UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

Assessing Brain Functional Connectivity in Parkinson's Disease using Explainable Artificial Intelligence Methods

Leonor Maria de Oliveira Pires

Mestrado em Engenharia Biomédica e Biofísica

Dissertação orientada por:
Professor Doutor Alexandre Andrade

2023

Para o meu avô Zé Fábio e a minha gata Mimi,
com muita saudade e carinho no coração.

"Nothing in life is to be feared.
It is only to be understood.
Now is the time to understand more,
so we may fear less."

Marie Curie

Agradecimentos

Neste trabalho estão refletidos cinco anos de preserverança, dedicação e paixão. Nada disto seria possível sem mostrar a minha gratidão aos que me acompanharam, desde o início ou não, até ao fim, e que sempre estarão do meu lado para o que vier a seguir.

Em primeiro lugar, quero agradecer ao meu orientador Professor Alexandre Andrade, por me orientar e apoiar na aventura de trabalhar com dados da doença de Parkinson, mesmo que durante várias vezes ao longo do percurso parecesse praticamente impossível lidar com estes.

Agradeço ainda ao meu colega e amigo Henrique Lopes, o meu parceiro de tese, que várias vezes ouviu as minhas dúvidas, no trabalho ou em mim própria, e nunca deixou de me dar uma palavra amiga para me apoiar e ajudar a seguir em frente. Também agradeço à Thaïs, que esteve pelo IBEB apenas 3 meses, mas que demonstrou-se sempre disponível para ajudar com toda a simpatia.

Também não posso deixar de agradecer aos meus colegas *senpai* Duarte Saraiva e João Mendes, por sempre me esclarecerem as minhas questões. As suas palavras e ajuda foram indispensáveis para a concretização desta dissertação.

À minha família: os meus pais, o meu suporte e porto de abrigo, as minhas irmãs Marcela, Júlia, Amélia e Madalena, e a minha sobrinha Laura, as meninas do meu coração, para sempre a minha inspiração e quem espero orgulhar, o meu cunhado Nuno, que para mim é como um irmão mais velho, a minha tia Fabiana, que está longe mas sempre presente em mim para me apoiar, e os meus avós, as pessoas mais trabalhadoras deste mundo que sempre me motivaram a dar o meu melhor.

Aos meus animais de companhia, que à sua forma me apoiaram durante esta etapa: Mimi, Pickays e a jovem Blue.

Um grande obrigada à minha amiga do coração, Ana João. Pelo apoio incondicional, pelas longas chamadas ao telemóvel, e por me ajudares a acreditar em mim mesma. E ao *Shalom* mais um grande obrigada: Andreia, Brito, Chico, Dani, David, Diogo, Inês, Mariana, Pipa e Toni. As minhas amigas e os meus amigos, que nos juntámos tarde, mas que para sempre fiquemos unidos.

Agradeço ainda, por sempre ficarem do meu lado, os meus amigos algarvios Mariana e João Francisco.

Um agradecimento a todos os amigos que apareceram durante o meu percurso académico, desde a minha turma ao NE2B2.

Por fim, a ti, Diogo, o maior dos obrigadas não é suficiente. Por todo o amor, apoio incondicional e carinho. *Wherever you are in the world, I swear that I will find you again.*

Abstract

Parkinson's disease (PD) is a neurodegenerative disease characterised by dopaminergic neuron loss and α-synuclein accumulation. It exhibits both motor symptoms (such as tremors, bradykinesia, and rigidity) and non-motor symptoms. Diagnosis relies on clinical presentation and DaTScan, though their reliability varies. Functional magnetic resonance imaging (fMRI) and brain connectivity analysis have aided PD assessment. Studies have shown promise in diagnosing PD using deep learning (DL) but lack large-scale studies and transparency due to their black-box nature. Explainable AI (XAI) aims to provide understandable explanations for AI model decisions.

This dissertation proposes methods to assess functional connectivity in PD using a convolutional neural network (CNN) classifier and XAI.

Resting-state fMRI scans from the PPMI and ADNI data sets were pre-processed following an atlas composed of 14 resting-state networks. The FC matrices were computed through Pearson correlation coefficient and Fisher transform.

The FC matrices were fed to the ExtendedConnectomeCNN, optimised through 10-fold cross-validation, and tested, yielding a final model with 0.8214 accuracy, satisfactory performance metrics, balanced metrics, and interpretability potential.

Three XAI methods were applied: layer-wise relevance propagation (LRP), deconvolutional network (DeconvNet) and guided backpropagation. The LRP method provided more specific explanations, achieving higher AOPC value. Therefore, it is the method that better explains the classification of PD.

No basal ganglia network alterations were found, but changes in dorsal and ventral default mode, and posterior salience networks – involved in PD pathophysiology – were identified as potential biomarkers.

An attempt to perform transfer learning by training a model on the larger ABIDE set was executed. The model presented a poor performance and was not generalising, hence, we disregarded this possibility