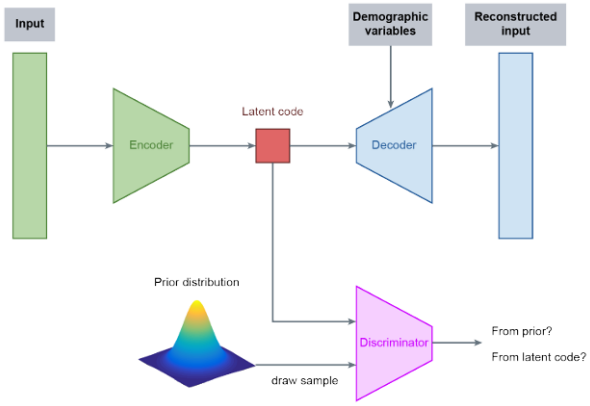


Figure 2.14: Structure of an AAE. The subject data is inputted into the encoder and then mapped to the latent code. This latent code is fed to the decoder with the demographic data, and then the decoder generates a reconstruction of the original data. During the model's training, the discriminator predicts if its input data came from the latent code (fake data) or randomly sampled from the chosen prior distribution, such as Gaussian distribution (real data). Based on these predictions, the AAE forces the encoder to produce a latent code similar to the prior distribution. (from [99])

Adversarial variational bayes is a type of AAE that introduces an auxiliary discriminative network on VAEs, connecting GANs and VAEs. This architecture allows to obtain a maximum-likelihood assignment for the generative model and the correct posterior distribution over the latent variables, in the nonparametric limit [111].

2.4.3 Fuzzy Clustering

Besides the generation of a normative model, it may be useful to cluster data that passed through the normative model in order to find groups and subgroups inside the continuous spectrum of NPDs. For that purpose, fuzzy logic seems to be very tempting. The use of fuzzy logic is promising for dealing with uncertain information without acute boundaries [112]. Real medical life applications give less importance to boolean variables, and fuzzy logic may be useful since it can take values between 0 and 1, which represent degrees of uncertainty [113]. In a fuzzy system, fuzzy values are often linguistic variables whose values are words or sentences. Fuzzification converts a binary value into a fuzzy value, with the operation of a membership function, which outputs a value between 0 and 1. Fuzzy-based if-then rules are applied, and the results of the individual rules are combined using fuzzy set operations. The final result, a fuzzy value, is converted to a binary value using the defuzzification method [113].

A fuzzy clustering algorithm can attribute more than one cluster to a particular object, which

may be useful for addressing the NPDs diagnosis paradigm since different diseases may have similar symptoms [113]. The use of clustering algorithms enables the generation of subgroups of NPDs that may help the clinical diagnosis. Fuzzy C Means Clustering [114], Possibilistic C Means Algorithm [115], Fuzzy Possibilistic C-Means Algorithm [116], and Possibilistic Fuzzy C Means Algorithm [117] are engaging fuzzy clustering algorithms that can address the problem of the continuous spectrum of NPDs [118, 119, 113].

2.5 State of The Art on Functional Brain Networks, and Normative Model for the Diagnosis of Neuropsychiatric Disorders

Neuropsychiatry deals with disorders at the intersection of neurology and psychiatric. NPDs are complex and not entirely understood, making them difficult to diagnose [7].

In an interesting review on NPDs, Taber *et al* [7] described that most NPDs diagnoses are made solely on the basis of the history and physical examination of the patient, but that neuroimaging exams can be helpful. Functional imaging should be considered for patients whose clinical symptoms are not expected for the working diagnosis. Additionally, structural imaging may also provide useful information. The authors referred that neuropsychiatric patients frequently have one or more of five symptom constellations: depression (includes apathy), psychosis (includes disturbances of perception and/or thought), agitation (includes anxiety and mania), behavioural dyscontrol (includes aggression and impulsivity), and cognitive disturbances (includes amnesia and dementia) [7]. Taber *et al* [7] further summarized the three major cortico-subcortical circuits important for neurobehavioural functions that are frequently impaired in NPDs [Figure A.1]. They also identified key aspects associating function to anatomical areas [Figure A.2] and highlighted that the incorporation of neuroimaging biomarkers, such as fMRI biomarkers, may provide an early diagnosis on some NPDs [7].

ADHD is the most common NPD in childhood. In 2020, the worldwide prevalence of persistent adult ADHD (with a childhood-onset) and symptomatic adult ADHD (regardless of a childhood-onset) was found to be 2.58% and 6.76%, respectively [120]. Besides, it is more frequent in males than in females, with a ratio of approximately 2:1. In the DSM-5 manual, ADHD is considered a neurodevelopment disorder and is characterized by impairing levels of inattention, disorganization, hyperactivity, and impulsivity [1]. It impairs several aspects of life, affecting social relationships, as well as academic and work performance [120]. The primary deficits of ADHD may affect social communication, limiting effective communication, social participation, and academic achievement. In addition, mild delays in language, motor, or social development often occur [1].

In 2017, Castellanos *et al* [121] did a research review on functional connectivity in ADHD. From their analysis, they highlighted that ADHD may be characterized by a decreased synchrony between the anterior and posterior nodes of the default mode network and by the interplay between several FBN, such as the default mode, executive control, and attention networks. By analysing some of the studies that led to those findings, it is possible to deeper explore the functional connectivity abnormalities in ADHD. The authors of [122] suggested that the long-range connections that link the dorsal anterior cingulate to precuneus and posterior cingulate are a possible candidate locus of dysfunction in ADHD. Differently, the authors of [123] proposed that the connectivity between the default mode network and ventral attention network are a key locus of dysfunction in ADHD. In another study [124], altered intra-network connectivity was observed in the default mode network, dorsal attention network, and visual network. The study from Sun *et al* [125] suggested that the development pattern of the interaction

between the dorsal anterior cingulate network and the default mode network is abnormal in ADHD. In a more recent review study [126], resting-state dysconnectivity in ADHD was also explored. The findings from that study supported that ADHD can be characterized as an inter- and intra- default mode network dysconnectivity disorder. In a very interesting paper, Aboitiz *et al* [127] explored the connection between the dynamics of the catecholaminergic signaling in the brain, and the activities of FBN that regulate behaviour. They discussed different lines of evidence including pharmacology, brain imaging and electrophysiology, that supported their proposal that involved abnormalities in the default mode network, ventral attention network, and salience network. A different study that also explored functional connectivity in ADHD with electroencephalography suggested that impairments in the ventral attention network contribute to the psychopathology of ADHD. Another study [128] that evaluated the functional connectivity of children with ADHD suggested that the dysregulation of FBN in children with ADHD not only involves the dorsal attention network and default mode network but also the somatosensory, motor, visual, and auditory networks. Still, regarding ADHD, it must be noticed that several publications and researches demonstrated language alterations in ADHD patients [129].

BD affects about 45 million people worldwide and is characterized by manic, hypomanic and major depressive episodes [130, 1]. In terms of symptomatology, family history, and genetics, bipolar and related disorders constitute a bridge between the SCZ spectrum and other psychotic disorders and depressive disorders. Family history is one of the strongest and most consistent risk factors for BD. In addition, there are some differences related to gender. As an example, females are more likely to experience depressive symptoms than males. The lifetime risk of suicide is increased in individuals with BD. This NPD decreases the quality of life since it leads to cognitive impairments, and prejudices work performance. Vocational difficulties may also be present in BD individuals. Considering that some ADHD symptoms overlap the symptoms of mania, it is possible to misdiagnose those NPDs [1].

Recently, Yoon *et al* [131] did a review study on altered functional connectivity in BD. They suggested that the pathophysiology of BD is influenced by disrupted intra-network and inter-network functional connectivity. The default mode, central executive, and salience networks, in particular, exhibit network hypoconnectivity. In addition, the connectivity between the default mode and the salience network was greater, and the connectivity between the salience network and the central executive networks was reduced when compared with healthy individuals. Recently, Wang *et al* [132] provided new insights into the neuropathology of BD affective and cognitive deficits, by analysing dynamic functional connectivity. In agreement to previous studies, they suggested that the salience and the basal ganglia networks play important roles in the neuropathology of BD patients, particularly in dysfunctional emotional processing and regulation. They also highlighted the role of the cerebellum network, default mode and sensory networks (sensorimotor, visual, and auditory networks) on the neuropathology of affective and cognitive deficits in BD. The default mode network can underline cognitive and affective processing problems, and in their study, it is pointed its possible role on depressive relapses. The cerebellum plays a role in the regulation of emotion, affect, and cognitive processes which are related to common symptoms in BD. On the contrary, the sensory networks mostly involved in motor and visual functions are not classical symptoms of BD, but also play a role in the neuropathology of BD [132].

Shashwath *et al* [133] studied the functional connectivity of BD and SCZ patients on five FBN (fronto/occipital, anterior default mode, meso/paralimbic, fronto-temporal/paralimbic, and sensory-motor). They found that the functional connectivity between the anterior default mode and fronto/occipital networks was abnormal in both groups. Besides, the functional connectivity between the meso/paralimbic and the sensory-motor networks and between the meso/paralimbic and the fronto-temporal/paralimbic networks was atypical in SCZ and BD patients, respectively. The study from Lois

et al [134] also provided novel insights into the role of the meso/paralimbic network in BD. Besides, they also referred that the majority of rs-fMRI studies in BD demonstrated abnormal ventral prefrontal cortex connectivity with the amygdala and other subcortical areas (thalamus, striatum) [134, 135, 136]. The authors of [136] also studied functional connectivity of both SCZ and BD patients. Their study focused on thalamic regions. For SCZ, they found disrupted thalamic communication to the frontal lobe, possibly affecting higher-order cognitive processes. For both SCZ and BD, increased connectivity between thalamic and sensory regions was observed, which highlights the potential role of the thalamus on psychosis sensory disruptions [136].

SCZ is a NPD that affects 20 million people worldwide [130]. According to the DSM-5, it belongs to the SCZ spectrum and other psychotic disorders, and it is characterized by two or more of the following domains: delusions, hallucinations, disorganized thinking (speech), grossly disorganized or abnormal motor behaviour (including catatonia), and negative symptoms (which include diminished emotional expression and avolition). The characteristic symptoms of SCZ involve a range of cognitive, behavioural and emotional dysfunctions. Considering that the diagnosis comprises the recognition of a constellation of signs and symptoms, individuals diagnosed with SCZ vary on most features since it is a highly heterogeneous NPD. In addition to the 5 characteristic symptoms, the assessment of cognition, depression, and mania symptom domains is crucial for differentiating between the various SCZ spectrum and other psychotic disorders. Genetic factors play a significant role in determining SCZ risk, and some risk alleles are shared by SCZ and other NPDs, such as BD. In addition, the age at the onset is usually later in females, and the symptoms tend to be more affect-laden among females. As in BD, the suicide risk is high. SCZ leads to significant social and occupational dysfunction. Thus, even when the cognitive skills are sufficient for the tasks at hand, it is difficult for the patients to keep a job. Additionally, it must be noticed that the distinction between SCZ and BD depends on the temporal relationship between the mood disturbance and the psychosis, and on the severity of the depressive or manic symptoms, which sometimes may lead to misdiagnosis.

Although the underlying causes and mechanisms of SCZ are uncertain, the hypothesis of functional brain "disconnection" is possible [137]. Jafri *et al* [138] demonstrated that SCZ patients had slightly more variability in functional connectivity than controls. Thus, healthy brain function is characterized by the presence of distinct, directed relationships among several FBN that are disrupted in SCZ, possibly reflecting cortical processing deficiencies [138, 9]. Garrity *et al* [139] compared the default mode network of SCZ patients with healthy subjects, and their results showed abnormalities in SCZ patients that correlated with both positive and negative symptoms. Compared with healthy subjects, patients with SCZ present altered temporal frequency and spatial location of the default mode network. [139, 9]. Besides the default mode network, the salience network is also an important FBN that seem to be involved in SCZ [140]. Calhoun *et al* [141] reviewed several articles and concluded that FBN can be used to identify patterns associated with NPDs. In their book, Shenton and Turetsky [9] suggested that SCZ-related deficits might arise from abnormalities in the quality or strength of functional connections among major nodes, such as the the the prefrontal–parietal, cingulate–opercular, temporal lobe, and the default mode network. Additionally, Liang *et al* [142] examined the functional connectivity at rest throughout the entire brain in SCZ. They found a large number of decreased functional connectivities in the insula, the temporal lobe (including the medial temporal structures), the prefrontal lobe and the corpus striatum. In particular, decreased functional connectivity was observed between the insula and all those regions (the prefrontal lobe, the temporal lobe and the corpus striatum). They demonstrated that the prefrontal, temporal lobe, corpus striatum, and insula may play an important role in the pathophysiology of SCZ. Additionally, they found increased connectivity between the cerebellum and other brain areas

[142]. In 2017, the review study from Bernard *et al* [143] provided further support for basal ganglia dysfunction in SCZ that may contribute to the symptoms and cognitive deficits presented in that NPD [143]. Li *et al* [144] did a very interesting study, where they related atypical FBN with genetic risk. They supported the theory that disconnectivity might be a possible endophenotype for SCZ. They found that SCZ patients and their first-degree relatives had similar disconnectivity patterns, even when the first-degree relatives did not present symptoms of the disorder. That was observed in the right executive control network and ventral default mode network, and that dysfunction was associated with symptoms of SCZ. They also highlighted the positive correlation between hallucinations and the linking between the left inferior frontal gyrus with the bilateral auditory cortex, right posterior temporal lobe, middle right anterior cingulate cortex, right ventral striatum, and left nucleus accumbens. Their study also showed hyperconnectivity between the bilateral thalamus and the language network, which may be associated with language, and consciousness abnormalities in patients with SCZ. Besides, functional connectivity between the higher visual network and right executive control network was positively correlated with the severity of positive symptoms in SCZ. Additionally, they found abnormalities between the right executive control network and the language network, and between the precuneus system and the ventral default mode network.

Calhoun and other colleagues [145] combined the temporal lobe and the default mode networks, derived from an auditory oddball task and a leave-one-out approach, to discriminate subjects with BD, chronic SCZ, and healthy controls. They were able to achieve an average sensitivity and specificity of 90% and 95%, respectively. Besides showing that the combination of those networks is a promising biomarker for SCZ and BD, this article demonstrates that FBN are an interesting tool for diagnostic classification even when two neuropsychiatric groups are included in the analysis [145, 9]. Recently, Xia *et al* [146] studied the similarities and differences in FBN among patients with SCZ, BD and major depressive disorder (MDD). They examined global network properties, whole-brain functional connectivity based on connectivity distances, and distance strength maps. In summary, functional network architectures exhibited a shift toward randomized configurations, with SCZ$>$ BD$>$ MDD. Furthermore, decreased short-range connectivity was observed for the different groups of patients, in the primary sensory and association cortices and the thalamus. Increases in medium/long-range connectivity were differentially localized within the prefrontal cortices among SCZ, BD and MDD patients. In their article, they highlighted different and common features of FBN across several patient groups, providing useful pieces of information for the diagnosis of NPDs [146].

Regarding SCZ classification-based articles, Arbabshirani *et al* [81] used FBN features for an automatic classification algorithm. Besides using inter-network connectivity features, they introduced the concept of autoconnectivity which captures the correlation of a timeseries with its lagged version. FBN were obtained using ICA, and after extracting and selecting features, the authors fed a support vector machine algorithm to perform dichotomous classification between healthy controls and SCZ patients. Their results, when combining functional network connectivity and autoconnectivity, reached an accuracy, sensitivity, and specificity of about 88%, 87%, and 90%, respectively. In a very different study, Oh *et al* [147] aimed to detect SCZ in structural functional magnetic resonance imaging (sMRI) datasets using a deep convolutional NN. They used a large and diversified dataset and obtained an area under the receiver operating characteristic curve (AUC) of 0.96. At the optimal operating point, sensitivity and specificity were 92% and 85%, respectively. The authors pointed out that compared with similar studies that used ML algorithms, their results were better, showing the promise of DL for medical diagnosis [147].

The diagnostic of NPDs using DL and ML methods is very promising. Commonly, it is mostly

based on a dichotomous classification approach. However, some studies already included a few numbers of other similar disorders to perform a classification task that simulates the clinical practice [19]. Nonetheless, to deal with the continuous spectrum of NPDs and to achieve a translation for clinical practice, the normative model approach previously mention in Section 2.4.2 seems to be more interesting [148, 149, 100, 150, 151, 152, 153, 101, 99, 154, 155, 156, 157, 158]. In Table. 2.1, some of the most interesting studies in the literature that used a normative model are summarized.

Table 2.1: Studies using normative model approach

Article	Analysed Conditions	Data Type	Normative Model Method
Kessler *et al* [154]	ADHD	rs-fMRI	Polynomial Regression
Marquand *et al* [149]	ADHD symptoms	task-fMRI	Gaussian Process Regression
Kia and Marquand [150]	SCZ, ADHD, BD	task-fMRI	Neural Processes
Wolfers *et al* [151]	SCZ, BD	sMRI	Gaussian Process Regression
Pinaya *et al* [101]	SCZ, ASD	sMRI	AE
Fernandes [158]	AD	rs-fMRI	AE
Zabihi *et al* [152]	ASD	sMRI	Gaussian Process Regression
Wolfers *et al* [155]	SCZ, BD	sMRI	Gaussian Process Regression
Pinaya *et al* [99]	MCI, AD	sMRI	AAE
Zabihi *et al* [156]	ASD	sMRI	Gaussian Process Regression
Kia *et al* [157]	ADHD, SCZ, BD, AD, MDD, MCI	sMRI	Hierarchical Bayesian Regression

sMRI is the type of data that is commonly used to study NPDs with the normative model approach. Nevertheless, the authors of the studies [154, 149, 150, 158] used fMRI to investigate NPDs. In a technical report, Fernandes [158] created a normative model using an AE with rs-fMRI data, but it tested the model on task-fMRI data from healthy and AD subjects. This study is relevant for this dissertation, since it was an inspiration for this project. Kessler *et al* [154] identified several limitations in their study, but their results were promising. Marquand *et al* [149] introduced a normative model for analysing heterogeneity in clinical groups. They mapped the relationship between trait impulsivity and reward-related brain activity in a large healthy group. Impulsivity and impairments in reward processing are core features of many disorders, including ADHD, and they used delay discounting to quantify impulsivity. Regarding methodology, they used a Gaussian process regression, which is the most common approach, to predict a set of biological response variables (e.g. neuroimaging) from a set of clinically relevant covariates (e.g. delay discounting), while estimating predictive confidence for every prediction. The normative model is estimated with Gaussian processes. This provides the ability to predict brain activity for any value of the clinical covariates along with measures of predictive confidence. Then, for each subject, a normative probability map that quantifies the deviation from the normative model at each brain region is computed, and a measure of abnormality for each subject using extreme value statistics is generated, which can be related to clinically relevant variables. In summary, they concluded that normative model provides a natural framework to study disorders at the individual participant level without dichotomous classification. [149]. In another study, Kia and Marquand [150] compared a DL model to a multi-task Gaussian process regression approach for generating the normative model. In fact, they used neural processes, which are latent variable models, and their results showed that the use of DL leads to substantial improvements on the normative model process [150]. Wolfers *et al* used a Gaussian process regression to generate a normative model in order to investigate differences in brain structure

among patients with SCZ and BD and to map the brain alterations linked to these disorders at the level of individual patients. Their results showed that deviations from the normative pattern were frequent in both disorders but highly heterogeneous at the level of the individual. On average, patients with SCZ had significantly reduced grey matter in frontal regions, cerebellum, and temporal cortex. Besides, patients with BD presented mean deviations in cerebellar regions. Their study found that group-level differences do not take into account inter-individual differences among patients with the same diagnosis. This finding suggests that the average patient analysis is a noninformative parameter due to high heterogeneity within the same NPDs group [151]. Zabihi *et al* [152] and Wolfers *et al* [155] did similar studies to that of the authors of [151]. They also used a Gaussian process regression to generate the normative model, and investigated brain structure alterations using sMRI. Zabihi *et al* [152] studied ASD and found few significant group differences between healthy and ASD subjects. Nevertheless, the ASD group showed highly individualized patterns of deviations in cortical thickness that were widespread across the brain. They reinforced the notion that the ASD group is highly heterogeneous and that by focusing on the average ASD participant, they do not take into account inter-individual variations, which are crucial for precision medicine [152]. Regarding the study of Wolfers *et al* [155], they tried to replicate a previous study [151] and their results were in agreement. They revealed robust group-level differences between patients and healthy controls, but only a few patients with SCZ and BD exhibited similar deviations from the expected pattern. Their findings support the idea that group-level differences analysis hides considerable individual differences [155]. Later, Zabihi *et al* [156] tried to solve a problem that they and other colleagues faced in previous studies [151, 152, 155] regarding high heterogeneity within the same pathological group. In fact, they focused on ASD, and considered of central importance the identification of subtypes within that NPDs to facilitate biomarker discovery. Thus, they considered that by combining a normative model with clustering they would be able to summarize these complex and highly individualized patterns of deviation into a small number of interpretable ASD subtypes. Their results identified five distinct and interpretable subtypes. Three of these clusters showed relatively widespread decreased cortical thickness and two showed relatively increased cortical thickness. These distinct and opposing expressions of cortical thickness explain high heterogeneity within the larger ASD group. Subdividing individuals with ASD into more homogenous subtypes seems to be interesting for clinical practice [156]. Recently, Kia *et al* [157] used hierarchical Bayesian regression to generate a normative model for evaluating neuroimaging data of six different NPDs. The main advantage of this method is that it can deal with data from different imaging devices. This provides the possibility to learn the normative range of structural and functional brain measures on large multi-site data and to recalibrate and reuse the learned model on local small data. Therefore, this study is the most universal approach published in the literature until now [157]. This fascinating and alternative approach for diagnosing NPDs was also proposed by Pinaya *et al* [101]. However, they used a deep AE for creating a normative model using structural imaging data from healthy people. Then, they used this model of healthy subjects for estimating individual differences of patients with SCZ and ASD. Their results were in agreement with the literature, revealing distinct patterns of neuroanatomical deviations for both disorders. They concluded that their approach provides a flexible and promising framework for assessing total and regional neuroanatomical deviations in neuropsychiatric populations [101]. Later, Pinaya *et al* [99] did a very similar study, but they used a different DL model to investigate AD and MCI. The AAE was used to generate the normative model, and overall their results were comparable with traditional classifiers. The use of DL enables the generation of a normative model that learns multiple levels of representation about the complex structure of the data and identifies the most important features of the healthy brain.

2.5.1 Summary

A detailed literature review on the functional connectivity abnormalities in ADHD, BD, and SCZ was performed. Additionally, some studies that led to promising results for the diagnosis of SCZ were exposed. Finally, the studies that used the normative model approach in the context of NPDs were also reviewed.

Although there are structural differences among neuropsychiatric patients, using rs-fMRI to assess functional connectivity in NPDs is also promising and should be explored further as a possible way to find NPDs biomarkers. Additionally, the use of DL methods for classification tasks is leading to good results. Still, the translation of dichotomous classification methods for clinical practice is difficult due to biological heterogeneity and undefined boundaries of NPDs. A normative model approach can deal with those inter-individual differences among patients with the same diagnosis by enabling individualized analysis. Also, the use of DL seems to improve the normative model. However, their diagnosis results are still far from the accuracy of the dichotomous classification articles. Besides, only a few recent scientific studies went further the normative model and clustered patients into different subgroups for studying that heterogeneity within a particular diagnosis.

Considering that, in this dissertation, rs-fMRI data is combined with a DL normative model approach and a clustering algorithm. Thus, it will be possible to study functional connectivity patterns that characterize each NPD, and by testing the normative model on an independent test set of SCZ patients, it will be possible to study the generalization of the model, and the heterogeneity within the SCZ individuals.

Chapter 3

Materials and Methods

3.1 Data Description

The brain rs-fMRI images used in this project were obtained from three public neuroimaging databases. The Consortium for Reliability and Reproducibility (CoRR) public database [159] was chosen because to it contains a large number of different datasets with data from healthy subjects, which is necessary for creating a generalized model of healthy subjects. The UCLA Consortium for Neuropsychiatric Phenomics LA5c Study (UCLA) public database [160] was chosen to train and test the normative model since it contains healthy and non-healthy subjects.The Center for Biomedical Research Excellence (COBRE) public database [161] was also chosen to evaluate the performance of trained model on an independent dataset.

The data from CoRR consist of 1725 healthy subjects from 36 different datasets. The data from UCLA consist of 130 healthy subjects, 43 ADHD subjects, 49 BD subjects, and 50 SCZ subjects. The data from COBRE consist of 91 healthy subjects, 9 BD subjects, 11 schizoaffective (SCZA) subjects and 74 SCZ subjects.

There was no phenotypic information about 1 subject from the XHCUMS dataset, 1 subject from the IPCAS3 dataset, and about all the 114 subjects from the DC1 dataset. Those datasets are part of the CoRR database. Consequently, those subjects were all excluded from the subsequent description of data. The phenotypic information of the other subjects, particularly their age and gender, was distributed by the number of subjects, as shown in Figure 3.1.

During a rs-fMRI scan, the participant is asked to refrain from performing a specific task. However, in terms of the instructions given to the participant about their eyes condition, rs-fMRI protocols can differ. The sample contains 845 healthy subjects that underwent scan with their EC, and 985 healthy subjects that had their eyes open while scanning (518 with the eyes fixated on an object, 125 with their eyes open towards a blank screen, 25 were stimulated by a word, and 317 were asked to keep their eyes open and relax). 94 non-healthy subjects were asked to maintain their EO-F, while 142 were asked to maintain their EO and relax.

The MRI scan of each dataset depends on the institutes where each study was performed. Figure 3.2 displays the distribution of the number of subjects by the manufacturer, model, field strength, and head coil of the specific scan equipment.

Besides the MRI scan, the parameters of scan are also useful to describe the data. Figure 3.3 displays the distribution of the number of subjects by the scan duration, TR, number of measurements and TE.

3. MATERIALS AND METHODS

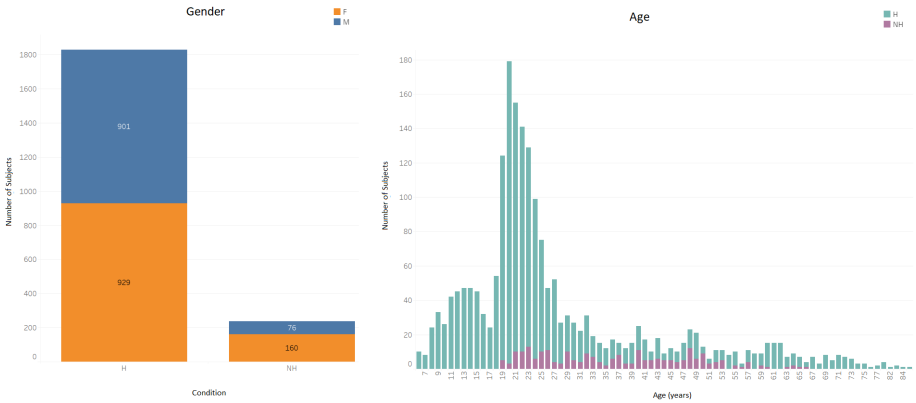


Figure 3.1: Phenotypic information about the subjects from CoRR, UCLA and COBRE databases. On the left, there is a graph showing the number of females (F - orange) and males (M - blue) for the group of healthy subjects (H - left bar), and for the group of non-healthy subjects (NH - right bar) consisting of the SCZ, BD, ADHD, and SCZA patients. A histogram is shown on the right, with the number of healthy (H - blue) and non-healthy (NH - purple) subjects distributed by their age.

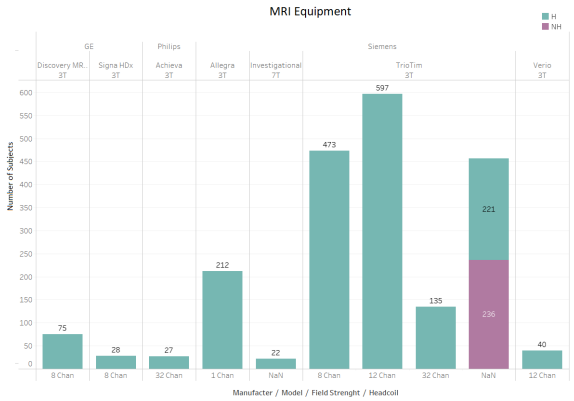


Figure 3.2: Distribution of the number of healthy subjects (H - blue), non-healthy subjects (NH - purple) by the fMRI scan. NaN stands for “no information” about a variable.

Figure 3.3: Distribution of the number of healthy subjects (H - blue), non-healthy subjects (NH - purple) by the scan duration, TR, number of measurements and TE.

3.1.1 Data Selection

The process of data selection was taken carefully since several factors influence fMRI data, including the MRI scanner, scan parameters, age, and gender. Although the data sample contains many healthy subjects, high variability is present between the different datasets.

The different scan equipment used by the various institutes results in a noticeable and unavoidable difference between the datasets. The scan parameters and the condition of the eyes during the scan, on the other hand, contribute to data diversity but are more controllable variables. As a result, conditions were set up to ensure a large sample size while reducing the variability caused by those factors.

First, the age range was defined as between 18 and 50 years old to exclude children and older adults.

As previously mentioned in section 2.3.1, eyes condition at scan influence fMRI data. Taking that into account, it was then decided to only select subjects that had their EO or EO-F during the rs-fMRI experiment since all non-healthy subjects included in the sample underwent the scan with their EO (UCLA database) or EO-F (COBRE database). The 25 subjects stimulated with a word were excluded because that experiment can be considered a task-fMRI. Although there may be residual differences between the other three open eyes conditions, they were all included in the study.

Despite the advantages of using short TRs and longer scan duration, as exposed in the Section 2.3.1, most studies acquire 5-12 min of rs-fMRI data at TR$<$2s. The selected sample of 866 individuals contained only 36 subjects that were underwent scan with TR lower than 2s and 115 subjects with scan time longer than 10 min. In addition, most subjects underwent scan with TR equal to 2s for approximately 5 min. Considering that, and knowing that the TR has an impact on fMRI data, it was decided to define a TR equal to 2s, as the inclusion of different TRs would not significantly increase the sample size. Despite the fact that longer scan durations are associated with higher data quality, the sample had insufficient data with longer acquisitions. Consequently, it would not improve the study in any way, and it might even increase variability. Therefore, to make the data more homogeneous, the data were trimmed to 150 time points, which is the smallest number of measurements [This step is supported by the studies presented in Appendix A.2]. There were no conditions set in place for TE.

It must also be noticed that the variability caused by the different MRI scans can not be managed in the clinical practice since we can not decide which MRI scan the hospital must have. Concerning that, different MRI scans were included in the sample to get a more robust model.

In the end, the sample was reduced to 712 subjects. Table 3.1 presents additional information about the selected sample

Table 3.1: Information about the data that was selected for the subsequent methods. Legend for eyes condition: 1- eyes open fixated on a target, 2- eyes open towards a blank screen, 3- eyes open

Dataset	Number of Subjects	Females (%)	Age Range	Eyes Condition	Manufacturer	Model	Headcoil	Field Strenght	TE (ms)	TR (s)	Acquisition Time (min:s)	Number of Time Points	Voxels (mm^3)
BNU3	H-48	50	18-30	2	Siemens	TrioTim	12 Chan	3T	30	2	8:06	240	3.5x3.5x3.5
HNU1	H-30	50	20-30	1	GE	Discovery MR750	8 Chan	3T	30	2	10:00	300	3.4x3.4x3.4
IPCAS1	H-30	70	18-24	1	Siemens	TrioTim	8 Chan	3T	30	2	6:54	205	4x4x4
IPCAS3	H-34	65	18-25	1	Siemens	TrioTim	8 Chan	3T	30	2	6:00	180	3.4x3.4x3
IPCAS4	H-20	50	21-28	1	GE	Discovery MR750	8 Chan	3T	30	2	6:04	180	3.5x3.5x3.5
MRN1	H-39	59	18-49	1	Siemens	TrioTim	12 Chan	3T	29	2	5:04	150	3.8x3.8x3.5
NYU2	H-81	36	18-48	3	Siemens	Allegra	1 Chan	3T	15	2	6:00	180	3x3x4
Utah1	H-16	0	18-39	2	Siemens	TrioTim	12 Chan	3T	28	2	8:06	240	3.4x3.4x3
Utah2	H-1	0	39	2	Siemens	TrioTim	12 Chan	3T	28	2	8:06	240	3.4x3.4x3
UCLA	H-130 ADHD-43 BD-49 SCZ-50	43	21-50	3	Siemens	TrioTim	NaN	3T	30	2	5:04	152	3x3x4
COBRE	H-74 BD-4 SCZA-8 SCZ-55	26	18-50	1	Siemens	TrioTim	NaN	3T	29	2	5:04	150	3.75x3.75x4.55

It must be noticed that the initial data sample contained 2182 subjects, and after data selection, only 712 subjects were retained. The renowned scientist Andre Ng once wrote that “if 80 percent of our work is data preparation, then ensuring data quality is the important work of a ML team”. Although more data usually leads to better performance, data quality makes the difference in ML models. Data selection was

performed in order to maximise the amount of data while maintaining data quality. The use of all data would probably not improve the model, since data variability would diminish the quality of data. This balance between the amount and quality of data supported this step of data selection, which resulted in a smaller data sample but guaranteed data quality.

3.2 Data Processing

3.2.1 Brain Extraction

Brain extraction of the T1-weighted structural images was the first step of image preprocessing. This step improves the registration of the functional to the anatomical image. It aims to remove non-brain tissues, such as the skull, dura and eyes, while retaining the brain tissues of the anatomical image. Brain extraction was performed with the Advanced Normalization Tools (ANTs) software [162]. After testing several templates, the OASIS template was selected for brain extract the T1-weighted images of all subjects, since it led to the finest results. During this step, 7 subjects did not have a T1-weighted image and were consequently excluded from the subsequent analysis.

3.2.2 Preprocessing of the Functional Image

After finishing the preparation of the anatomical images, the preprocessing of rs-fMRI images was handled within the FSL software [163]. The FSL's MELODIC tool was used to perform several steps at once: (i) it included discarding the first 4 volumes of functional data to ensure BOLD fMRI signal stabilization; (ii) it also included motion correction using MCFLIRT tool, which attenuates the effect of head motion by spatially registering each volume separately to the middle volume, and estimates motion parameters; (iii) spatial smoothing using a Gaussian kernel with 5 mm FWHM was also applied, which has the effect of blurring the images, and is achieved by calculating, at each voxel, a weighted average over the multiple voxels; (iv) in order to eliminate unwanted signal components from the timeseries of each voxel, a high-pass filter with a cut-off of 0.01 Hz was also applied to remove frequencies lower than the low-frequency fluctuations that dominate the BOLD signal; (v) it also generated the necessary transformations to align the functional data to the standard space without applying the registration; (vi) finally, single-session spatial-ICA with automatic dimensionality estimation was run, which allowed data to be unmixed into a set of independent components, each described by their own timecourse and a spatialmap [30].

After running the Melodic ICA, it was possible to analyse the outputs of the MCFLIRT tool in order to evaluate the motion parameters of each subject. For defining general criteria for head motion rejection, the minimum value of each voxel size was found to be $3 \times 3 \times 3$ mm^3. Considering that, data in which the maximum framewise displacement exceeded 3 mm, or motion in translation exceeded 3 mm, or angular deviation exceeded 3º were excluded from the sample. Following these criteria, 32 subjects were removed from the analysis. In addition, it was not possible to finish this step for 6 subjects due to some errors, including the lack of a rs-fMRI image. The 8 SCZA patients and the 4 BD patients from COBRE were also excluded due to their reduced sample size. In the end, 655 subjects continued to the subsequent analysis.

3.2.3 Data Cleaning

ICA was here used for signal-denoising, since it allows the detection of artefacts in the data, such as physiological noise, scan-related artefacts, or head motion that can not be corrected by standard techniques such as MCFLIRT. FSL's tool FIX was used to automatically detect and remove the non-brain signal components. FIX classifies each component into brain signal, or noise, by applying a trained dataset. For each ICA component, this tool generates several spatial and temporal features, each describing a different aspect of the data. These features are then fed into a multi-level classifier. The classifier is trained by performing hand-classification of the ICA components, on a large number of subjects. After training, the classifier can automatically classify new data and remove the noise components from the original data [84].

For creating a robust training set, ten healthy subjects from each dataset (and the subject from Utah2) were randomly selected, and hand label classification of the ICA components was performed. The trained dataset was then used by FIX for classification and removal of the noise component of all subjects in the sample.

3.2.4 Registration

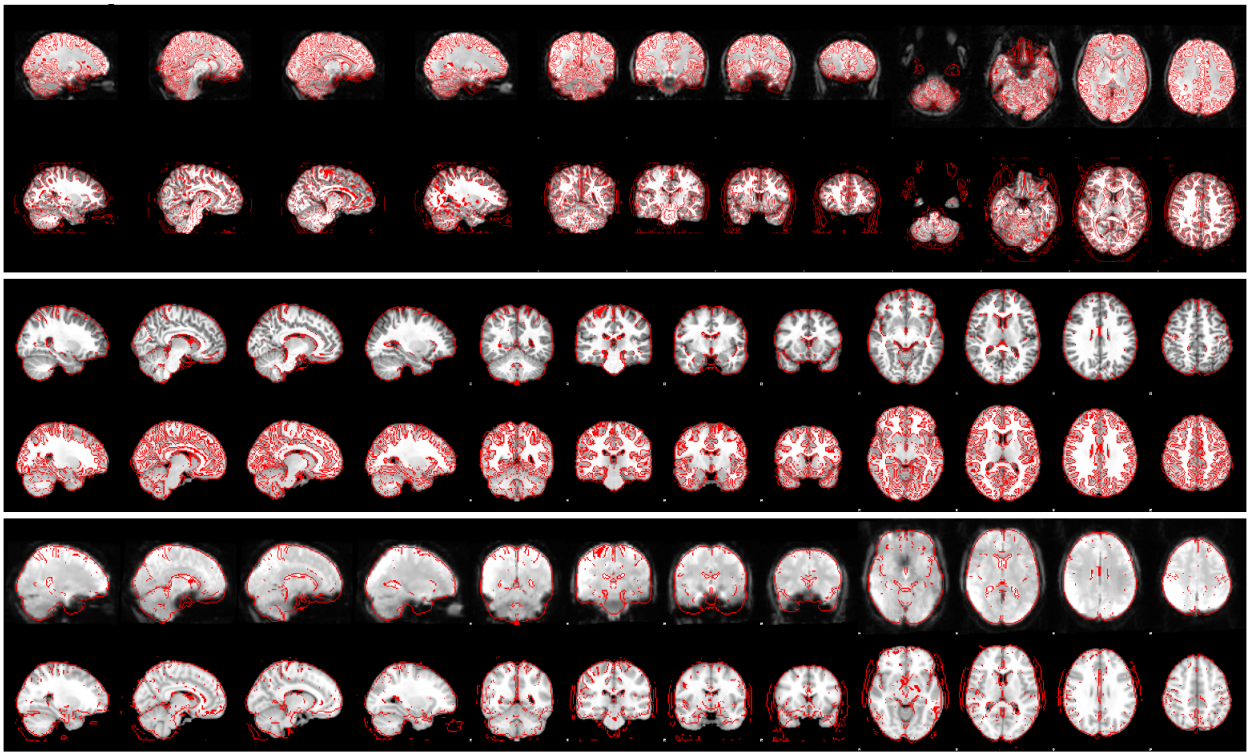


Figure 3.4: Example of the registration process for one patient. Top image: registration of the subject's functional image to the subject's anatomical image; Middle image: registration of the subject's anatomical image to the MNI152 standard space; Bottom image: registration of the combined information from the subject's functional image to the MNI152 standard space.

The cleaned 4D preprocessed functional image underwent registration to the individual T1-weighted structural image, and then to the MNI152 standard space. Registration is used to align images so that the voxel location corresponds to the anatomical location. It is useful for group analysis since it registers all subjects to the same standard space. First, the functional image is registered to the T1-weighted structural image through the boundary-based registration (BBR) method. It is based on a cost function, which focuses on the intensities around the white matter boundaries. The boundaries location

are determined from the structural image and are mapped onto the functional image, normally using a 6- degrees of freedom (DOF) spatial transformation (rigid-body transformation, which involves three rotations, and three translations). This rigid-body transformation is the most adequate for situations where the anatomy does not change, such as this registration of the functional to the structural image of the same individual [24]. Second, the T1-weighted structural image is spatially normalized to the MNI152 standard space, through a non-linear transformation (a transformation with more than 12-DOF). The non-linear transformation is initialized by an affine transformation, which consists of a linear transformation with 12- DOF (three rotations, three translations, three scalings and three shears), since it leads to significant improvements in robustness and accuracy. The affine transformation is done so that a good approximation of the position, orientation, and size of the brain can be established before applying the nonlinear steps. When there is a significant difference in anatomy or any type of local geometric change, these steps are critical because they allow the transformation to align anatomical structures from different subjects. [24]. In the end, this two-step registration is combined, and the individual's functional image is registered into the MNI152 standard space [Figure 3.4].

3.3 Functional Brain Networks

3.3.1 Dual Regression

To perform comparisons between a group of subjects, it is common to run a group-level analysis, to ensure that the components are the same across all subjects, and then map these group components back to individual subjects to get comparable individual information.

Considering that, dual regression was here applied. It used a group template and applied two subsequent regressions, using the registered functional data from each subject to derive the subject-specific maps.

The use of a group template available in the literature, instead of creating a specific group-ICA template, has the advantage of ensuring that the components used for dual regression are consistent with FBN previously identified in the literature.

Therefore, the ninety functional ROIs from *Shirer et al.* [44] were concatenated to obtain the fourteen individual networks identified in their study, using *fslmerge*. Those fourteen FBN were used as a template map for the dual regression analysis. In stage 1 of dual regression, the functional data were used as input data (dependent variable), and the components from the *Shirer et al.* template were the spatial regressors (independent variables). This regression outputted a set of subject-specific time series, describing the temporal pattern of each brain region for each subject. The second stage of dual regression used this outputted set of time series as independent variables, against the functional data (dependent variable), and outputted for each subject, one map for each original group-template component [Figure 3.5].

3.3.2 Network Matrices

To compare the functional connectivity across subjects, a graph-based connectivity analysis was performed. This approach allows the investigation of functional connectivity differences across individuals. Considering a graph as a diagram representation of nodes and edges, where nodes are the FBN and edges are the connections, it is possible to study the functional connectivity among pairs of FBN. The graph-based method involves four steps: the definition of nodes; the time series extraction from each node; the connectivity calculation (edges); and the creation of a network matrix that describes all possible pair-wise connections. The first two steps were previously performed with

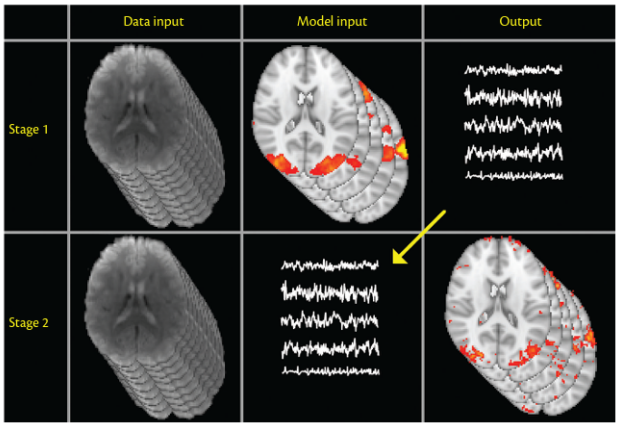


Figure 3.5: Dual regression uses two subsequent regressions, with the aim of outputting subject-specific spatial maps. The dependent variable on both stages of dual regression is the registered functional data. In stage 1, the independent variables are the components from the group template. In stage 2, the outputs of stage 1, the set of time series, are used as independent variables and the spatial maps for each subject are outputted [30].

FSL's dual regression (Section 3.3.1). For the two last steps and all subsequent methods, Python 3 was the selected programming language while Pycharm was the preferred environment to work with. For finding the edges between nodes, Pearson's correlation coefficient between the time series of each FBN was calculated. This metric provides a numerical value, which measures the linear correlation between two variables. The Pearson's correlation coefficient ranges between -1 (perfect negative correlation) to +1 (perfect positive correlation), and is calculated as follows:

$$r = \frac{\sum_{i=1}^{150}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{150}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{150}(y_i - \bar{y})^2}}, \tag{3.1}$$

where x and y are the time series of two brain regions, and i ranges between 1 and 150, since the number of points of each time series after trimming is equal to 150.

This metric was calculated between all the 14 nodes, creating a 14x14 functional network connectivity matrix for each subject, that gives information about inter-network correlations. Considering that Pearson's correlation coefficient has no information about the direction of functional connectivity, the network matrix is symmetric. Therefore, to avoid redundancy, the diagonal matrix, and one half of the network matrix was discarded. Taking into account that $(14 \times 14 - 14)/2 = 91$, the other half of the matrix was transformed into a vector containing 91 elements for feeding the normative model.

Figure 3.6 shows an example of a generated correlation lower-triangular matrix for a particular subject, where each element represents the Pearson's correlation between two pairs of FBN.

Those matrices inform us about the relationship between the BOLD signals that were extracted from different FBN. If those FBN are correlated, their BOLD signals should show similar patterns over time, indicating that they are functionally connected.

3.4 Normative Model

The AE architecture was chosen for the generation of the normative model out of the three possible architectures mentioned in Section 2.4.2. Although GANs seem to be quite interesting, their application

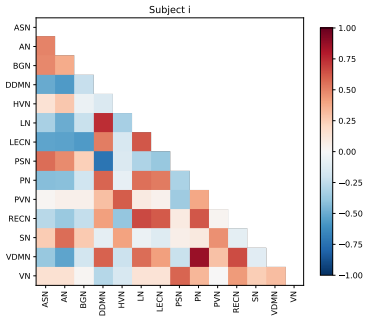


Figure 3.6: Example of a functional network connectivity lower-triangular matrix. The presented matrix was obtained after all steps of data processing.

for the normative model is more conceptually challenging, as it requires finding the most similar generated image to the presented image. The process is clearer for AEs, since the underlying idea is that at test time, pathological data will pass through the trained AE, and after being encoded to the latent representation (which implies data reduction), it will be difficult to reconstruct it similarly to the input. Furthermore, there are lack of studies that support the use of GANs for a normative model approach applied to NPDs. VAEs, and AAEs are also viable options, but since both can be considered upgraded AEs, it would be prudent to start with a simpler AE and gradually improve its architecture.

3.4.1 Data Split and Data Standardization

After defining the main architecture of the model, and before optimizing the AE particularities, the 655 subjects that remained after data processing were split into different sets of subjects Table 3.2. Before

Table 3.2: Data sets for the normative model. Data were split into several sets of subjects. H-Train represents the 366 healthy subjects from CoRR and UCLA databases that were used to train the AE. H-Test-U is the set of 39 healthy subjects from the UCLA database that were used to test the AE. SCZ-Test-U, BD-Test-U, and ADHD-Test-U are the sets of neuropsychiatric patients consisting of 47 SCZ patients, 45 BD patients, and 39 ADHD patients from the UCLA database, respectively. H-Test-C and SCZ-Test-C are the independent test sets consisting of 70 healthy subjects, and 49 SCZ patients, respectively.

Set Group	Dataset	Number	Females (%)	Age Range	Age Mean
H-Train	CoRR and UCLA	366	47 %	18-49	26
H-Test-U	UCLA	39	51 %	21-50	34
SCZ-Test-U	UCLA	47	26 %	22-49	37
BD-Test-U	UCLA	45	44 %	21-50	35
ADHD-Test-U	UCLA	39	49 %	21-50	32
H-Test-C	COBRE	70	30 %	18-50	35
SCZ-Test-C	COBRE	49	18 %	19-50	31

feeding the model with the 91 features that were obtained after calculating the Pearson's correlation coefficient between each pair of FBN, those features were scaled. Feature scaling can be performed by standardizing the data, with Z-score, for each feature:

$$z = \frac{x - \mu}{\sigma} \tag{3.2}$$

where x represents the original value of the feature for a specific subject, μ represents the mean value of the feature for all subjects, and σ the standard deviation value of the feature for all subjects.

To calculate the mean, and the standard deviation values for standardization, only the subjects belonging to the H-Train were used. After that, standardization was applied to all sets of subjects.

3.4.2 Autoencoder Optimization

For optimization, only the training set fed the model, to avoid bias on the analysis. 10-fold cross-validation was applied to validate the optimized model while studying its generalization. It consisted of separating the training set into 10 different sets with an equal number of subjects. The proportion of databases was maintained equal for the different folds. The model was trained and tested ten times, using for each time a different fold as the validation set.

The latent code for the AE architecture applied to the normative model should have a lower number of neural units, to ensure that the reconstruction of the input data for pathological subjects fails more frequently than for healthy subjects. A large latent code would provide more flexibility in reconstructing the pathological image. Considering that, and the concepts covered in Section 2.4.1.1, and Section 2.4.1.2, for the first step of model optimization, the hyperparameters were defined and several AEs with 3 hidden layers, differing in the number of neural units in the hidden layers were tested. The AE was implemented using the *Keras* Python library. To select the number of neural units in each layer, a logical interpretation was performed. Consider that each feature corresponds to a correlation between two FBN and that it is known that 14 FBN are presented in the analysis. If we selected 13 features, we would guarantee that each of the 14 FBN would influence those 13 features. Therefore, 25 features would have the influence of each FBN at least two times. The FBN would influence the 91 features 13 times, since all features would be included in the analysis. Taking that into account that data is reconstructed from the latent code, it would make sense to guarantee that it would be influenced by all FBN. Several architectures were tested using this reasoning while maintaining the other parameters of the AE. At this time, the hyperparameters were defined considering the literature considerations. The activation function of the hidden layers was the Leaky ReLu, while a linear activation function was applied to the output layer. L2 regularization parameter of 10^{-6} and a dropout value of 0.5 was applied (it must be noticed that in Python the referred dropout value is the fraction of neural units to drop). Mean squared error was the loss function, the weights were initialized following He initialization, and the model was trained for 1000 epochs, using Adam optimization ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-7}$, and a learning rate of 0.0005).

To evaluate model performance, the mean squared error between the reconstructed data and the data that was inputted in the model was calculated for each subject (mean squared error per subject (MSE_s):

$$MSE_s = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}i - yi)^2, \tag{3.3}$$

where n is the number of features (91), $\hat{y}i$ is the reconstructed feature i, and yi is the inputted feature i.

For each run of the cross-validation, a boxplot was constructed for the training set and the validation set. After cross-validation, the mean of the median MSE_s values that were determined for all of the 10 runs was calculated for the training set and the validation set. The difference between the mean of the median MSE_s values of the validation set and the mean of the median MSE_s values of the training set was also determined. The mean of the median MSE_s values of the training set was used to evaluate the performance of the model on reconstructing the data, and the difference between the mean of the median MSE_s values of the validation and the training set was used to evaluate the generalization of the model. Considering that a conjugation between both metrics is crucial to decide which one is the best model

architecture, and by knowing that the range of their values was different, both metrics were normalized.

$$x_{norm} = \frac{x_{original} - min}{max - min} \tag{3.4}$$

where x_{norm} is the normalized value of the evaluation metric that is being normalized, $x_{original}$ is the original value of the evaluation metric that is being normalized, and min and max are the minimum, and maximum values of the evaluation metric that is being normalized, respectively.

After electing the best number of neural units for each hidden layer, the elected model was optimized by varying the other parameters of the model (number of epochs, dropout value, L2 parameter, activation function, and learning rate). The same evaluation metrics were used, and the best AE architecture was selected.

3.4.3 Testing the Autoencoder

After defining the best model architecture, the AE was trained using the 366 healthy subjects belonging to the H-Train set. Subjects from the H-Test-U, SCZ-Test-U, BD-Test-U, ADHD-Test-U were used to test the AE.

To evaluate how well the model worked, the mean squared error between the reconstructed data and the data that was inputted in the AE was calculated for each subject of the different test sets (MSE_s in Equation 3.3). The MSE_s corresponds to a global value per subject, representing the reconstruction error. Then, all subject's MSE_s values were used to generate a boxplot for each test group, and the median values were compared between the different test sets. Besides, a Shapiro-Wilk test was applied to each group of subjects, in order to test the null hypothesis that the data was drawn from a normal distribution. Then, an adequate statistical test was applied between the MSE_s values of different groups of subjects, to evaluate if the differences between groups were statistically significant.

Additionally, the inputted data vector was subtracted to the outputted data vector for each subject, and the absolute value of that difference was calculated. A 91 row vector for each subject was created, representing the absolute difference between the reconstructed data and the original data for each feature. Those difference vectors were binarized. The threshold was decided after analysing the distribution values of all difference vectors. The 91-row binarized vectors were reshaped into a 14x14 symmetric binarized matrix. The binarized matrices only retained the higher values. The network degree of the binarized matrices was calculated for each subject, using graph theory. For each test group, the median value of the number of degrees was calculated for each FBN. In the end, for each test group, a 14 row vector was generated. This vector represented the median value of the number of degrees for each FBN.

Afterwards, to study which pairs of FBN were abnormal in each test set, the mean squared error was calculated for each feature, giving a global metric per feature (mean squared error per feature (MSE_f).

$$MSE_f = \frac{1}{N}\sum_{i=1}^{N}(\hat{f}i - fi)^2, \tag{3.5}$$

where N is the number of subjects in each test set, $\hat{f}i$ is the reconstructed feature of subject i, and fi is the inputted feature of subject i

To visualize the pairs of FBN that presented a higher MSE_f, the generated vector containing 91 elements was reshaped into a 14x14 lower matrix for each test set. Each matrix was thresholded, by only retaining the 10% of features with higher MSE_f. Furthermore, to get the pairs of FBN that were specific to each test set, the matrix of the H-Test-U was subtracted to each of the other matrices. This matrix was

also thresholded and only maintained the 10% of features with higher values.

3.4.4 Clustering

After analysing the pairs of FBN that were abnormal for each test group, the clustering task was initiated. This task was performed in an unsupervised way. No labels were given to the model, and the inputted data contained both healthy subjects and SCZ patients from the UCLA database. The model was trained in order to find the best position for the selected number of clusters.

The fuzzy c-means algorithm was chosen to cluster the data. Fuzzy logic principles can be used to cluster data, where each point belong to a cluster centre with a percentage ranging from 0 to 100. The cluster to which the data belongs is frequently assumed to be the cluster with the highest percentage for that data point. As a result, each subject was assigned to a specific cluster. The python library *Scikit Fuzzy* was used to implement the clustering algorithm. The parameters m, error and maximiter were respectively 2, 0.005, and 1000, following the example from the documentation of *Scikit Fuzzy*. The number of clusters was defined after several tests. The difference vectors that had previously been calculated were used as input data. The clustering algorithm does not benefit from the inclusion of the 91 features as input. Starting with a small number of features and gradually increasing it is more reasonable for a clustering task. The clustering algorithm was designed with the intention of testing the normative model with the COBRE database, which only included healthy subjects and SCZ patients. As a result, the first features chosen for the cluster were those that were worse reconstructed for the SCZ patients from UCLA. On the COBRE database, it would be possible to investigate the generalization of the model. In addition, the model's performance would be improved if the clustering algorithm was trained solely using the healthy subjects and the SCZ subjects from UCLA. Thereby, the ADHD and BD patients were excluded from this task. Several features and number of clusters were tested, and the best model for distinguishing healthy subjects from SCZ patients of UCLA was selected.

3.4.5 Schizophrenia case study

To study the generalization of the normative model, an independent test set was passed through the AE. Subjects from the COBRE database consisted of 49 SCZ patients and 70 healthy subjects. The two groups were used to test the AE, by applying the same methods that were applied to the UCLA database. A comparison between the results obtained for the SCZ-Test-U and the results obtained for the SCZ-Test-C was pursued. Additionally, data was tested in the trained clustering algorithm. Each subject was assigned to the cluster with the highest percentage of membership.

Chapter 4

Results and Discussion

This chapter presents the dissertation's results and discusses their findings. Firstly, the AE optimization results are displayed. Then, the pathological differences between groups of subjects are exhibited for the UCLA database. Afterwards, a case study with the COBRE database is also presented, to study the generalization of the model.

4.1 Autoencoder optimization

AE optimization can be divided into two major parts. First, it was decided that the AE would have 3 hidden layers, and the goal was to find the best model architecture in terms of the number of neural units presented in each of the hidden layers. Secondly, the hyperparameters of the selected model architecture were tuned, to optimize the AE generalization, and performance.

The normalized mean of the median MSE_s values of the training set for the 10 folds of cross-validation (CV-MSE-Train), and the normalized difference between the mean of the median MSE_s values of the training set and the mean of the median MSE_s values of the validation set for the 10 folds of cross-validation (CV-MSE-Difference) is shown in Figure 4.1 for 14 different arrangements of the hidden layers.

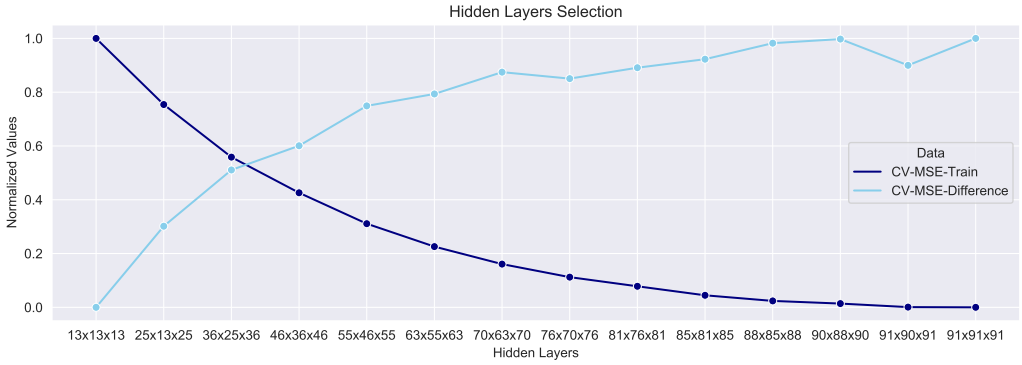


Figure 4.1: Normalized evaluation metrics (CV-MSE-Train - normalized mean of the median MSE_s values of the training set for the 10 folds of cross-validation; CV-MSE-Difference - the normalized difference between the mean of the median MSE_s values of the training set and mean of the median MSE_s values of the validation set for the 10 folds of cross-validation), for 14 different arrangements of the hidden layers.

As expected, the AE that reduces more its dimensionality is the one with the highest CV-MSE-Train and lowest CV-MSE-Difference. On contrary, the AE that does not reduce its dimensionality is the one with the lowest CV-MSE-Train and highest CV-MSE-Difference. A simple examination of Figure

4.1 demonstrates that selecting a model that is perfect in both performance and generalization when reconstructing healthy data is impossible. For developing an effective normative model, both evaluation metrics are critical. Overfitting occurs when CV-MSE-Train is the smallest and the CV-MSE-Difference is the largest, since the model is very good at reconstructing data from the training set but fails on the validation set. Underfitting occurs at the other extreme, where the CV-MSE-Train is the largest and the CV-MSE-Difference is the smallest, since the model is bad at reconstructing data from the training set, and the reconstruction error for the validation set is very similar to the reconstruction error of the training set. Given this, the best model would strike a balance between the two evaluation metrics. An efficient normative model must fail to reconstruct pathological data, when compared with healthy data from a test set. If the model is not generalizable it will fail in both test sets from pathological and healthy subjects. If the model can not reconstruct healthy data from the training set, it won't be able to learn healthy characteristics, and it will fail in both pathological and healthy test sets. Taking this into account, the model with the hidden layers 36x25x36 is the one that best balances both evaluation metrics. However, it would be interesting to test some other networks. The networks that were close to the network 36x25x36 were 25x13x25 and 46x36x46. As a result, some other architectures combining those networks were investigated.

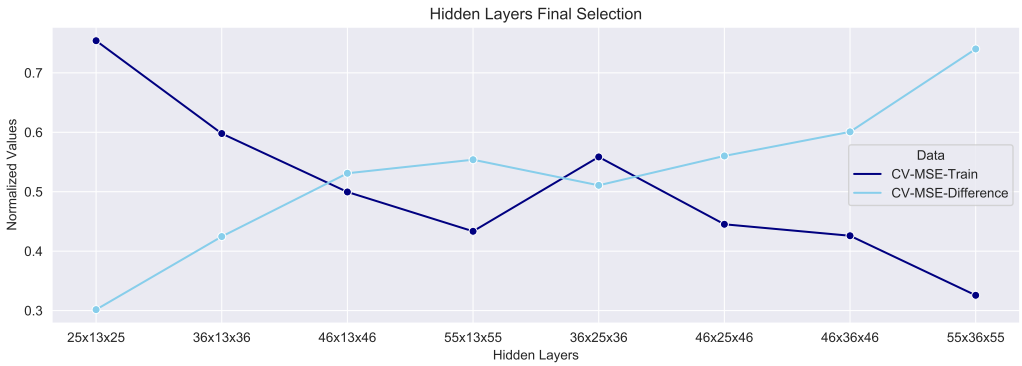


Figure 4.2: Normalized evaluation metrics (CV-MSE-Train - normalized mean of the median MSE_s values of the training set for the 10 folds of cross-validation; CV-MSE-Difference - the normalized difference between the mean of the median MSE_s values of the training set and mean of the median MSE_s values of the validation set for the 10 folds of cross-validation), for networks that combine the hidden layer architecture of networks 25x13x25, 36x25x36 and 46x36x46

Figure 4.2 displays the evaluation metrics for those networks. The network with the hidden layers 46x13x46 was chosen as the best architecture for the AE since it is the one that better balances both metrics.

After defining the layers of the network, several hyperparameters were changed. Table 4.1 shows the hyperparameters that were selected for the different networks. The hyperparameters that were tuned were the number of epochs, the learning rate for Adam, the activation function, the dropout parameters and the L2 parameter.

Network 1 is the standard network, and its hyperparameters are those that were used for the previous selection of the hidden layers. The evaluation parameters of network 2 are both higher when compared with network 1. As a consequence, Leaky ReLu was decided to be the activation function. For the learning rate, 0.0005 is still the best option, because 0.001 and 0.0001 cause an unbalancing of the evaluation metrics. Regarding the number of epochs and the L2 parameter, network 7 appears to be the best option. The dropout value of 0.25 is not beneficial since it also unbalances the evaluation metrics. In the end, network 7 was the selected AE.

Table 4.1: Results of the second step of optimization, which was performed by changing several hyperparameters

Network	Epochs	Learning Rate	Activation Function	Dropout Parameter	L2 Parameter	CV-MSE-Train	CV-MSE-Difference
1	1000	0.0005	Leaky ReLu	0.5	10^{-6}	0.4996	0.5311
2	1000	0.0005	ReLu	0.5	10^{-6}	0.6970	0.5531
3	1000	0.0001	Leaky ReLu	0.5	10^{-6}	0.6692	0.3802
4	1000	0.001	Leaky ReLu	0.5	10^{-6}	0.4787	0.6058
5	1000	0.0005	Leaky ReLu	0.5	10^{-4}	0.5071	0.5209
6	1000	0.0005	Leaky ReLu	0.5	10^{-5}	0.5025	0.5247
7	1000	0.0005	Leaky ReLu	0.5	10^{-7}	0.5025	0.5158
8	500	0.0005	Leaky ReLu	0.5	10^{-6}	0.5524	0.5222
9	500	0.0005	Leaky ReLu	0.5	10^{-5}	0.5553	0.5171
10	500	0.0005	Leaky ReLu	0.5	10^{-7}	0.5518	0.5272
11	1000	0.0005	Leaky ReLu	0.25	10^{-6}	0.2657	0.7516
12	1000	0.0005	Leaky ReLu	0.25	10^{-5}	0.2695	0.7288
13	1000	0.0005	Leaky ReLu	0.25	10^{-7}	0.2657	0.7554

4.2 Testing the Autoencoder

The UCLA database was used to test the trained AE. The 91-element vector of each subject was fed into the model, which produced a reconstructed vector.

Figure 4.3 represents the data reconstruction of 5 subjects from different set groups. Notably, the reconstructed data is smoother than the inputted data for all subjects. However, the pattern seems to be maintained for all subjects, which demonstrates that the AE can reconstruct the data. Visually, it is not possible to establish if a particular set group is worse reconstructed than the others because only 1 subject of each set group is presented. Therefore, the MSE_s was calculated, and boxplots were generated to perform a group analysis.

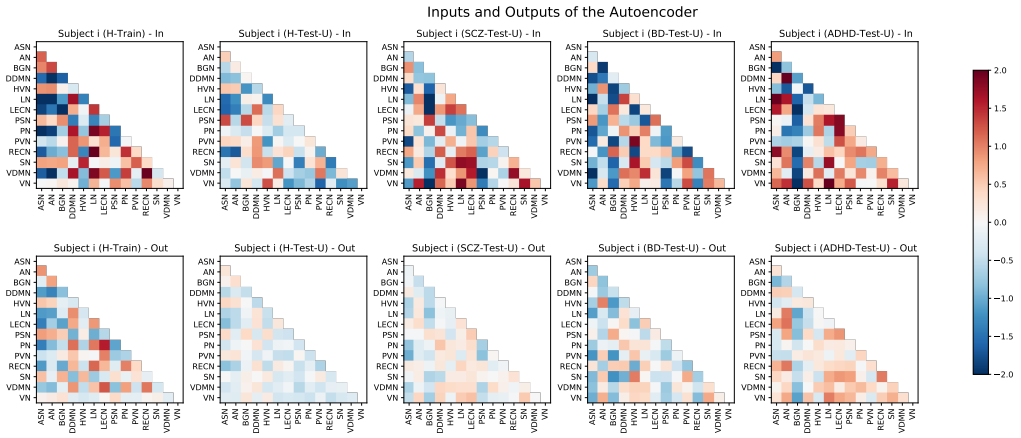


Figure 4.3: Example of the input standardized vector reshaped to a lower triangular matrix, and the reconstructed vector reshaped to a lower-triangular matrix. Each matrix represents the correlation coefficient between pairs of FBN.

Figure 4.4 exhibits the boxplots for each group of subjects. It is evident that the median MSE_s value of the H-Train is the lowest, as expected. The boxplot of the H-Test-U is similar to the boxplot of the H-Train, which demonstrates that the model is generalizable and that overfitting does not occur. The median MSE_s values of the SCZ-Test-U, BD-Test-U, and ADHD-Test-U are higher than the median MSE_s value of H-Test-U. Then, it is possible to conclude that the AE is able to distinguish the group of healthy subjects from the groups of pathological patients, demonstrating that the model learned healthy characteristics. Nonetheless, it is also notorious that in all groups of subjects, there are subjects with higher and lower MSE_s values, demonstrating that in some cases the model reconstructed better data from pathological patients than data from healthy subjects.

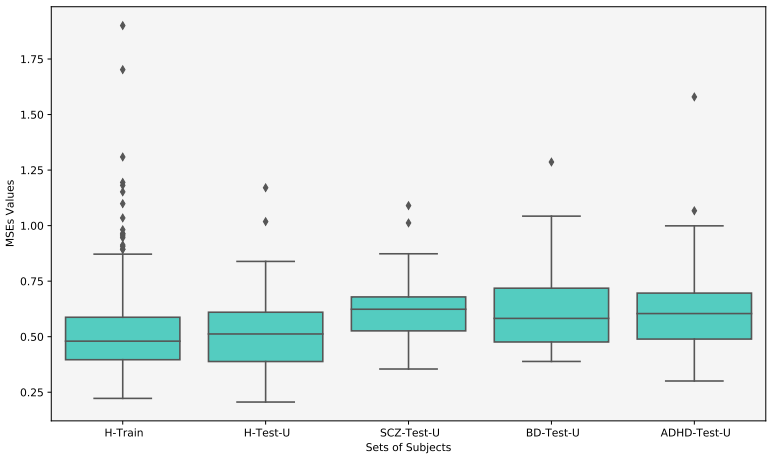


Figure 4.4: Boxplots presenting the MSE_s values of all subjects, for each group

The Shapiro-Wilk test rejected the null hypothesis that the data was normal distributed for all groups of subjects presented in Figure 4.4. Thus, Mann-Whitney was the selected nonparametric test to evaluate if the differences between the groups were statistically significant.

Table 4.2: Mann-Whitney test between the MSE_s values data of the groups of subjects

Group I	Group II	p-value
H-Train	H-Test-U	0.258
H-Train	SCZ-Test-U	1.06×10^{-6}
H-Train	BD-Test-U	1.59×10^{-5}
H-Train	ADHD-Test-U	1.71×10^{-5}
H-Test-U	SCZ-Test-U	0.002
H-Test-U	BD-Test-U	0.011
H-Test-U	ADHD-Test-U	0.007
SCZ-Test-U	BD-Test-U	0.339
SCZ-Test-U	ADHD-Test-U	0.452
BD-Test-U	ADHD-Test-U	0.373

Considering a p-value=0.05, it is noticeable that the differences between the groups of patients and the healthy subjects are statistically significant. On contrary, the null hypothesis is not rejected for the statistical test between the two groups of healthy subjects. The same is observed for the groups of patients, which was also expected. These findings show that the normative model was able to learn healthy characteristics and distinguish patients from healthy individuals.

Graph theory was also applied to perform group analysis, by transforming the difference vectors into binarized matrices. Figure 4.5 shows a boxplot containing the difference values of all subjects used to test the AE. Binarization of values higher than the third quartile value was considered since it would represent about 25% of the higher values. Thus, the selected threshold was 1. The elements of the matrix that were retained represent the pairs of FBN that were worse reconstructed, which are those that are probably functionally abnormal.

Figure 4.6 shows the number of FBN distributed by the median number of degrees. It is observed that the sets from the patients are right-shifted when compared with the healthy subjects. Considering that the number of degrees represents the number of connections to a node, the SCZ, BD, and ADHD

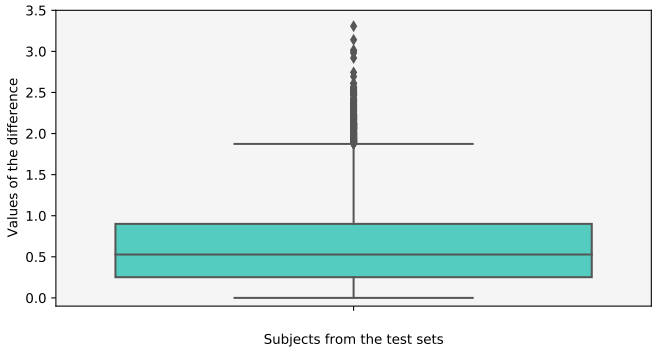


Figure 4.5: Boxplot with the difference values of all subjects used to test the AE.

patients have more FBN with more connections. Actually, this means that there are more pairs of FBN worse reconstructed for the test sets of patients.

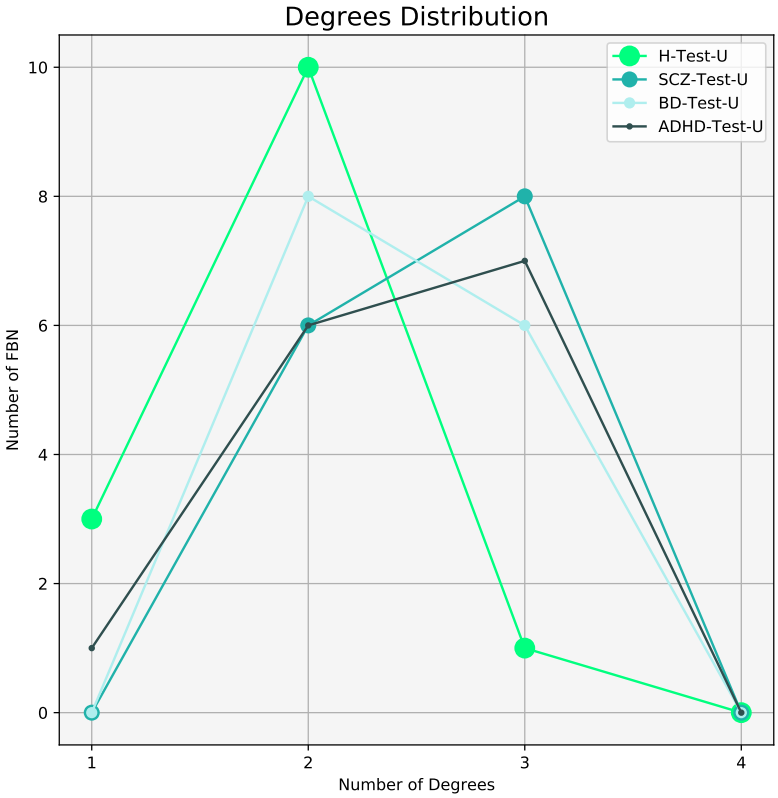


Figure 4.6: Distribution of the number of FBN for the number of degrees, for the H-Test-U, SCZ-Test-U, BD-Test-U, and ADHD-Test-U groups

The FBN median number of degrees were also regionally analysed. Figure 4.7 displays the median number of degrees for each FBN. The FBN that contain a higher number of degrees are those that were worse reconstructed. Considering that some FBN have more degrees than others, not all FBN are reconstructed in the same way. The ASN, HVN, LN, RECN, VDMN and VN are the more abnormal FBN in the ADHD group. The ASN, BGN, DDMN, HVN, LN, and PSN are the more atypical FBN

in the BD group. The ASN, BGN, HVN, LN, LECN, PSN, RECN and VDMN are the more divergent FBN in the SCZ group. The results demonstrate that different NPDs have different characteristics and that the normative model can determine the FBN that are more abnormal for each group. Nevertheless, even the group of healthy subjects present some FBN that are worse reconstructed than others. This may be justified by the fact that some FBN are more stable among healthy subjects than others, and that their functional connectivity was learned easier by the normative model. Even though, it must be noted that the threshold influences the results. Therefore, since the method for determining the threshold was not very precise, caution should be taken when analysing the FBN that appear to be worse reconstructed for each group.

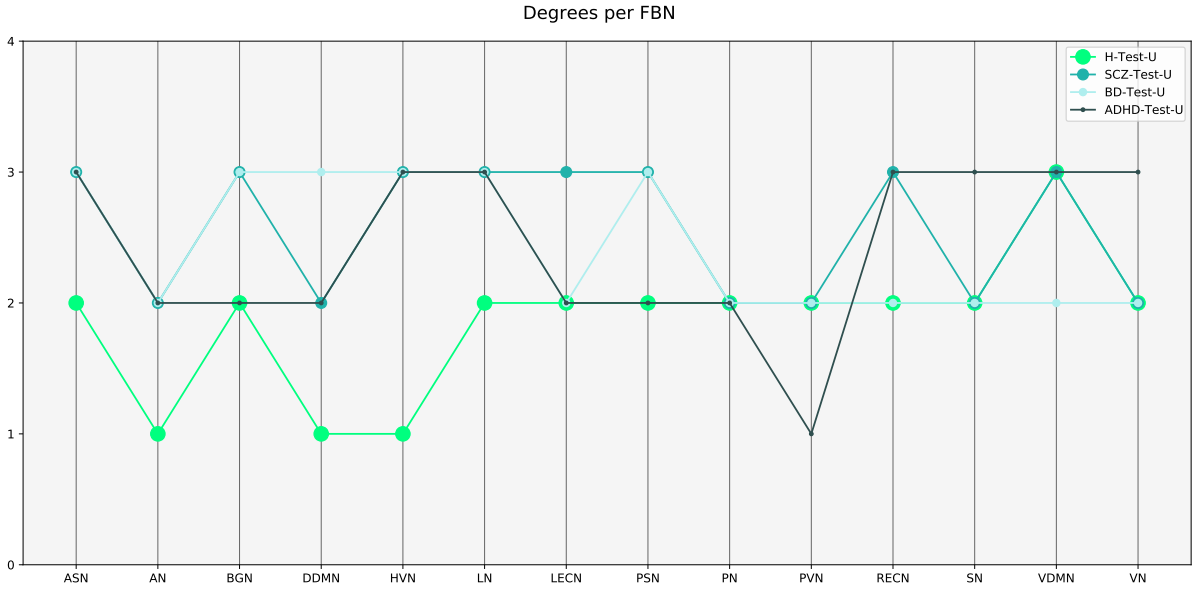


Figure 4.7: Representation of the median number of degrees for each FBN, for the H-Test-U, SCZ-Test-U, BD-Test-U, and ADHD-Test-U groups

Afterwards, the MSE_f was calculated for each pair of FBN, and the resulting matrix for each set of subjects is presented in Figure 4.8. The results support the perception that the normative model was not able to reconstruct all pairs of FBN equally. Actually, there are some FBN that appear to be more difficult to reconstruct than others. The pairs LN-HVN, SN-LN, and VDMN-PN are worse reconstructed for the H-Test-U, and are also highlighted for the other three test sets. This probably means that the functional connectivity of those pairs of FBN varies even among healthy subjects. Some FBN present more abnormal connectivity, than others, for each test set. This is consistent with the results from graph theory analysis, since most of the pairs of FBN that are highlighted in Figure 4.8 involve the FBN that were considered to be atypical following graph theory analysis for each test set [Figure 4.7]. For ADHD, several pairs that were highlighted involved the ASN. For BD, the BGN and the LN are the FBN that are more evident. For the SCZ, the BGN, HVN and LN are the FBN that are more noticeable, which corroborates some of the findings of graph theory analysis.

Considering that the H-Test-U matrix also contains elements with a high MSE_f, it was considered important to subtract the MSE_f matrix of the healthy subjects to each of the other MSE_f matrices from the groups of patients. Therefore, the regions that are highlighted in Figure 4.9 are specific of that group of patients, since a group of healthy subjects from the same database was used to exclude pairs of FBN that are worse reconstructed for both healthy subjects and patients.

Some pairs of FBN have been smoothed after subtracting the MSE_f matrix from the H-Test-U, such

4. RESULTS AND DISCUSSION

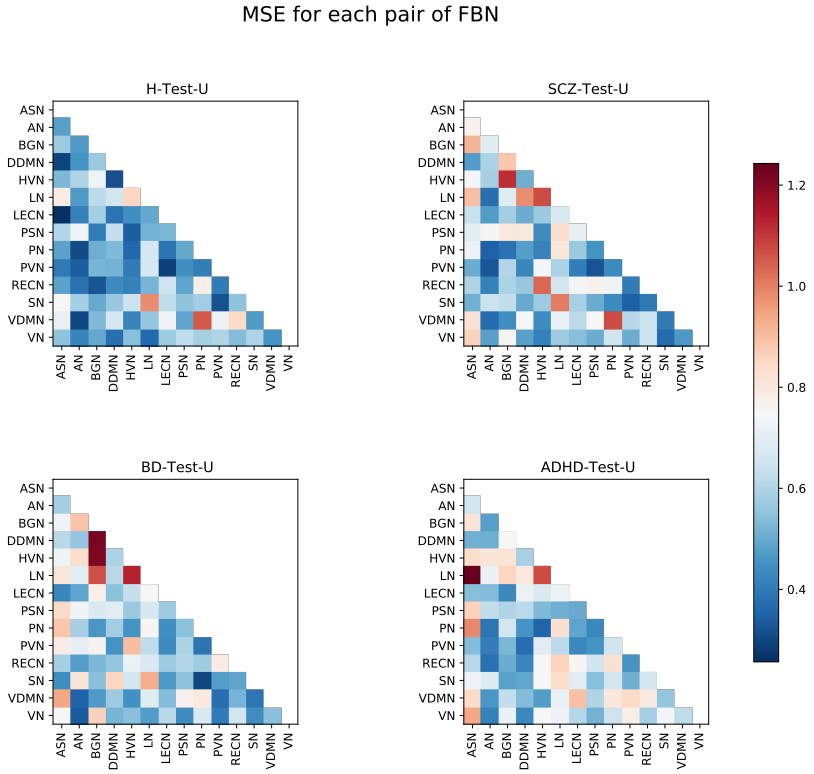


Figure 4.8: Matrices representing the MSE_f for each pair of FBN, for the four test sets. The colorbar ranges between the lower value and the higher value from the MSE_f for each pair of FBN of all test sets.

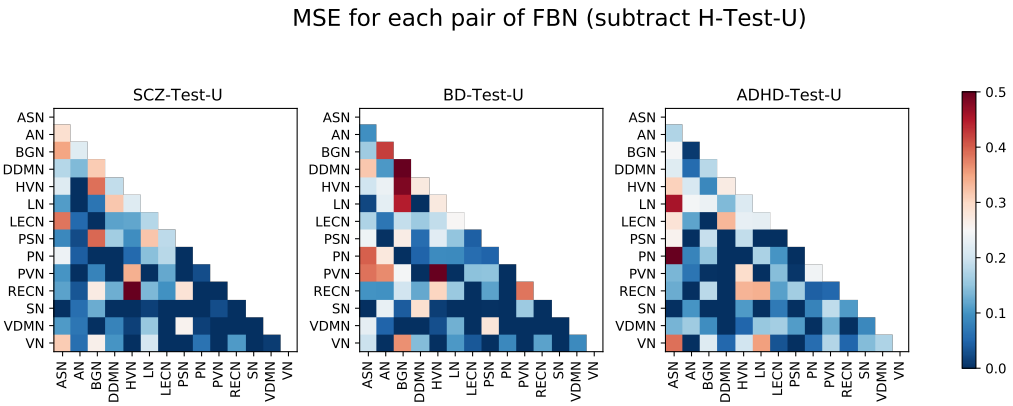


Figure 4.9: SCZ-Test-U, BD-Test-U, and ADHD-Test-U matrices represented in Figure 4.8, subtracted by the H-Test-U matrix.

as the LN-SN and the PN-VDMN. Both features were emphasized on Figure 4.8 for all test sets, in particular for the SCZ-Test-U, and were not significant on Figure 4.9. On contrary, some pairs of FBN were still highlighted after that calculation, such as the pairs BGN-DDMN, BGN-HVN and BGN-LN for the BD-Test-U.

To analyze those results in detail, Figure 4.10 was displayed. It shows the 10% of worst reconstructed pairs of FBN for each test, before and after subtracting the MSE_f matrix of healthy subjects. For each test group, those pairs of FBN that are highlighted in both situations should be considered atypical FBN for that group of neuropsychiatric patients, since they were easy to reproduce for the healthy subjects. The pairs of FBN that were considered to be atypical before subtracting the MSE_f matrix of healthy subjects, and were not present after subtraction, are probably the pairs of FBN that have not a stable

connectivity pattern even among healthy individuals. Thus, although the connectivity between those pairs of FBN seems to be atypical for the patient's group, they may not be characteristic of that NPD since the normative model was not even able to reconstruct them well for the group of healthy subjects. However, it is also possible that those pairs of FBN have been badly reconstructed for that specific group of healthy subjects due to some other reason, and are still specific of the group of patients. Regarding those pairs of FBN, the main finding is that they are not useful to distinguish the group of patients from the group of healthy subjects that were tested in the AE. Finally, the pairs of FBN that were not highlighted before the subtraction, and were present after, are those that were easy to learn by the normative model, indicating that they are stable among healthy people, but are probably abnormal in the group of patients.

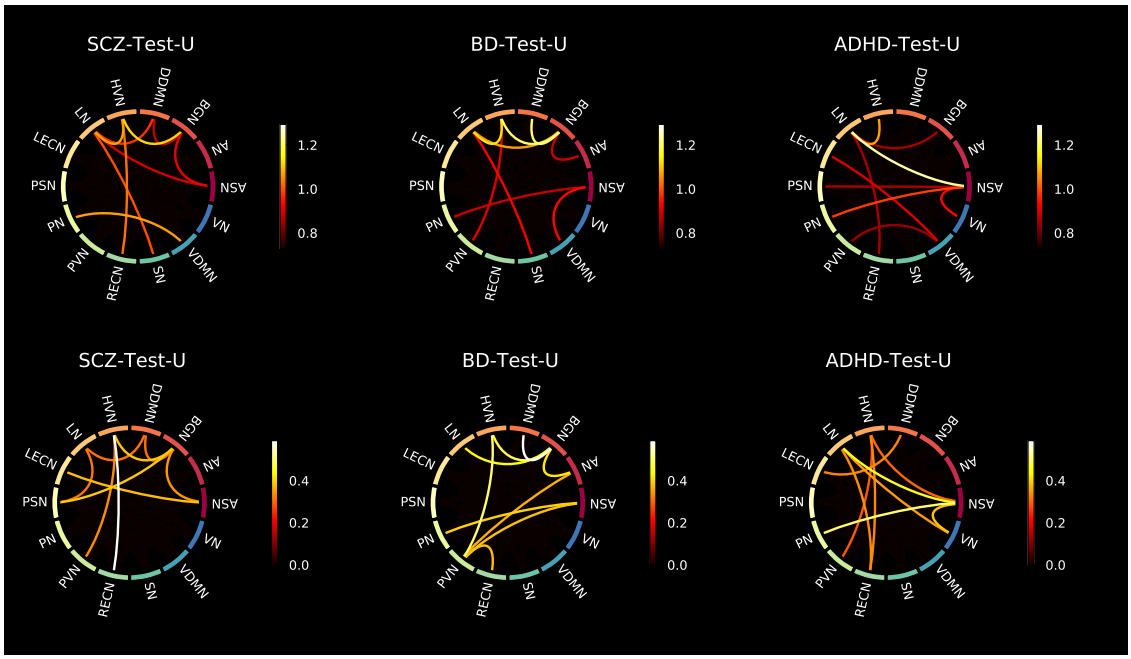


Figure 4.10: Connectogram exposing the 10% of pairs of FBN that were worse reconstructed for each group without (top) and with (bottom) subtraction of the H-Test-U MSE_f matrix

For the SCZ-Test-U, the correlations between the ASN-BGN, BGN-DDMN, BGN-HVN, DDMN-LN, and HVN-RECN are pairs of FBN that characterize the group, since they are present before and after subtraction of the MSE_f matrix of the healthy subjects. The ASN-LECN, BGN-PSN, LN-PSN, and HVN-PVN pairs of FBN also characterize the SCZ group. However, the MSE_f of those pairs of FBN was low for the healthy group. This suggests that the connectivity pattern between those FBN is important to characterize the healthy pattern, and the normative model really tried to minimise the reconstruction error of those features. Thus, for the SCZ-Test-U the reconstruction error of those pairs of FBN was not very high, because the normative model strives to reconstruct those features well. Nevertheless, when the reconstruction error of those features for the SCZ group is compared to the reconstruction error of the H-Test-U group, the conclusion is that those features possibly characterize the SCZ-Test-U. On the contrary, the PN-VDMN, LN-SN, ASN-LN, and HVN-LN may not characterize the SCZ-Test-U, because the normative model was also not able to learn to reconstruct those features well for the healthy individuals, meaning that those pairs of FBN are not decisive to characterize the healthy pattern. Thus, even though they are not well reconstructed for the SCZ group, that does not suggest that those pairs of FBN are important to characterize SCZ.

According to the literature, default mode and salience networks appear to play a role in SCZ, which is consistent with the findings of this study, especially given the FBN match previously reported in Section

2.5. Considering that the anterior and posterior insula are located in the ASN, and PSN, respectively, that corpus striatum belongs to basal ganglia, that the LN is made up of middle temporal regions, and that the LECN combine regions of the temporal and prefrontal lobes, the pairs of FBN corresponding to the ASN-BGN, BGN-PSN, LN-PSN, ASN-LECN that were highlighted in the SCZ-Test-U follow the findings of [142], previously refereed in Section 2.5. The role of basal ganglia in SCZ has been also previously reported [143], which is also found in the results of this study. The abnormal connectivity between the HVN and the RECN was also reported in previous studies [144], and that pair of FBN was the most abnormal in this group of SCZ patients.

For the BD-Test-U, the AN-BGN, BGN-DDMN, BGN-HVN, BGN-LN, ASN-PN, HVN-PVN pairs of FBN were maintained before and after subtraction of the MSE_f H-Test-U matrix, and characterize the BD group. The ASN-PVN, AN-PVN, and PVN-RECN are also useful to differentiate BD patients from healthy subjects since they were highlighted after subtracting the matrix of the healthy test subjects. On the contrary, the HVN-LN, ASN-VDMN, and LN-SN are not useful to distinguish the BD-Test-U from the H-test-U since they are also badly reconstructed for healthy subjects. The role of the BGN is evident from the results, which is in agreement with some studies in the literature. Actually, it has been previously suggested that basal ganglia may play an important role in the dysfunctional emotional processing and regulation of BD patients [132]. Nevertheless, whereas in this study the BGN is highlighted as the FBN that is more functionally abnormal in BD, in the literature is has been demonstrated that some other FBN appear to be more relevant in BD, such as the default mode, central executive and salience networks [131]. Although the ASN also appears to be relevant for the neuropathology of BD, the executive control networks do not seem to be that important. There are parts of the default mode network (DDMN, PN, and LN) that seem to be abnormal in BD, but they do not seem to play a role as relevant as the BGN. Unfortunately, the limbic and cerebellum networks were not included in this study, and they also seem to play a role in BD according to the literature [134, 132]. The thalamus was also previously considered to play a role in BD, which may justify the atypical functional connectivity of the BGN since it includes thalamic regions [134, 135, 136]. The role of the visual regions, which include the PVN and the HVN, also characterize BD. Although visual regions have been previously reported to be related to the neuropathology of BD, not much focus has been given to the FBN related to visual processes [132].

The ASN-LN, ASN-PN, ASN-VN, and LN-RECN pairs of FBN are representative in the ADHD-Test-U. In addition, the ASN-HVN, DDMN-LECN, HVN-PVN, HVN-RECN, and LN-VN also seem to characterize the ADHD group, because their reconstruction error is distinguishable from the reconstruction error of the healthy subjects. Differently, the BGN-LN, LN-HVN, ASN-PSN, LECN-VDMN, and PVN-VDMN pairs of FBN are not useful to distinguish the ADHD group from the healthy group, and it is not possible to conclude that their functional connectivity pattern is useful to characterize the ADHD group. The pair ASN-PN presents the highest MSE_f among the ADHD group, and in a previous study, the link between the dorsal anterior cingulate to the precuneus was considered a possible candidate locus for the dysfunction of ADHD [122]. The interplay of the default mode, executive control and attention networks [121] in ADHD was also observed in this dissertation, considering that the LECN and the RECN make up the executive control networks, the ASN and the VN are the attention networks, and the DDMN and the LN belong to the default mode network. All those FBN were found to be involved in alteration of the functional connectivity pattern, characteristic of ADHD. Additionally, the visual networks (HVN, and PVN) also seem to be involved in the neuropathology of ADHD. This is in line with a recent study [128], that highlighted the role of somatosensory FBN in ADHD. The language alterations reported in ADHD patients are also in line with the results since the LN seems to be functional abnormal in ADHD [129].

Several results are in agreement with the literature, which demonstrates the ability of the normative model on identifying functional connectivity patterns of the several NPDs.

It is interesting to notice that the correlation ASN-LN was highlighted for the ADHD-Test-U before and after subtraction of the MSE_f matrix of the H-Test-U, but was excluded for the SCZ-Test-U after subtraction of the matrix of the H-Test-U. By visualizing Figure 4.8, it is noticeable that the MSE_f of that pair of FBN is one of the highest for the H-Test-U, and that is why it is smoothed for the SCZ-Test-U after subtraction of the healthy matrix. For the ADHD-Test-U, that feature is maintained after subtraction because the MSE_f is still really high, even when compared with the healthy group. Therefore, although the connectivity between those FBN may not be very stable among healthy individuals, it seems to be even more atypical for the ADHD patients. Nevertheless, care should be taken on concluding that this abnormality is specific to the ADHD group. Similar reasoning must be performed about the correlation BGN-LN, since it was highlighted on the BD-Test-U before and after subtraction of the matrix of the H-Test-U, but it was excluded for the ADHD-Test-U after subtraction of the matrix of the H-Test-U. Although the reasons are similar to those enunciated for the ASN-LN pair, in this situation the MSE_f of that feature is not so high for the H-Test-U. Additionally, the MSE_f of that feature is also not so high on the ADHD-Test-U. Thus, it is more secure to conclude that the BGN-LN connectivity can characterize BD, than to enunciate that the ASN-LN characterizes ADHD. Nonetheless, it must be remembered that the H-Test-U represents a specific group of healthy subjects and that some pairs of FBN that were smoothed after subtraction of the MSE_f H-Test-U matrix may still be characteristic of the disorder, while others that appear after subtraction may not.

Furthermore, particular attention should be given to the MSE_f values of the pairs of FBN that are considered to characterize each group of patients. Considering that the matrices were thresholded to maintain the 10% of worst reconstructed features, rather than defining a threshold value, the MSE_f values of the remained pairs of FBN is not the same. Thus, some pairs of FBN should be considered more representative of a particular group than others.

4.3 Clustering

As previously enunciated, only the group of healthy subjects and SCZ patients from UCLA were used to train the model for the clustering task. Considering the goal of distinguishing SCZ patients from healthy subjects, the clustering task was performed by testing several combinations of the features that appeared to be more effective at distinguishing healthy individuals from SCZ patients in the UCLA database. It is more reasonable for a clustering task to start with a few number of features, and increase it step by step. Therefore, several cluster models were tested using combinations of the features: 49 - HVN-RECN, 23 - BGN-PSN, 8 - BGN-HVN, 15 - ASN-LECN, 1 - ASN-BGN, 40 - HVN-PVN, 26 - LN-PSN, 13 - DDMN-LN, and 5 - BGN-DDMN. Those features are ordered from the higher to the lower MSE_f. Here are presented some of the attempts and respective results. Figure 4.11 shows the results for combinations of features 49, 23, and 8, for two, three, four and five clusters.

The findings appear to show that using four clusters to distinguish healthy subjects from SCZ patients from the UCLA database produces better results. Therefore, more combinations of features were tested, while selecting four clusters [Figure 4.12].

Other combinations could have been tried, and a different clustering model could have been used. However, the combination of the features 5 - BGN-DDMN, 8 - BGN-HVN, 23 - BGN-PSN, and 49 - HVN-RECN with four clusters seems to lead to the best result on discriminating the SCZ-Test-U from the H-Test-U. This clustering task is quite difficult, and optimizing it would go beyond the scope of this

investigation. The initial objectives for this dissertation were thought to be met by a simple application of the clustering algorithm.

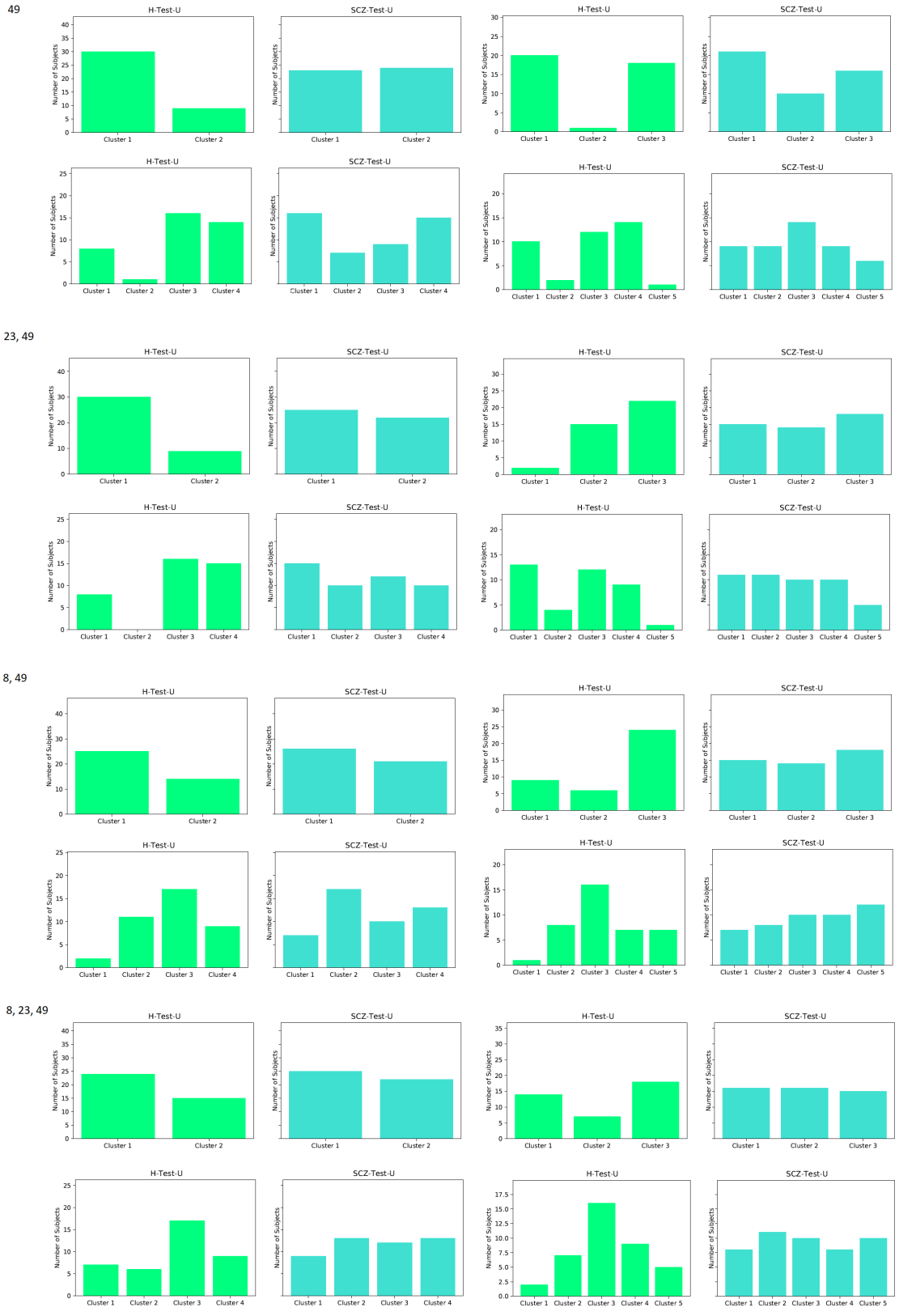


Figure 4.11: Results of several attempts of clustering for two, three and five clusters. Note: 49 - HVN-RECN, 23 - BGN-PSN, 8 - BGN-HVN. Where H-Test-U represents the group of healthy subjects from the UCLA database, and SCZ-Test-U represents the number of SCZ patients from the UCLA database.

Figure 4.12: Results of several attempts of clustering for 4 clusters. Note: 49 - HVN-RECN, 23 - BGN-PSN, 8 - BGN-HVN, 15 - ASN-LECN, 1 - ASN-BGN, 40 - HVN-PVN, 26 - LN-PSN, 13 - DDMN-LN, and 5 - BGN-DDMN. Where H-Test-U represents the group of healthy subjects from the UCLA database, and SCZ-Test-U represents the number of SCZ patients from the UCLA database.

4.4 Schizophrenia Case Study

The trained AE was tested in an independent test set. The healthy subjects and the SCZ patients from COBRE were used to evaluate the generalization of the model. Therefore, in order to compare the results, those are presented in combination with the results from the H-Test-U, and SCZ-Test-U.

Figure 4.13 displays the boxplots of the MSE_s values. The differences between the boxplots of H-Test-C, and SCZ-Test-C are not as notorious as those from the UCLA database. Nevertheless, the SCZ-Test-C boxplot is shifted up when compared to the the H-Test-C boxplot, which may indicate that the normative model has more difficulty in reconstructing data from the SCZ patients.

For the COBRE database, the Shapiro-Wilk test also rejected the null hypothesis that the data was normal distributed. Thus, the Mann-Whitney test was also applied [Table 4.3].

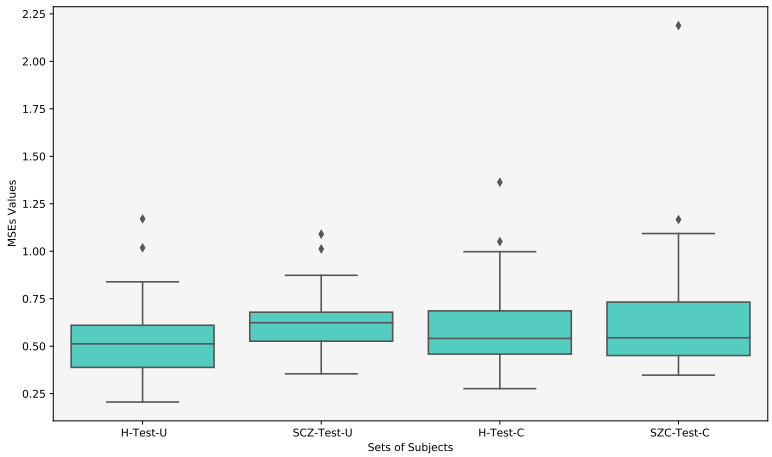


Figure 4.13: Boxplots presenting the MSE_s values of all subjects, for healthy and SCZ subjects from the UCLA and COBRE database

Table 4.3: Mann-Whitney test between the MSE_s values data of the groups of subjects, for the COBRE database

Group I	Group II	p-value
H-Test-U	H-Test-C	0.100
H-Test-U	SCZ-Test-C	0.049
H-Test-C	SCZ-Test-C	0.324

Considering a p-value=0.05, the null hypothesis is not rejected for the comparison between the groups of healthy subjects. On contrary, the differences between the SCZ patients from the COBRE database and the healthy subjects from UCLA database are statistically significant. However, when comparing the both groups of the COBRE database, the differences between the SCZ patients and the healthy individuals are not statistically significant. This suggests that the model is generalizable, but that the model's ability to distinguish healthy individuals from patients in an independent database is less accurate.

Figure 4.14 shows the distribution of the number of regions by the median number of degrees is similar for both sets. It seems that the H-Test-C contains less abnormal FBN than the SCZ-Test-C.

On the contrary, Figure 4.15 shows some distinct differences, by examining each FBN separately. Both groups of SCZ patients have similar patterns, which suggests that the model is able to characterize SCZ. On contrary, there are few similarities between the patterns of both groups of healthy subjects.

Analysing the connectivity between the pairs of FBN, it is possible to deeper explore those findings. Figure 4.16 shows that in general the pairs of FBN that were worse reconstructed for the H-Test-C have a lower reconstruction error than those from the SCZ-Test-C and SCZ-Test-U. In addition,both SCZ groups have a similar pattern, with high reconstruction error for many of the same pairs of FBN. This indicates that the normative model was successful in identifying characteristics of SCZ. Differently, the two groups of healthy subjects do not present similar patterns. This was expected since the model was learned through a healthy population. The pairs of FBN shared between those healthy groups that were poorly reconstructed may represent patterns of connectivity that are not stable in healthy individuals (eg. the ASN-LN, and the PN-VDMN). Other pairs of FBN that appear in one of the healthy groups but not the other could be associated with the specific dataset. For example, the pair LECN-RECN seems to be unique to the COBRE database, as it is highlighted in both COBRE groups, but not in any of the other

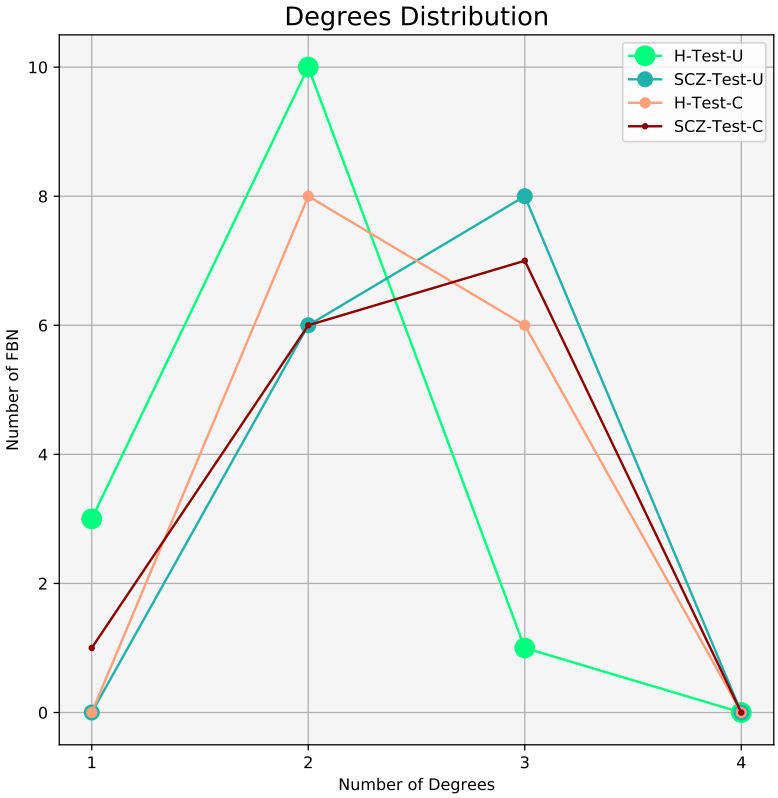


Figure 4.14: Distribution of the number of FBN for the number of degrees, for the H-Test-C, SCZ-Test-C groups

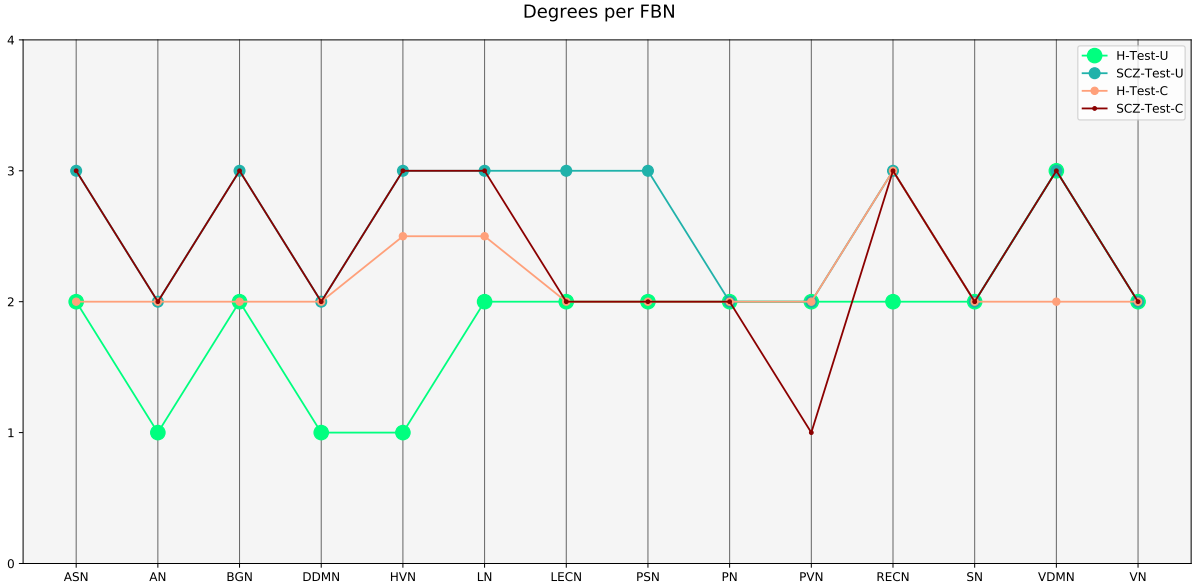


Figure 4.15: Representation of the median number of degrees for each FBN, for the H-Test-U, SCZ-Test-U, H-Test-C, SCZ-Test-C groups

test groups.

It should be noted that the H-Test-U MSE_f matrix was not subtracted to the SCZ-Test-C. This step was done for the UCLA sets since the goal was to figure out which pairs of FBN could be used to

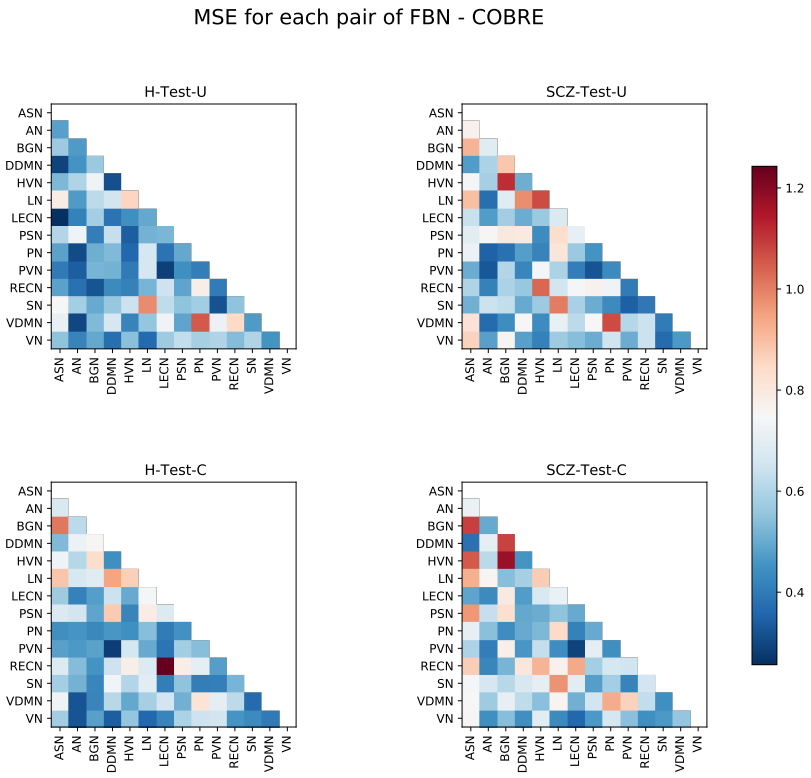


Figure 4.16: Matrices representing the MSE_f for each pair of FBN, for the healthy (left) and SCZ (right) test sets of the UCLA (top) and COBRE (bottom) databases.

distinguish those test sets. The data from COBRE is used as an independent test set. Thus, there is no benefit in subtracting the H-Test-U MSE_f matrix to the data from COBRE, since it would not lead to significant findings.

Figure 4.17 shows the connectograms of the SCZ patients from UCLA and COBRE. From the nine highlighted pairs of FBN, six are shared between the two groups. However, it must be observed that some of those shared pairs of FBN, were smoothed after subtracting the H-Test-U MSE_f matrix to the SCZ-Test-U MSE_f matrix. Nonetheless, this suggests again that the normative model was able to identify SCZ patterns.

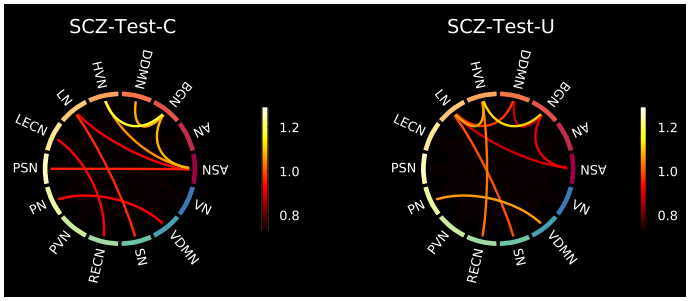


Figure 4.17: Connectogram exposing the 10% of pairs of FBN that were worse reconstructed for the SCZ patients of the COBRE (left) and UCLA (right) database

The COBRE database was then tested on the cluster that was trained with SCZ patients and healthy subjects from UCLA database. The results are presented in Figure 4.18. There are differences between

subjects, since there are subjects distributed across all clusters for both test groups. Regarding the UCLA database, it is noticeable that cluster 2 seems to be more SCZ-specific and cluster 3 appears to be characteristic of healthy subjects. Nevertheless, the clustering shows that subgroups may exist and that differentiating SCZ subjects from healthy subjects is complex. COBRE database follows a similar pattern, with several SCZ patients and few healthy subjects in cluster 2. For the COBRE database, both cluster 1 and cluster 3 seem to be characteristic of healthy subjects. Those results are appropriate since cluster 1 and cluster 3 are the two clusters with less representation of SCZ patients for both databases. Nonetheless, cluster 1 contains a small number of healthy UCLA subjects, demonstrating the divergence among healthy people. Cluster 4 does not seem to be characteristic of a particular group of subjects and supports the idea that there is high variability between subjects from the same group. Overall, the performance of the cluster model seems to be good, since it grouped healthy and SCZ subjects similarly for both databases. Those results may be justified by the fact that the features that were used for the clustering process were also poorly reconstructed for the SCZ-Test-C. In addition, by visualizing Figure 4.16, it is possible to denote that the 4 features appear to be highlighted in both SCZ matrices when compared to the healthy matrices from the same database.

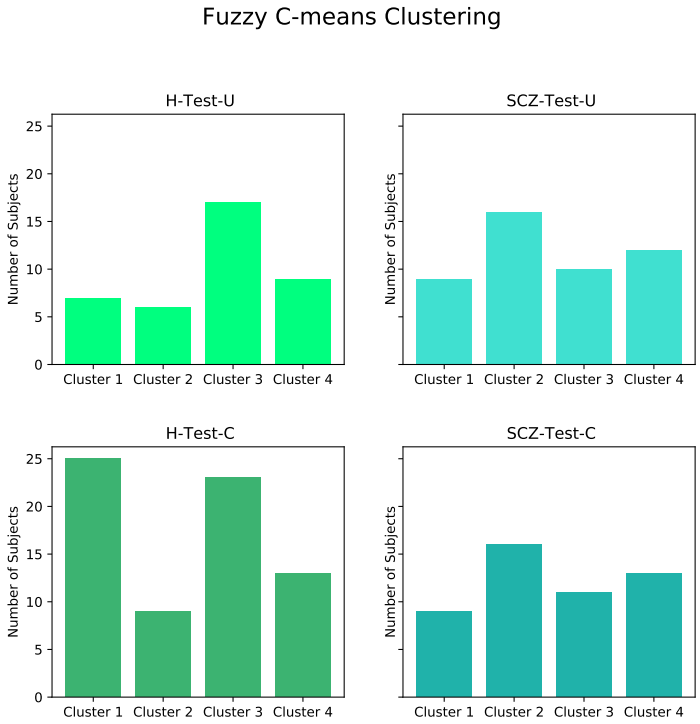


Figure 4.18: Fuzzy C-means clustering of the individual's difference vectors of the H-Test-U, SCZ-Test-U, H-Test-C, and SCZ-Test-C groups, using the features BGN-DDMN, BGN-HVN, BGN-PSN, and HVN-RECN.

Actually, those correlations of pairs of FBN that appear to be characteristic of SCZ patients are in line with previous studies, as highlighted in Section 2.5 [144, 143, 140]. Considering that the clustering task generated subgroups, those features that characterize SCZ patients were also analysed at the individual-level. The absolute values of the difference between the reconstructed and inputted data for those pairs of FBN, for each subject, are presented in Figure 4.19 and Figure 4.20, for the UCLA and COBRE databases, respectively.

The results indicate that SCZ individuals present a general higher error of reconstruction, but that the pattern is not constant among individuals. There is high heterogeneity among individuals of the same

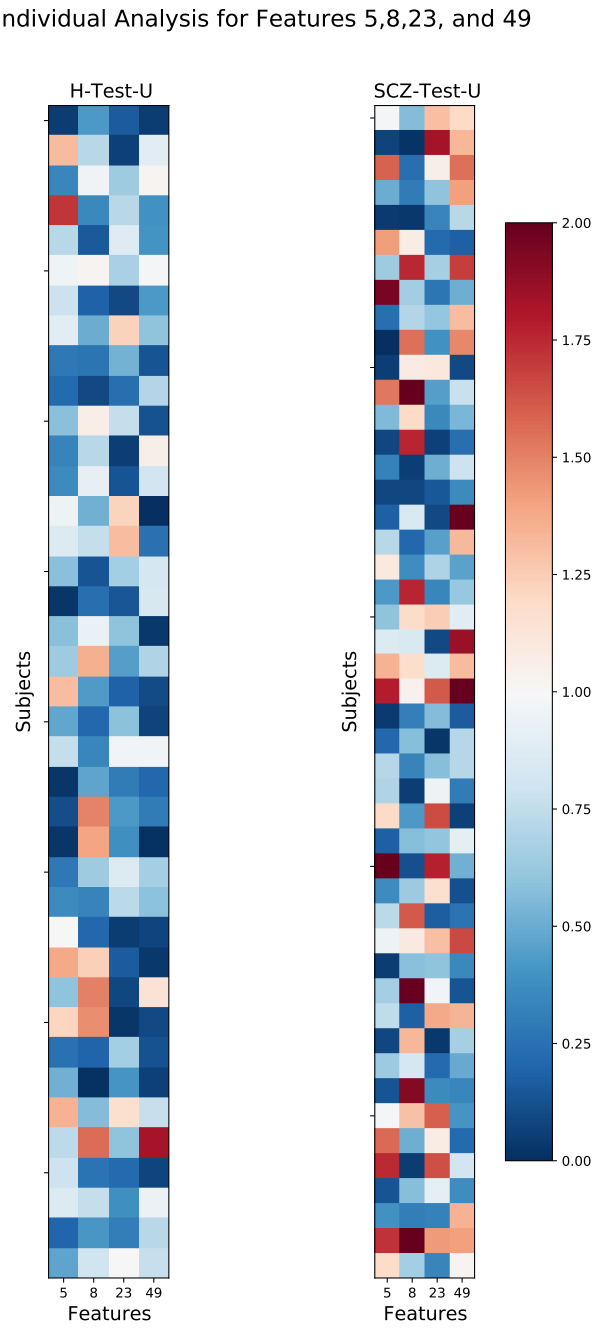


Figure 4.19: Individual-level analysis of the absolute values of the difference between the reconstructed and inputted data for features BGN-DDMN (5), BGN-HVN (8), BGN-PSN (23), and HVN-RECN (49) from subjects of the UCLA database

group, which supports the existence of subgroups. The characteristic symptoms of SCZ involve a range of cognitive, behavioural, and emotional dysfunctions. However, no single symptom is specific to the disorder. SCZ diagnosis involves the recognition of a constellation of signs and symptoms associated with occupational or social impairments. Therefore, SCZ can be considered a heterogeneous NPD, since several individuals diagnosed with SCZ will vary in most of the features that describe the disorder [1]. Consider a SCZ patient A that is mainly characterized by the presence of hallucinations and delusions and a SCZ patient B that presents disorganized thinking and negative symptoms. Assuming that those

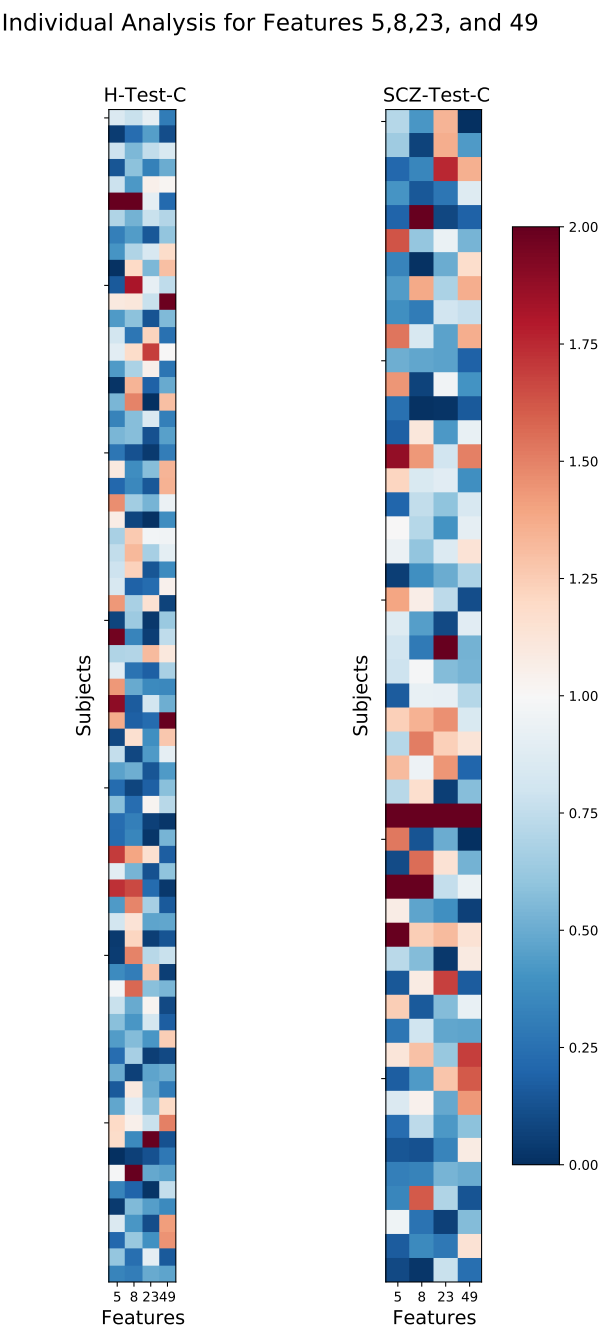


Figure 4.20: Individual-level analysis of the absolute values of the difference between the reconstructed and inputted data for features BGN-DDMN (5), BGN-HVN (8), BGN-PSN (23), and HVN-RECN (49) from subjects of the COBRE database

symptoms are in part associated with functional connectivity abnormalities, it would be expected that the functional connectivity patterns of patients A and B would be different. Thus, the results are in line with what was expected, because the functional connectivity patterns that were identified at the group-level do not resist to the individual-level analysis. The group of SCZ patients is made up of several patients A and B. Some of them present similar patterns of functional connectivity, which may be related to a similar manifestation of SCZ. Other patients are very different in terms of functional connectivity, possibly due to different manifestations of the NPD. The DSM-5 authors have already recognized SCZ as a spectrum

of NPD, but the diagnostic criteria are still too wide. Subgroups seem to still be present in patients that are diagnosed with SCZ.

Nonetheless, this hypothesis is also true for healthy subjects, since the functional connectivity pattern is also not equal for all individuals. Considering that personality traits may be related to patterns of functional connectivity and that those patterns are also dependent on gender, those results are expected [164]. Even a group of healthy subjects can be divided into several subgroups, which may be related to personality traits, age, or gender, among other factors.

The results presented in Figure 4.19 and Figure 4.20 also support the finding of the cluster algorithm [Figure 4.18]. When looking at the features at the individual-level, it is clear that the reconstruction error of some healthy individuals is higher than the reconstruction error of some SCZ patients. As consequence, it is natural that some SCZ patients have been clustered into the group that contains more healthy subjects, and that some healthy individuals have been clustered into the group that contains more SCZ patients.

Overall, those results demonstrate that the group-level differences do not resist to the individual-level analysis. Therefore, the concept of a group-pattern disguise the inter-individual differences presented in the same NPD group. This study also supports the hypothesis that the average patient concept falls apart at the individual-level in the field of neuropsychiatry [151, 152]. The findings show that in order to better understand the neuropathology of NPDs and develop effective therapeutic approaches, researchers must look beyond the average patient. Considering that NPDs develop over a continuous spectrum, a precision-based approach focused on the specific functional network changes of individual patients is promising for a translation to clinical practice.

Chapter 5

Conclusions and Future Work

Nowadays, the diagnosis of NPDs is mostly based on the DSM-5, since no ML applications have been able to face the continuous spectrum of NPDs. Neuropsychiatry constitutes a fascinating and abstract world, that is still not clearly understood. NPDs boundaries have been considered undefined, which means that several symptoms domains are shared between different NPDs, and that performing a differential diagnosis is quite complex. Herein, a DL-based normative model was created from rs-fMRI data of healthy subjects to deal with the unclear frontiers of NPDs. This approach expects the model to "fail" when tested in patients, allowing the identification of functional connectivity anomalies that characterize each NPD.

Firstly, FBN were studied in detail, and a literature review was pursued in order to define an adequate workflow and compare the dissertation results with findings previously identified in the literature.

Secondly, considering that rs-fMRI data would be used for feeding the DL-based normative model, data was carefully selected by defining particular conditions that would maximise the balance between the amount and quality of data. Therefore, we guaranteed that the impact of scan parameters would not markedly influence the results of the dissertation and that we would have enough data to train a DL model.

Thirdly, considering that the exploration of different methodologies for processing rs-fMRI is part of a different research topic, it was decided to apply well-established methodologies for processing rs-fMRI data, using FSL. Thus, the scope of this dissertation was focused on the identification of functional connectivity patterns through a DL normative model approach, instead of improving data processing methodologies.

Fourthly, the normative model was created and optimized, to guarantee the performance and the generalization of the AE. This model was tested on SCZ, ADHD, and BD patients. The results demonstrated that the functional connectivity abnormalities found at the group-level were in agreement with the literature for the three evaluated NPDs.

Lastly, the model was tested on an independent test set of SCZ and healthy subjects. It was possible to conclude that the normative model was generalizable since there were several similarities between the functional connectivity patterns identified in both SCZ groups of patients. Those findings were reinforced by the results of the clustering task, which lead to the identification of four clusters. The clustering results were also very similar for both databases. The BGN-DDMN, BGN-HVN, BGN-PSN, and HVN-RECN pairs of FBN can be considered those that best characterize the group of SCZ patients, since they were all worse reconstructed for the group of SCZ patients, and were used to feed the clustering algorithm that produced the best results. Nevertheless, it must be noticed that the group-level pattern does not resist the individual-level analysis, which suggests that there is heterogeneity within the same group. Actually,

the results of the clustering task suggested the existence of subgroups. Therefore, and considering that NPDs develop over a continuous spectrum, a precision-based medicine approach focused on the specific functional network changes of individual patients, could be more advantageous than a general group-based diagnostic classification that sometimes leads to misdiagnosis. A personalised diagnosis would allow for personalised therapy, which would improve the quality of life of the neuropsychiatric patients.

Some limitations of this study should also be refereed. First, although the parameters of acquisition that were defined led to the best balance between the amount and quality of data, longer scan duration and higher temporal resolutions are both important factors for more reliable results. Secondly, the template that was used for dual regression was not specific to the data used in this study. Nonetheless, it allowed for a comparison with the literature, and since it included more FBN than several studies using rs-fMRI in the context of NPDs, it resulted in findings that are novel to the field of functional inter-connectivity analysis. However, this study did not explore within network connectivities, which have been found to be abnormal in a number of NPDs. Incorporating those metrics could have aided in the improvement of the analysis.

Overall, this study met the objectives set forth at the outset of this dissertation. It is demonstrated that functional connectivity has the potential to be used to distinguish patients from healthy subjects and that different NPDs are characterized by different functional connectivity patterns. The DL-based normative model approach here used in this study demonstrates to be promising. It establishes a bridge between DL models and precision medicine, which may be the key for the translation of these models to clinical practice.

This study raises new research questions that could be investigated in future studies. The relation between functional connectivity abnormalities and the manifestation of the NPD was not explored. The normative model approach provides an intuitive match between the manifestations of the disorder and the brain abnormalities. By looking beyond the group-level findings and relating the clinical manifestations of the patients to the functional connectivity abnormalities discovered through the normative model, it would be possible to perform a more precise-based medicine, focused on the individual. This association could also lead to the finding of several subgroups within each NPD. Another gap that was not explored yet in the literature is the association of the common binary-based classification methods with a normative model. A pipeline that combines the accuracy of binary-classification models with the normative model's ability to perform individual-level analyses and link them to clinical manifestations would be promising for clinical practice.

References

[1] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders 5*, vol. 21. American Psychiatric Association, 2013.

[2] M. Bajouco, D. Mota, M. Coroa, S. Caldeira, V. Santos, and N. Madeira, "The quest for biomarkers in schizophrenia: From neuroimaging to machine learning," *International Journal of Clinical Neurosciences and Mental Health*, vol. 3, no. 4(Suppl. 3), 2017.

[3] R. de Filippis, E. A. Carbone, R. Gaetano, A. Bruni, V. Pugliese, C. Segura-Garcia, and P. De Fazio, "Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: A systematic review," *Neuropsychiatric Disease and Treatment*, vol. 15, pp. 1605–1627, 2019.

[4] M. S. García-Gutiérrez, F. Navarrete, F. Sala, A. Gasparyan, A. Austrich-Olivares, and J. Manzanares, "Biomarkers in psychiatry: Concept, definition, types and relevance to the clinical reality," *Frontiers in Psychiatry*, vol. 11, p. 432, 2020.

[5] A. S. David and T. Nicholson, "Are Neurological and Psychiatric Disorders Different?," *The British Journal of Psychiatry : The Journal of Mental Science*, vol. 207, no. 5, pp. 373–374, 2015.

[6] "National Institutes of Health (US); Biological Sciences Curriculum Study. NIH Curriculum Supplement Series; Bethesda (MD): National Institutes of Health (US); Information About Mental Illness And The Brain. [online] Ncbi.nlm.nih.gov. Available at: https://www.ncbi.nlm.nih.gov/books/NBK20369/ [Accessed at 19 November 2020] 2007."

[7] K. H. Taber, R. A. Hurley, and S. C. Yudofsky, "Diagnosis and treatment of neuropsychiatric disorders," *Annual Review of Medicine*, vol. 61, pp. 121–133, 2010.

[8] K. Miyoshi, Y. Morimura, and K. Maeda, eds., *Neuropsychiatric disorders*. Springer Science & Business Media, 2010.

[9] M. Shenton and B. Turetsky, eds., *Understanding neuropsychiatric disorders: Insights from neuroimaging*. Cambridge University Press, 2010.

[10] P. Sachdev, "International neuropsychiatric association," *Neuropsychiatric Disease and Treatment*, vol. 1, no. 3, pp. 191–192, 2005.

[11] V. L. Feigin, E. Nichols, T. Alam, (...), and T. Vos, "Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the global burden of disease Study 2016," *The Lancet Neurology*, vol. 18, no. 5, pp. 459–480, 2019.

REFERENCES

[12] H. Ritchie and M. Roser, "Mental Health. [online] Our World in Data. Available at: https://ourworldindata.org/mental-health [Accessed at 19 November 2020]," 2018.

[13] "Reports and Data. Neurodiagnostics Market Share, Growth — Industry Analysis [online] Available at: https://www.reportsanddata.com/report-detail/neurodiagnostics-market [Accessed at January 2021] 2020."

[14] "Market Data Forecast. Behaviour/Mental Health Software Market Size — 2020 - 2025. [online] Available at: https://www.marketdataforecast.com/market-reports/mental-health-software-market [Accessed at 19 November 2020] 2020."

[15] "World Health Organization. Substantial investment needed to avert mental health crisis [online] Available at: https://www.who.int/news/item/14-05-2020-substantial-investment-needed-to-avert-mental-health-crisis [Accessed at November 2020] 2020."

[16] M. Butler, T. A. Pollak, A. G. Rooney, B. D. Michael, and T. R. Nicholson, "Neuropsychiatric complications of covid-19," *The BMJ*, vol. 371, m3871, 2020.

[17] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "Causability and Explainability of Artificial Intelligence in Medicine," *WIREs Data Mining and Knowledge Discovery*, vol. 9, no. 4, e1312, 2019.

[18] M. S. Keshavan, D. W. Morris, J. A. Sweeney, G. Pearlson, G. Thaker, L. J. Seidman, S. M. Eack, and C. Tamminga, "A dimensional approach to the psychosis spectrum between bipolar disorder and shizophrenia: The schizo-bipolar scale," *Schizophrenia Research*, vol. 133, no. 1-3, pp. 250–254, 2011.

[19] S. Vieira, W. Pinaya, and A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications," *Neuroscience and Biobehavioral Reviews*, vol. 74, no. Pt A, pp. 58–75, 2017.

[20] D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A.-S. LaMantia, J. O. McNamara, and S. M. Williams, *Neuroscience, 3rd Edition*. Massachusetts: Sinauer Associates Inc Publishers, 2004.

[21] G. D. Hammer and S. J. McPhee, eds., *Pathophysiology of disease: An introduction to clinical medicine, 7th edition*. McGraw-Hill Education Medical, 2014.

[22] A. N. Healey, "Speculation on the neuropsychology of teleoperation: Implications for presence research and minimally invasive surgery," *Presence: Teleoperators and Virtual Environments*, vol. 17, no. 2, pp. 199–211, 2008.

[23] A. G. Filler, "The history, development and impact of computed imaging in neurological diagnosis and neurosurgery: CT, MRI, and DTI," *Nature Precedings*, no. 3267.5, 2009.

[24] M. Jenkinson, M. Chappell, J. Bijsterbosch, S. Smith, C. Beckmann, B. Macintosh, and T. Okell, *Introduction to neuroimaging analysis*. Oxford Neuroimaging Primers, 2017.

[25] B. H. Brown, R. H. Smallwood, D. C. Barber, P. V. Lawford, and D. R. Hose, *Medical physics and biomedical engineering*. CRC Press, 1999.

[26] R. W. Brown, Y.-C. N. Cheng, E. Mark Haacke, M. R. Thompson, R. Venkatesan, D. Manager, M. Applications Engineering, and P. Scientist, *Magnetic resonance imaging: Physical principles and sequence design, 2nd Edition*. John Wiley & Sons, 2014.

[27] V. P. Grover, J. M. Tognarelli, M. M. Crossey, I. J. Cox, S. D. Taylor-Robinson, and M. J. McPhail, "Magnetic resonance imaging: Principles and techniques: Lessons for clinicians," *Journal of Clinical and Experimental Hepatology*, vol. 5, no. 3, p. 246, 2015.

[28] S. Currie, N. Hoggard, I. J. Craven, M. Hadjivassiliou, and I. D. Wilkinson, "Understanding MRI: Basic MR physics for physicians," *Postgraduate Medical Journal*, vol. 89, no. 1050, pp. 209–223, 2013.

[29] L. Schad, "Problems in texture analysis with magnetic resonance imaging.," *Dialogues in Clinical Neuroscience*, vol. 6, no. 2, pp. 235–242, 2004.

[30] J. Bijsterbosch, S. Smith, and C. Beckmann, *Introduction to resting state fMRI functional connectivity*. Oxford Neuroimaging Primers, 2017.

[31] G. H. Glover, "Overview of functional magnetic resonance imaging," *Neurosurgery Clinics of North America*, vol. 22, no. 2, pp. 133–139, 2011.

[32] J. M. Greve, "The BOLD effect," *Methods in Molecular Biology*, vol. 771, pp. 153–169, 2011.

[33] B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar mri," *Magnetic Resonance in Medicine*, vol. 34, no. 4, pp. 537–541, 1995.

[34] J. D. Power, B. L. Schlaggar, and S. E. Petersen, "Studying brain organization via spontaneous fMRI signal," *Neuron*, vol. 84, no. 4, pp. 681–696, 2014.

[35] A. S. Eddin, J. Wang, S. Sargolzaei, W. D. Gaillard, and M. Adjouadi, "ICA-based connectivity on brain networks using fMRI," in *International IEEE/EMBS Conference on Neural Engineering, NER*, pp. 391–394, 2013.

[36] N. Qiang, Q. Dong, Y. Sun, B. Ge, and T. Liu, "Deep variational autoencoder for modeling functional brain networks and ADHD identification," in *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, vol. 2020-April, pp. 554–557, IEEE Computer Society, 2020.

[37] A. R. Franco, M. V. Mannell, V. D. Calhoun, and A. R. Mayer, "Impact of analysis methods on the reproducibility and reliability of resting-state networks," *Brain Connectivity*, vol. 3, no. 4, pp. 363–374, 2013.

[38] L. J. Hearne, J. B. Mattingley, and L. Cocchi, "Functional brain networks related to individual differences in human intelligence at rest," *Scientific Reports*, vol. 6, no. 1, pp. 1–8, 2016.

[39] M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon, "Functional connectivity in the resting brain: A network analysis of the default mode hypothesis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 1, pp. 253–258, 2003.

[40] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen, and M. E. Raichle, "The human brain is intrinsically organized into dynamic, anticorrelated functional networks,"

Proceedings of the National Academy of Sciences of the United States of America, vol. 102, no. 27, pp. 9673–9678, 2005.

[41] J. S. Damoiseaux, S. A. Rombouts, F. Barkhof, P. Scheltens, C. J. Stam, S. M. Smith, and C. F. Beckmann, "Consistent resting-state networks across healthy subjects," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 37, pp. 13848–13853, 2006.

[42] W. W. Seeley, V. Menon, A. F. Schatzberg, J. Keller, G. H. Glover, H. Kenna, A. L. Reiss, and M. D. Greicius, "Dissociable intrinsic connectivity networks for salience processing and executive control," *Journal of Neuroscience*, vol. 27, no. 9, pp. 2349–2356, 2007.

[43] S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, N. Filippini, K. E. Watkins, R. Toro, A. R. Laird, and C. F. Beckmann, "Correspondence of the brain's functional architecture during activation and rest," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 31, pp. 13040–13045, 2009.

[44] W. R. Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, and M. D. Greicius, "Decoding subject-driven cognitive states with whole-brain connectivity patterns," *Cerebral Cortex*, vol. 22, no. 1, pp. 158–165, 2012.

[45] V. Kiviniemi, T. Starck, J. Remes, X. Long, J. Nikkinen, M. Haapea, J. Veijola, I. Moilanen, M. Isohanni, Y. F. Zang, and O. Tervonen, "Functional segmentation of the brain cortex using high model order group PICA," *Human Brain Mapping*, vol. 30, no. 12, pp. 3865–3886, 2009.

[46] M. De Luca, C. F. Beckmann, N. De Stefano, P. M. Matthews, and S. M. Smith, "fMRI resting state networks define distinct modes of long-distance interactions in the human brain," *NeuroImage*, vol. 29, no. 4, pp. 1359–1367, 2006.

[47] C. F. Beckmann, M. DeLuca, J. T. Devlin, and S. M. Smith, "Investigations into resting-state connectivity using independent component analysis," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1457, pp. 1001–1013, 2005.

[48] R. Salvador, J. Suckling, M. R. Coleman, J. D. Pickard, D. Menon, and E. Bullmore, "Neurophysiological architecture of functional magnetic resonance images of human brain," *Cerebral Cortex*, vol. 15, no. 9, pp. 1332–1342, 2005.

[49] M. van den Heuvel, R. Mandl, and H. Hulshoff Pol, "Normalized cut group clustering of resting-state fMRI data," *PLoS ONE*, vol. 3, no. 4, e2001, 2008.

[50] A. R. Laird, P. M. Fox, S. B. Eickhoff, J. A. Turner, K. L. Ray, D. R. Mckay, D. C. Glahn, C. F. Beckmann, S. M. Smith, and P. T. Fox, "Behavioral interpretations of intrinsic connectivity networks," *Journal of Cognitive Neuroscience*, vol. 23, no. 12, pp. 4022–4037, 2011.

[51] B. T. Thomas Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, B. Fisch, H. Liu, and R. L. Buckner, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of Neurophysiology*, vol. 106, no. 3, pp. 1125–1165, 2011.

[52] B. Voloh, R. Knoebl, B. Y. Hayden, and J. Zimmermann, "Oscillations as a window into neuronal mechanisms underlying dorsal anterior cingulate cortex function," *International Review of Neurobiology*, vol. 158, pp. 311–335, 2021.

[53] X. Wang, Q. Wu, L. Egan, X. Gu, P. Liu, H. Gu, Y. Yang, J. Luo, Y. Wu, Z. Gao, and J. Fan, "Anterior insular cortex plays a critical role in interoceptive attention," *eLife*, vol. 8, e42265, 2019.

[54] T. Torrico and S. Munakomi, "Neuroanatomy, thalamus [online] Available at: https://www.ncbi.nlm.nih.gov/books/NBK542184/ [Accessed at 17 May 2021]," 2021.

[55] H. Ekhtiari, P. Nasseri, F. Yavari, A. Mokri, and J. Monterosso, "Neuroscience of drug craving for addiction medicine: From circuits to therapies," *Progress in Brain Research*, vol. 223, pp. 115–141, 2016.

[56] S. Lee, T. Parthasarathi, and J. W. Kable, "The ventral and dorsal default mode networks are dissociably modulated by the vividness and valence of imagined events," *Journal of Neuroscience*, vol. 41, no. 24, pp. 5243–5250, 2021.

[57] S. Marek and N. U. Dosenbach, "The frontoparietal network: function, electrophysiology, and importance of individual precision mapping," *Dialogues in Clinical Neuroscience*, vol. 20, no. 2, p. 133, 2018.

[58] R. E. Beaty, M. Benedek, S. Barry Kaufman, and P. J. Silvia, "Default and executive network coupling supports creative idea production," *Scientific Reports*, vol. 5, no. 1, pp. 1–14, 2015.

[59] S. J. Coen, A. R. Hobson, and Q. Aziz, "Processing of gastrointestinal sensory signals in the brain," *Physiology of the Gastrointestinal Tract*, vol. 1, pp. 689–702, 2012.

[60] E. T. Rolls, "The neuroscience of emotional disorders," *Handbook of Clinical Neurology*, vol. 183, pp. 1–26, 2021.

[61] D. Borsook, N. Maleki, and R. Burstein, "Migraine," *Neurobiology of Brain Disorders: Biological Basis of Neurological and Psychiatric Disorders*, pp. 693–708, 2015.

[62] S. Chenji, S. Jha, D. Lee, M. Brown, P. Seres, D. Mah, and S. Kalra, "Investigating default mode and sensorimotor network connectivity in amyotrophic lateral sclerosis," *PLoS ONE*, vol. 11, no. 6, e0157443, 2016.

[63] S. Vossel, J. J. Geng, and G. R. Fink, "Dorsal and ventral attention systems: Distinct neural circuits but collaborative roles," *The Neuroscientist*, vol. 20, no. 2, p. 150, 2014.

[64] J. K. Mai and G. Paxinos, eds., *The human nervous system*. Elsevier Inc., 2012.

[65] R. L. DeLaPaz, "Echo-planar imaging.," *Radiographics*, vol. 14, no. 5, pp. 1045–1058, 1994.

[66] M. Poustchi-Amin, S. A. Mirowitz, J. J. Brown, R. C. McKinstry, and T. Li, "Principles and applications of echo-planar imaging: A review for the general radiologist," *Radiographics*, vol. 21, no. 3, pp. 767–779, 2001.

REFERENCES

[67] K. R. A. V. Dijk, T. Hedden, A. Venkataraman, K. C. Evans, S. W. Lazar, and R. L. Buckner, "Intrinsic functional connectivity as a tool for human connectomics: Theory, properties, and optimization," *Journal of Neurophysiology*, vol. 103, no. 1, p. 297, 2010.

[68] R. M. Birn, E. K. Molloy, R. Patriat, T. Parker, T. B. Meier, G. R. Kirk, V. A. Nair, M. E. Meyerand, and V. Prabhakaran, "The effect of scan length on the reliability of resting-state fMRI connectivity estimates," *NeuroImage*, vol. 83, pp. 550–558, 2013.

[69] L. Demetriou, O. S. Kowalczyk, G. Tyson, T. Bello, R. D. Newbould, and M. B. Wall, "A comprehensive evaluation of increasing temporal resolution with multiband-accelerated protocols and effects on statistical outcome measures in fMRI," *NeuroImage*, vol. 176, pp. 404–416, 2018.

[70] H. Jahanian, S. Holdsworth, T. Christen, H. Wu, K. Zhu, A. B. Kerr, M. J. Middione, R. F. Dougherty, M. Moseley, and G. Zaharchuk, "Advantages of short repetition time resting-state functional MRI enabled by simultaneous multi-slice imaging," *Journal of Neuroscience Methods*, vol. 311, pp. 122–132, 2019.

[71] N. Huotari, L. Raitamaa, H. Helakari, J. Kananen, V. Raatikainen, A. Rasila, T. Tuovinen, J. Kantola, V. Borchardt, V. J. Kiviniemi, and V. O. Korhonen, "Sampling rate effects on resting state fMRI metrics," *Frontiers in Neuroscience*, vol. 13, p. 279, 2019.

[72] N. Zhao, L.-X. Yuan, X.-Z. Jia, X.-F. Zhou, X.-P. Deng, H.-J. He, J. Zhong, J. Wang, and Y.-F. Zang, "Intra- and inter-scanner reliability of voxel-wise whole-brain analytic metrics for resting state fMRI," *Frontiers in Neuroinformatics*, vol. 12, p. 54, 2018.

[73] A. J. Thomas and D. Bathula, "Reducing inter-scanner variability in multi-site fMRI data: Exploring choice of reference activation map and use of correction functions," in *International Conference on Computing, Communication and Automation (ICCCA)*, pp. 1187–1192, IEEE, 2015.

[74] C. Yan, D. Liu, Y. He, Q. Zou, C. Zhu, X. Zuo, X. Long, and Y. Zang, "Spontaneous brain activity in the default mode network is sensitive to different resting-state conditions with limited cognitive load," *PLOS ONE*, vol. 4, no. 5, e5743, 2009.

[75] K. Specht, "Current challenges in translational and clinical fMRI and future directions," *Frontiers in Psychiatry*, vol. 10, p. 924, 2019.

[76] N. K. Aurich, J. O. Filho, A. M. da Silva, and A. R. Franco, "Evaluating the reliability of different preprocessing steps to estimate graph theoretical measures in resting state fMRI data," *Frontiers in Neuroscience*, vol. 9, p. 48, 2015.

[77] S. E. Joel, B. S. Caffo, P. C. Van Zijl, and J. J. Pekar, "On the relationship between seed-based and ICA-based measures of functional connectivity," *Magnetic Resonance in Medicine*, vol. 66, no. 3, pp. 644–657, 2011.

[78] L. Wu, A. Caprihan, J. Bustillo, A. Mayer, and V. Calhoun, "An approach to directly link ICA and seed-based functional connectivity: Application to schizophrenia," *NeuroImage*, vol. 179, pp. 448–470, 2018.

[79] D. M. Cole, S. M. Smith, and C. F. Beckmann, "Advances and pitfalls in the analysis and interpretation of resting-state FMRI data," *Frontiers in Systems Neuroscience*, vol. 4, p. 8, 2010.

[80] W. S. Sohn, K. Yoo, Y. B. Lee, S. W. Seo, D. L. Na, and Y. Jeong, "Influence of ROI selection on resting functional connectivity: An individualized approach for resting fMRI analysis," *Frontiers in Neuroscience*, vol. 9, p. 280, 2015.

[81] M. R. Arbabshirani, E. Castro, and V. D. Calhoun, "Accurate classification of schizophrenia patients based on novel resting-state fMRI features," in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC*, vol. 2014, pp. 6691–6694, IEEE, 2014.

[82] L. Griffanti, G. Douaud, J. Bijsterbosch, S. Evangelisti, F. Alfaro-Almagro, M. F. Glasser, E. P. Duff, S. Fitzgibbon, R. Westphal, D. Carone, C. F. Beckmann, and S. M. Smith, "Hand classification of fMRI ICA noise components," *NeuroImage*, vol. 154, pp. 188–205, 2017.

[83] R. H. Pruim, M. Mennes, D. van Rooij, A. Llera, J. K. Buitelaar, and C. F. Beckmann, "ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data," *NeuroImage*, vol. 112, pp. 267–277, 2015.

[84] G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith, "Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers," *NeuroImage*, vol. 90, pp. 449–468, 2014.

[85] L. Griffanti, G. Salimi-Khorshidi, C. F. Beckmann, E. J. Auerbach, G. Douaud, C. E. Sexton, E. Zsoldos, K. P. Ebmeier, N. Filippini, C. E. Mackay, S. Moeller, J. Xu, E. Yacoub, G. Baselli, K. Ugurbil, K. L. Miller, and S. M. Smith, "ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging," *NeuroImage*, vol. 95, pp. 232–247, 2014.

[86] K. Li, L. Si, B. Cui, X. Ling, B. Shen, and X. Yang, "Altered intra- and inter-network functional connectivity in patients with persistent postural-perceptual dizziness," *NeuroImage: Clinical*, vol. 26, 102216, 2020.

[87] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, S. M. Smith, and D. C. Van Essen, "A multi-modal parcellation of human cerebral cortex," *Nature*, vol. 536, no. 7615, pp. 171–178, 2016.

[88] H. S. Parmar, *Machine learning in functional magnetic resonance neuroimaging analysis*. PhD thesis, 2020.

[89] Y. Bengio, I. Goodfellow, and A. Courville, *Deep learning*. Massachusetts, USA: MIT press, 2015.

[90] K. G. Kim, "Book review: Deep learning," *Healthcare Informatics Research*, vol. 22, no. 4, p. 351, 2016.

[91] F. Emmert-Streib, Z. Yang, H. Feng, S. Tripathi, and M. Dehmer, "An introductory review of deep learning for prediction models with big data," *Frontiers in Artificial Intelligence*, vol. 3, p. 4, 2020.

[92] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–27, 2009.

[93] C. Cao, F. Liu, H. Tan, D. Song, W. Shu, W. Li, Y. Zhou, X. Bo, and Z. Xie, "Deep learning and its applications in biomedicine," *Genomics, Proteomics and Bioinformatics*, vol. 16, no. 1, pp. 17–32, 2018.

[94] D. P. Kingma and J. Lei Ba, "ADAM: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2015.

[95] G. Hinton and T. Tieleman, "Lecture 6.5 - rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural Networks for Machine Learning*, vol. 4, pp. 26–31, 2012.

[96] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, 2011.

[97] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, vol. 9, pp. 249–256, JMLR Workshop and Conference Proceedings, 2010.

[98] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.

[99] W. Pinaya, C. Scarpazza, R. Garcia-Dias, S. Vieira, L. Baecker, P. da Costa, A. Redolfi, G. Frisoni, M. Pievani, V. Calhoun, J. Sato, and A. Mechelli, "Normative modelling using deep autoencoders: a multi-cohort study on mild cognitive impairment and Alzheimer's disease," *bioRxiv preprint bioRxiv:2020.02.10.931824*, 2020.

[100] A. F. Marquand, S. M. Kia, M. Zabihi, T. Wolfers, J. K. Buitelaar, and C. F. Beckmann, "Conceptualizing mental disorders as deviations from normative functioning," *Molecular Psychiatry*, vol. 24, no. 10, pp. 1415–1424, 2019.

[101] W. Pinaya, A. Mechelli, and J. R. Sato, "Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study," *Human Brain Mapping*, vol. 40, no. 3, pp. 944–954, 2019.

[102] S. Vieira, W. Pinaya, R. Garcia-Dias, and A. Mechelli, "Deep neural networks," *Machine Learning: Methods and Applications to Brain Disorders*, pp. 157–172, 2019.

[103] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *arXiv preprint arXiv:1406.2661*, vol. 27, 2014.

[104] S. Kazeminia, C. Baur, A. Kuijper, B. van Ginneken, N. Navab, S. Albarqouni, and A. Mukhopadhyay, "GANs for medical image analysis," *Artificial Intelligence in Medicine*, vol. 109, 101938, 2020.

[105] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial Networks to guide marker discovery," in *International conference on information processing in medical imaging*, pp. 146–157, Springer, Cham, 2017.

[106] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," *arXiv preprint arXiv:2003.05991*, 2020.

[107] L. Girin, S. Leglaive, X. Bie, J. Diard, T. Hueber, and X. Alameda-Pineda, "Dynamical variational autoencoders: A comprehensive review," *arXiv preprint arXiv:2008.12595*, 2020.

[108] P. Salehi, A. Chalechale, and M. Taghizadeh, "Generative adversarial networks (GANs): An overview of theoretical model, evaluation metrics, and recent developments," *arXiv preprint arXiv:2005.13178*, 2020.

[109] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[110] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," *arXiv preprint arXiv:1511.05644*, 2015.

[111] L. Mescheder, S. Nowozin, and A. Geiger, "Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks," in *34th International Conference on Machine Learning, ICML*, vol. 5, pp. 3694–3707, International Machine Learning Society (IMLS), 2017.

[112] H. Singh, M. M. Gupta, T. Meitzler, Z. G. Hou, K. K. Garg, A. M. Solo, and L. A. Zadeh, "Real-life applications of fuzzy logic," *Advances in Fuzzy Systems*, vol. 2013, 581879, 2013.

[113] S. Sujamol, S. Ashok, and U. Krishna Kumar, "Fuzzy based machine learning: A promising approach," *CSI Commun. Knowl. Digest for IT Community*, vol. 41, no. 8, pp. 21–25, 2017.

[114] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.

[115] R. Krishnapuram and J. M. Keller, "The possibilistic C-means algorithm: Insights and recommendations," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 3, pp. 385–393, 1996.

[116] N. R. Pal, K. Pal, and J. C. Bezdek, "Mixed c-means clustering model," in *IEEE International Conference on Fuzzy Systems*, vol. 1, pp. 11–21, IEEE, 1997.

[117] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 517–530, 2005.

[118] C. S. George Lee and C. T. Lin, "Supervised and unsupervised learning with fuzzy similarity for neural-network-based fuzzy logic control systems," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, vol. 1992-January, pp. 688–693, IEEE, 1992.

[119] S. R. Price, S. R. Price, and D. T. Anderson, "Introducing fuzzy layers for deep learning," in *IEEE International Conference on Fuzzy Systems*, vol. 2019-June, IEEE, 2019.

[120] P. Song, M. Zha, Q. Yang, Y. Zhang, X. Li, and I. Rudan, "The prevalence of adult attention-deficit hyperactivity disorder: A global systematic review and meta-analysis," *Journal of Global Health*, vol. 11, pp. 1–9, 2021.

[121] F. X. Castellanos and Y. Aoki, "Intrinsic functional connectivity in attention-deficit / hyperactivity disorder: A science in development," *Biological Psychiatry : Cognitive Neuroscience and Neuroimaging*, vol. 1, no. 3, p. 253, 2016.

[122] F. X. Castellanos, D. S. Margulies, C. Kelly, L. Q. Uddin, M. Ghaffari, A. Kirsch, D. Shaw, Z. Shehzad, A. Di Martino, B. Biswal, E. J. Sonuga-Barke, J. Rotrosen, L. A. Adler, and M. P.

Milham, "Cingulate - precuneus interactions: A new locus of dysfunction in adult attention-deficit/hyperactivity disorder," *Biological Psychiatry*, vol. 63, no. 3, p. 332, 2008.

[123] C. Sripada, D. Kessler, Y. Fang, R. C. Welsh, K. Prem Kumar, and M. Angstadt, "Disrupted network architecture of the resting brain in attention-deficit/hyperactivity disorder," *Human Brain Mapping*, vol. 35, no. 9, p. 4693, 2014.

[124] D. Kessler, M. Angstadt, R. C. Welsh, and C. ipada, "Modality-spanning deficits in attention-deficit/hyperactivity disorder in functional networks, gray matter, and white matter," *The Journal of Neuroscience*, vol. 34, no. 50, pp. 16555–16566, 2014.

[125] L. Sun, Q. Cao, X. Long, M. Sui, X. Cao, C. Zhu, X. Zuo, L. An, Y. Song, Y. Zang, and Y. Wang, "Abnormal functional connectivity between the anterior cingulate and the default mode network in drug-naïve boys with attention deficit hyperactivity disorder," *Psychiatry Research*, vol. 201, no. 2, pp. 120–127, 2012.

[126] B. Sutcubasi, B. Metin, M. K. Kurban, Z. E. Metin, B. Beser, and E. Sonuga-Barke, "Resting-state network dysconnectivity in ADHD: A system-neuroscience-based meta-analysis," *The World Journal of Biological Psychiatry*, vol. 21, no. 9, pp. 662–672, 2020.

[127] F. Aboitiz, T. Ossandón, F. Zamorano, B. Palma, and X. Carrasco, "Irrelevant stimulus processing in ADHD: Catecholamine dynamics and attentional networks," *Frontiers in Psychology*, vol. 5, p. 183, 2014.

[128] H. Lin, Q. Lin, H. Li, M. Wang, H. Chen, Y. Liang, X. Bu, W. Wang, Y. Yi, Y. Zhao, X. Zhang, Y. Xie, S. Du, C. Yang, and X. Huang, "Functional connectivity of attention-related networks in drug-naïve children with ADHD," *Journal of Attention Disorders*, vol. 25, no. 3, pp. 377–388, 2021.

[129] C. Brites, "ADHD and impact on language," in *ADHD - From Etiology to Comorbidity*, ch. 5, IntechOpen, 2020.

[130] "World Health Organization. Mental disorders. [online] Available at: https://www.who.int/news-room/fact-sheets/detail/mental-disorders [Accessed at December 2021] 2019."

[131] S. Yoon, T. D. Kim, J. Kim, and I. K. Lyoo, "Altered functional activity in bipolar disorder: A comprehensive review from a large-scale network perspective," *Brain and Behavior*, vol. 11, no. 1, e01953, 2021.

[132] J. Wang, Y. Wang, H. Huang, Y. Jia, S. Zheng, S. Zhong, L. Huang, and R. Huang, "Abnormal intrinsic brain functional network dynamics in unmedicated depressed bipolar II disorder," *Journal of affective disorders*, vol. 253, pp. 402–409, 2019.

[133] S. A. Meda, A. Gill, M. C. Stevens, R. P. Lorenzoni, D. C. Glahn, V. D. Calhoun, J. A. Sweeney, C. A. Tamminga, M. S. Keshavan, G. Thaker, and G. D. Pearlson, "Differences in resting-state fMRI functional network connectivity between schizophrenia and psychotic bipolar probands and their unaffected first-degree relatives," *Biological psychiatry*, vol. 71, no. 10, p. 881, 2012.

[134] G. Lois, J. Linke, and M. Wessa, "Altered functional connectivity between emotional and cognitive resting state networks in euthymic bipolar I disorder patients," *PLoS ONE*, vol. 9, no. 10, e107829, 2014.

[135] A. Anand, Y. Li, Y. Wang, M. J. Lowe, and M. Dzemidzic, "Resting state corticolimbic connectivity abnormalities in unmedicated bipolar disorder and unipolar depression," *Psychiatry research*, vol. 171, no. 3, p. 189, 2009.

[136] K. C. Skåtun, T. Kaufmann, C. L. Brandt, N. T. Doan, D. Alnæs, S. Tønnesen, G. Biele, A. Vaskinn, I. Melle, I. Agartz, O. A. Andreassen, and L. T. Westlye, "Thalamo-cortical functional connectivity in schizophrenia and bipolar disorder," *Brain imaging and behavior*, vol. 12, no. 3, pp. 640–652, 2018.

[137] K. Friston and C. Frith, "Schizophrenia: A disconnection syndrome?," *Clinical Neuroscience*, vol. 3, pp. 89–97, 1995.

[138] M. J. Jafri, G. D. Pearlson, M. Stevens, and V. D. Calhoun, "A method for functional network connectivity among spatially independent resting-state components in schizophrenia," *NeuroImage*, vol. 39, no. 4, pp. 1666–1681, 2008.

[139] A. G. Garrity, G. D. Pearlson, K. McKiernan, D. Lloyd, K. A. Kiehl, and V. D. Calhoun, "Aberrant "default mode" functional connectivity in schizophrenia," *American Journal of Psychiatry*, vol. 164, no. 3, pp. 450–457, 2007.

[140] Y. Wang, W. Tang, X. Fan, J. Zhang, D. Geng, K. Jiang, D. Zhu, Z. Song, Z. Xiao, and D. Liu, "Resting-state functional connectivity changes within the default mode network and the salience network after antipsychotic treatment in early-phase schizophrenia," *Neuropsychiatric Disease and Treatment*, vol. 13, pp. 397–406, 2017.

[141] V. D. Calhoun, K. A. Kiehl, and G. D. Pearlson, "Modulation of temporally coherent brain networks estimated using ICA at rest and during cognitive tasks," *Human Brain Mapping*, vol. 29, no. 7, pp. 828–838, 2008.

[142] M. Liang, Y. Zhou, T. Jiang, Z. Liu, L. Tian, H. Liu, and Y. Hao, "Widespread functional disconnectivity in schizophrenia with resting-state functional magnetic resonance imaging," *NeuroReport*, vol. 17, no. 2, pp. 209–213, 2006.

[143] J. A. Bernard, C. E. Russell, R. E. Newberry, J. R. Goen, and V. A. Mittal, "Patients with schizophrenia show aberrant patterns of basal ganglia activation: Evidence from ALE meta-analysis," *NeuroImage: Clinical*, vol. 14, pp. 450–463, 2017.

[144] P. Li, T. T. Fan, R. J. Zhao, Y. Han, L. Shi, H. Q. Sun, S. J. Chen, J. Shi, X. Lin, and L. Lu, "Altered brain network connectivity as a potential endophenotype of schizophrenia," *Scientific Reports*, vol. 7, no. 1, pp. 1–9, 2017.

[145] V. D. Calhoun, P. K. Maciejewski, G. D. Pearlson, and K. A. Kiehl, "Temporal lobe and "default" hemodynamic brain modes discriminate between schizophrenia and bipolar disorder," *Human Brain Mapping*, vol. 29, no. 11, pp. 1265–1275, 2008.

[146] M. Xia, F. Womer, M. Chang, Y. Zhu, Q. Zhou, E. K. Edmiston, X. Jiang, S. Wei, J. Duan, K. Xu, Y. Tang, Y. He, and F. Wang, "Shared and distinct functional architectures of brain networks across psychiatric disorders," *Schizophrenia Bulletin*, vol. 45, no. 2, pp. 450–463, 2018.

[147] J. Oh, B.-L. Oh, K.-U. Lee, J.-H. Chae, and K. Yun, "Identifying schizophrenia using structural MRI with a deep learning algorithm," *Frontiers in Psychiatry*, vol. 11, p. 16, 2020.

REFERENCES

[148] W. Huizinga, D. H. Poot, M. W. Vernooij, G. V. Roshchupkin, E. E. Bron, M. A. Ikram, D. Rueckert, W. J. Niessen, and S. Klein, "A spatio-temporal reference model of the aging brain," *NeuroImage*, vol. 169, pp. 11–22, 2018.

[149] A. F. Marquand, I. Rezek, J. Buitelaar, and C. F. Beckmann, "Understanding heterogeneity in clinical cohorts using normative models: Beyond case-control studies," *Biological Psychiatry*, vol. 80, no. 7, pp. 552–561, 2016.

[150] S. M. Kia and A. F. Marquand, "Neural processes mixed-effect models for deep normative modeling of clinical neuroimaging data," in *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning, PMLR*, vol. 102, pp. 297–314, arXiv, 2019.

[151] T. Wolfers, N. T. Doan, T. Kaufmann, D. Alnæs, T. Moberget, I. Agartz, J. K. Buitelaar, T. Ueland, I. Melle, B. Franke, O. A. Andreassen, C. F. Beckmann, L. T. Westlye, and A. F. Marquand, "Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models," *JAMA Psychiatry*, vol. 75, no. 11, pp. 1146–1155, 2018.

[152] M. Zabihi, M. Oldehinkel, T. Wolfers, V. Frouin, D. Goyard, E. Loth, T. Charman, J. Tillmann, T. Banaschewski, G. Dumas, R. Holt, S. Baron-Cohen, S. Durston, S. Bölte, D. Murphy, C. Ecker, J. K. Buitelaar, C. F. Beckmann, and A. F. Marquand, "Dissecting the heterogeneous cortical anatomy of autism spectrum disorder using normative models," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 4, no. 6, pp. 567–578, 2019.

[153] G. Ziegler, G. R. Ridgway, R. Dahnke, and C. Gaser, "Individualized Gaussian process-based prediction and detection of local and global gray matter abnormalities in elderly subjects," *NeuroImage*, vol. 97, no. 100, pp. 333–348, 2014.

[154] D. Kessler, M. Angstadt, and C. Sripada, "Growth charting of brain connectivity networks and the identification of attention impairment in youth," *JAMA Psychiatry*, vol. 73, no. 5, pp. 481–489, 2016.

[155] T. Wolfers, J. Rokicki, D. Alnæs, P. Berthet, I. Agartz, S. M. Kia, T. Kaufmann, M. Zabihi, T. Moberget, I. Melle, C. F. Beckmann, O. A. Andreassen, A. F. Marquand, and L. T. Westlye, "Replicating extensive brain structural heterogeneity in individuals with schizophrenia and bipolar disorder," *Human brain mapping*, vol. 42, no. 8, pp. 2546–2555, 2021.

[156] M. Zabihi, D. Floris, S. M. Kia, T. Wolfers, J. Tillmann, A. Llera, C. Moessnang, T. Banaschewski, R. Holt, S. Baron-Cohen, E. Loth, T. Charman, T. Bourgeron, D. Murphy, C. Ecker, J. Buitelaar, C. Beckmann, and A. Marquand, "Fractionating autism based on neuroanatomical normative modeling," *Translational Psychiatry*, vol. 10, no. 1, pp. 1–10, 2020.

[157] S. M. Kia, H. Huijsdens, R. Dinga, T. Wolfers, M. Mennes, O. A. Andreassen, L. T. Westlye, C. F. Beckmann, and A. F. Marquand, "Hierarchical Bayesian regression for multi-site normative modeling of neuroimaging data," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 12267 LNCS, pp. 699–709, Springer, Cham, 2020.

[158] M. Fernandes, "Detecting Abnormal Functional Connectivity Patterns Using Stacked Autoencoders," *technical report, Instituto de Biofisica e Engenharia Biomedica*, 2018.

[159] X. N. Zuo, J. S. Anderson, P. Bellec, R. M. Birn, B. B. Biswal, J. Blautzik, J. C. Breitner, R. L. Buckner, V. D. Calhoun, F. X. Castellanos, A. Chen, B. Chen, J. Chen, X. Chen, S. J. Colcombe, W. Courtney, R. C. Craddock, A. Di Martino, H. M. Dong, X. Fu, Q. Gong, K. J. Gorgolewski, Y. Han, Y. He, Y. He, E. Ho, A. Holmes, X. H. Hou, J. Huckins, T. Jiang, Y. Jiang, W. Kelley, C. Kelly, M. King, S. M. LaConte, J. E. Lainhart, X. Lei, H. J. Li, K. Li, K. Li, Q. Lin, D. Liu, J. Liu, X. Liu, Y. Liu, G. Lu, J. Lu, B. Luna, J. Luo, D. Lurie, Y. Mao, D. S. Margulies, A. R. Mayer, T. Meindl, M. E. Meyerand, W. Nan, J. A. Nielsen, D. O'Connor, D. Paulsen, V. Prabhakaran, Z. Qi, J. Qiu, C. Shao, Z. Shehzad, W. Tang, A. Villringer, H. Wang, K. Wang, D. Wei, G. X. Wei, X. C. Weng, X. Wu, T. Xu, N. Yang, Z. Yang, Y. F. Zang, L. Zhang, Q. Zhang, Z. Zhang, Z. Zhang, K. Zhao, Z. Zhen, Y. Zhou, X. T. Zhu, and M. P. Milham, "An open science resource for establishing reliability and reproducibility in functional connectomics," *Scientific Data*, vol. 1, no. 1, pp. 1–13, 2014.

[160] K. J. Gorgolewski, J. Durnez, and R. A. Poldrack, "Preprocessed Consortium for Neuropsychiatric Phenomics dataset," *F1000Research*, vol. 6, p. 1262, 2017.

[161] "COBRE Phase 3 — The Mind Research Network (MRN) https://www.mrn.org/common/cobre-phase-3 [Accessed at August 2021]."

[162] B. B. Avants, N. Tustison, and H. Johnson, "Advanced Normalization Tools (ANTS) Release 2.x," *Insight j*, vol. 2, no. 365, pp. 1–41, 2014.

[163] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *NeuroImage*, vol. 62, no. 2, pp. 782–790, 2012.

[164] A. D. Nostro, V. I. Müller, D. P. Varikuti, R. N. Pläschke, F. Hoffstaedter, R. Langner, K. R. Patil, and S. B. Eickhoff, "Predicting personality from network-based resting-state functional connectivity," *Brain structure & function*, vol. 223, no. 6, pp. 2699–2719, 2018.

Appendix A

Appendix

A.1 Additional Information

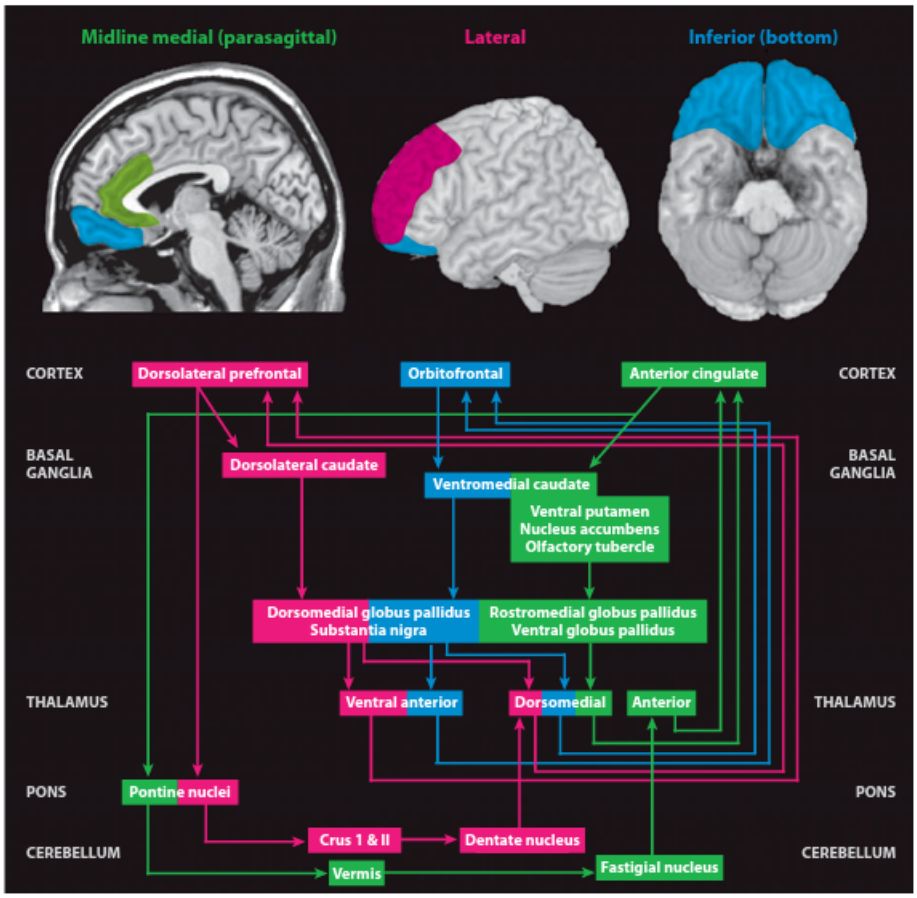


Figure A.1: The three major cortico-subcortical circuits important for neurobehavioural functions that are frequently impaired in NPD are colour-coded onto representative magnetic resonance images (top) and the diagram below. The circuit that begins/ends in the dorsolateral prefrontal cortex (pink) mediates executive functions such as organization, planning, and attention. The circuit that begins/ends in the orbitofrontal cortex (blue) mediates socially appropriate behaviour, impulse control, and empathy. The circuit that begins/ends in the anterior cingulate cortex (green) produces motivation by balancing the inhibitory input of the supplemental motor area with its own stimulus that supports wakefulness and arousal. (from [7])

Anatomic area	Summary of functions
Dorsolateral prefrontal	Cognition, executive function, focused attention
Orbital prefrontal	Social conduct, insight, judgment, mood
Parietal	Sensation, speech production/conduction, and deficit recognition
Mesial temporal lobe, hippocampus, parahippocampus, amygdala	Memory formation and storage
Basal ganglia	Suppression/modulation of involuntary movements; contributes to memory, cognition, behavior, and mood
Thalamus	Key "relay station" for memory, emotion, cognition, behavior, motor, and sensory functions
Hypothalamus	Physiological response to emotional stimuli; temperature control, sleep, water metabolism, hormone secretion, satiety, circadian rhythms
Pons	Contains locus ceruleus (norepinephrine production) and portions of the reticular formation (alertness)
Cerebellum	Equilibrium, fine motor coordination; associated with cognition and executive functions

Figure A.2: Functional anatomy associations. (from [7])

A. APPENDIX

	Anatomical Location of Functional ROIs
Auditory	Left Superior Temporal Gyrus, Heschl's Gyrus
	Right Superior Temporal Gyrus
	Right Thalamus
Basal Ganglia	Left Thalamus, Caudate
	Right Thalamus, Caudate, Putamen
	Left Inferior Frontal Gyrus
	Right Inferior Frontal Gyrus
	Pons
PCC/MPFC	Medial Prefrontal Cortex, Anterior Cingulate Cortex, Orbitofrontal Cortex
	Left Angular Gyrus
	Right Superior Frontal Gyrus
	Posterior Cingulate Cortex, Precuneus
	Midcingulate Cortex
	Right Angular Gyrus
	Left and Right Thalamus
	Left Hippocampus
	Right Hippocampus
V2	Left Middle Occipital Gyrus, Superior Occipital Gyrus
	Right Middle Occipital Gyrus, Superior Occipital Gyrus
Language	Inferior Frontal Gyrus
	Left Middle Temporal Gyrus
	Left Middle Temporal Gyrus, Angular Gyrus
	Left Middle Temporal Gyrus, Superior Temporal Gyrus, Supramarginal Gyrus, Angular Gyrus
	Right Inferior Frontal Gyrus
	Right Supramarginal Gyrus, Superior Temporal Gyrus, Middle Temporal Gyrus
	Left Crus I
Left DLPFC/Parietal	Left Middle Frontal Gyrus, Superior Frontal Gyrus
	Left Inferior Frontal Gyrus, Orbitofrontal Gyrus
	Left Superior Parietal Gyrus, Inferior Parietal Gyrus, Precuneus, Angular Gyrus
	Left Inferior Temporal Gyrus, Middle Temporal Gyrus
	Right Crus I
	Left Thalamus
Sensorimotor	Left Precentral Gyrus, Postcentral Gyrus
	Right Precentral Gyrus, Postcentral Gyrus
	Right Supplementary Motor Area
	Left Thalamus
	Bilateral Lobule IV, Lobule V, Lobule VI
	Right Thalamus
Posterior Insula	Left Middle Frontal Gyrus
	Left Supramarginal Gyrus, Inferior Parietal Gyrus
	Left Precuneus
	Right Midcingulate Cortex
	Right Superior Parietal Gyrus, Precuneus
	Right Supramarginal Gyrus, Inferior Parietal Gyrus
	Left Thalamus
	Lobule VI
	Left Posterior Insula, Putamen
	Right Thalamus
	Lobule VI
	Right Posterior Insula
Precuneus	Midcingulate Cortex, Posterior Cingulate Cortex
	Precuneus
	Left Angular Gyrus
	Right Angular Gyrus
V1	Calcarine Sulcus
	Left Thalamus
Right DLPFC/Parietal	Right Middle Frontal Gyrus, Right Superior Frontal Gyrus
	Right Middle Frontal Gyrus
	Right Inferior Parietal Gyrus, Supramarginal Gyrus, Angular Gyrus
	Right Superior Frontal Gyrus
	Left Crus I, Crus II, Lobule VI
	Right Caudate
Insula/dACC	Left Middle Frontal Gyrus
	Left Insula
	Anterior Cingulate Cortex, Medial Prefrontal Cortex, Supplementary Motor Area
	Right Middle Frontal Gyrus
	Right Insula
	Left Lobule VI, Crus I
	Right Lobule VI, Crus I
RSC/MTL	Left Retrosplenial Cortex, Posterior Cingulate Cortex
	Left Middle Frontal Gyrus
	Left Parahippocampal Gyrus
	Left Middle Occipital Gyrus
	Right Retrosplenial Cortex, Posterior Cingulate Cortex
	Precuneus
	Right Superior Frontal Gyrus, Middle Frontal Gyrus
	Right Parahippocampal Gyrus
	Right Angular Gyrus, Middle Occipital Gyrus
	Right Lobule IX
IPS/FEF	Left Middle Frontal Gyrus, Superior Frontal Gyrus, Precentral Gyrus
	Left Inferior Parietal Sulcus
	Left Frontal Operculum, Inferior Frontal Gyrus
	Left Inferior Temporal Gyrus
	Right Middle Frontal Gyrus
	Right Inferior Parietal Lobule
	Right Frontal Operculum, Inferior Frontal Gyrus
	Right Middle Temporal Gyrus
	Left Lobule VIII, Lobule VIIb
	Right Lobule VIII, Lobule VIIb
	Right Lobule VI, Crus I

Figure A.3: Information about the anatomy of the FBN from Shirer *et al* [44]. Auditory- AN; Basal Ganglia- BGN; posterior cingulate cortex/medial prefrontal cortex- DDMN; V2- HVN; Language- LN; Left dorsolateral prefrontal/parietal cortices - LECN; Sensorimotor- SN; Posterior Insula- PSN; Precuneus- PN; V1- PVN; Right dorsolateral prefrontal/parietal cortices- RECN; Insula/dorsal anterior cingulate cortex- ASN; retrosplenial cortex/medial temporal lobe- VDMN; intraparietal sulcus / frontal eye fields - VN (from [44])

A.2 Scientific Output

Section presenting the abstracts and posters presented at the 1st ISMRM Iberian Chapter Annual Meeting, June 16-17, 2021 (“Effect of BOLD signal data trimming on functional connectivity metrics”), and at the ESMRMB 38th Annual Scientific Meeting, October 7-9, 2021 (“BOLD signal data trimming results in increased functional connectivity”).

Effect of BOLD signal data trimming on functional connectivity metrics

D Saraiva[1], HA Ferreira[1]

1-Institute of Biophysics and Biomedical Engineering, Faculty of Sciences of the University of Lisbon

Abstract

INTRODUCTION

In the Big Data Era, with a lack of comparable functional neuroimaging data, scientists try to combine data that may differ in terms of scan parameters and trim all data to the same number of time points. However, the effects of trimming BOLD signal data are still poorly understood.

METHODS

Functional MRI data from thirty subjects were pre-processed for five different time points, and individual functional connectivity matrices were generated. The correlation matrices were binarized for several thresholds, which excluded weak correlations and included both positive and negative correlations for the analysis. Graph metrics were calculated to study functional connectivity differences between different time points. Small-worldness was evaluated using the sigma and omega coefficients [1, 2]. A comparison between the brain, random, and lattice networks was also performed.

RESULTS

A tendency of decreasing the network degree of the global network for more time points was observed. Regarding small-worldness, while the sigma coefficient was greater than one for all observations and had an increasing pattern for more time points, the omega coefficient was positive for all observations and trended to zero. In addition, the omega coefficient was smaller and greater for lattice and random networks, respectively.

DISCUSSION & CONCLUSION

A decrease of the number of edges with an increase of time points can be interpreted as a loss of stronger correlations with time. Those dynamical alterations are also observed for the small-worldness metrics. Finally, although we observe a tendency for an increase of small-world characteristics with time points, when compared with other networks, we observe that the overall characteristics of brain networks are almost constant and that trimming BOLD signal data leads to small differences in terms of functional connectivity analysis.

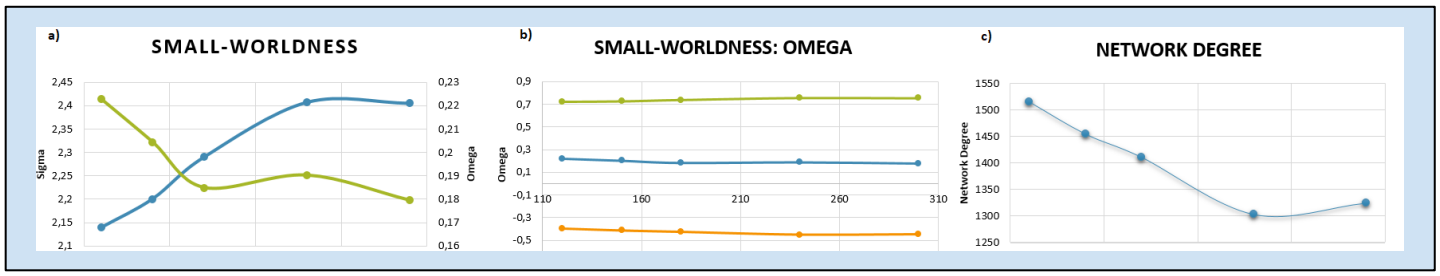


Figure 1 *– Results for a correlation coefficient threshold of 0.35 for positive correlations, and -0.35 for negative correlations. a) Evaluation of sigma and omega coefficients variations with time points, which shows a tendency of increasing small-world characteristics for more time points. b) Comparison of the brain, random, and lattice networks omega coefficient changes with time points, which demonstrates that brain networks keep the small-world characteristics independently of the number of time points. c) Variation of the global brain network degree with time points, which shows a decreasing trend.*

References

[1] – Humphries, MD et al, “Network ‘Small-World-Ness’: A Quantitative Method for Determining Canonical Network Equivalence” PLoS One, vol. 3,4 (2008): e2051.

[2] - Telesford, QK et al. “The ubiquity of small-world networks.” Brain connectivity vol. 1,5 (2011): 367-75.

Effect of BOLD Signal Data Trimming on Functional Connectivity Metrics

Duarte Saraiva[1], Hugo Ferreira[1]
1- Institute of Biophysics and Biomedical Engineering, Faculty of Sciences of the University of Lisbon

Introduction

Big Data Era → Machine Learning Robust Studies

Problem: Lack of Comparable fMRI Data

Trimming Different fMRI Data → More Comparable Data

What is the effect of trimming fMRI data?

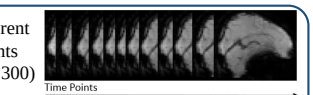
Methods

Acquisition

Rest-fMRI data from 30 healthy subjects (15 females and 15 males, age range: 20-30 years)

Processing

Trimming for 5 different numbers of time points (120, 150, 180, 240, 300)



- Motion correction (MCFLIRT tool)
- High-pass filtering (100 s)
- Spatial smoothing (FWHM = 5 mm)
- Registration to MNI152 standard space
- ICA-cleaning with FSL-FIX
- Dual regression with Shirer *et al* [1] template

Pearson's correlation between the time series extracted from dual regression

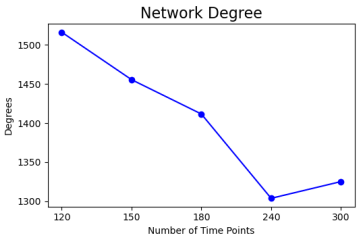
Threshold (± 0.35 correlation coefficient) and binarization for both positive and negative correlations

Graph Metrics:
- Small-Worldness: Sigma and Omega [2][3]
- Network Degree

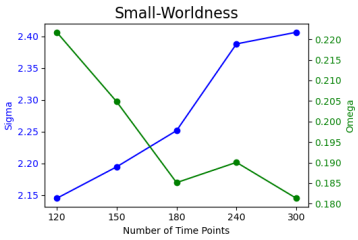
Sigma and Omega compare the clustering coefficient (C) and the characteristic path length (L) of a given network (g) with random (rand) and lattice (latt) networks. Sigma = $(C_g / C_{rand}) / (L_g / L_{rand})$; Omega = $L_{rand} / L_g - C_g / C_{latt}$

Sigma: > 1 → small-world | Omega: ≈ -1 → lattice; ≈ 0 → small-world; ≈ 1 → random

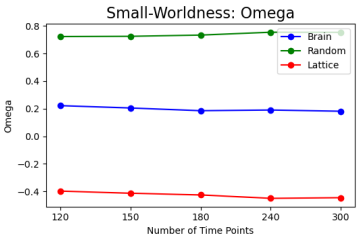
Results and Discussion



- There is a tendency of **decreasing the network degree** of the global brain network for more time points. This indicates that there is a loss of stronger correlations for more time points.



- The **sigma coefficient was greater than one** for all observations**,** and had an **increasing pattern**; The **omega coefficient** was positive for all observations and **trended to zero.** Both results suggest that brain tends to increase small-world characteristics with time.



- The **brain omega coefficient was smaller and greater** than the **random and lattice** omega coefficients, respectively. This demonstrates that, although there are changes, the small-world characteristics of the brain are maintained.

Conclusion

- Loss of stronger correlations with time
- Increase of small-world characteristics with time

The overall characteristics of brain networks are almost constant

Trimming leads to small differences in terms of functional connectivity metrics

References: [1] - Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD: Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb Cortex (2012); [2] - Humphries, MD et al, "Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence" PLoS One, vol. 3,4 (2008): e2051. [3] - Telesford, QK et al. "The ubiquity of small-world networks." Brain connectivity vol. 1,5 (2011): 367-75.

This work was financially supported by Fundação para a Ciência e Tecnologia (FCT) under the projects UID/BIO/00645/2019 and DSAIPA/DS/0065/2018.

133

BOLD signal data trimming results in increased functional connectivity

Saraiva D.[1], Ferreira H. [1]
[1]University of Lisbon, Faculty of Sciences, Lisbon, Portugal

INTRODUCTION

For performing robust and generalizable machine learning functional neuroimaging studies, combining multi-site data has been essential. Considering the impact of scan parameters on fMRI images, researchers usually trim multi-site data to the same number of time points. However, the effects of trimming BOLD signal data in terms of functional connectivity (FC) are still poorly understood.

METHODS

Resting-state functional MRI data from thirty healthy subjects were pre-processed for five different numbers of time points. Individual FC matrices were generated by performing dual regression with the brain template from *Shirer et al* [1]. The correlation matrices were binarized for several thresholds, which excluded weak correlations and included both positive and negative correlations for the analysis. To study FC differences between different numbers of time points, network degrees were computed for each brain region, and the degree distributions were compared. In addition, the median degree numbers per subject and per brain region were subtracted between different trimming conditions. Wilcoxon tests were then performed to evaluate which brain region degrees were significantly different between different trimming conditions.

RESULTS

Increasing degree numbers for the global network were observed, together with a non-significant right-shifting trend of the degree distributions, associated with fewer time points (Figure 1 left). Additionally, when comparing the degree distribution per brain region, it was observed that the majority of resting-state networks show increased degree numbers, whilst only the auditory network showed a decreased degree for fewer time points (Figure 1 right).

DISCUSSION

The overall increase in the degree numbers for fewer time points, suggests that there are fewer stronger correlations between brain regions with longer scanning times. In fact, some brain networks seem to present a more dynamic pattern than others, namely the higher visual, language, and left executive control networks. In contrast, the auditory, the visuospatial, the basal ganglia, and the dorsal default mode networks present a more stable and less dynamic pattern. This may be related to environmental demands (e.g. at rest during MRI scanning, visual stimuli) that engage some networks to change communication patterns more than others. Finally, although the overall characteristics of the global networks seem to be maintained, care should be taken to account for the trimming effects on FC at the regional level.

REFERENCES

[1] Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD: Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb Cortex (2012)

BOLD Signal Data Trimming Results In Increased Functional Connectivity

Duarte Saraiva[1], Hugo Ferreira[1]
1- Institute of Biophysics and Biomedical Engineering, Faculty of Sciences of the University of Lisbon

Introduction

Generalizable machine learning studies require multi-site fMRI data

Impact of scan parameters

How does it affect functional connectivity metrics?

Solution: **trimming!**

Methods

Rest-fMRI data from 30 healthy subjects (15 females and 15 males, age range: 20-30 years), with 300 time points of acquisition

Data was trimmed, in order to get five different number of time points conditions (300, 240, 180, 150, 120)

- Discard first 4 volumes
- Motion correction (MCFLIRT tool)
- High-pass filtering (100 s)
- Spatial smoothing (FWHM=5 mm)
- ICA-cleaning with FSL-FIX

Registration to the MNI152 standard space

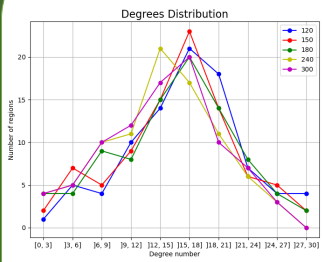
Dual regression with 88 regions from the Shirer *et al* [1] template

Pearson's correlation between the time series extracted from dual regression, followed by matrix binarization with a correlation coefficient of 0.35

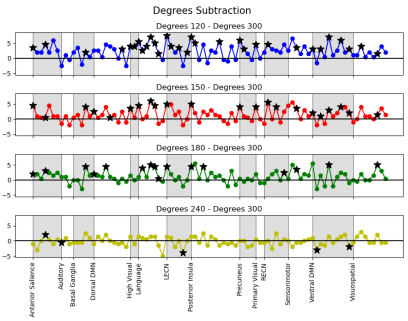
Graph metrics: network degree calculation for each subject's binarized correlation matrix

- Distribution of the degree number, per brain regions
- For each brain region, the median value of the number of degrees of all subjects was subtracted between different trimming conditions
- Wilcoxon tests were performed to evaluate which brain region degrees were significantly different between different trimming conditions

Results



For fewer time points, it is observed a non-significant right-shifting trend of the degree distributions, and an increase of the number of degrees



For fewer time points, the majority of resting-state networks showed increased degree numbers. Black stars represent brain regions that are significantly different

Conclusion

Degree numbers for fewer time points ↑ → Strong correlations between brain regions, for longer scanning times ↓

↑ higher visual, left executive control, language — dynamic

↓ auditory, visuospatial, sensorimotor, basal ganglia — dynamic

Although the overall characteristics of the global networks seem to be maintained, care should be taken to account for the trimming effects on functional connectivity at the regional level

References: [1] - Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD: Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb Cortex (2012);
This work was financially supported by Fundação para a Ciência e Tecnologia (FCT) under the projects UID/BIO/00645/2019 and DSAIPA/DS/0065/2018.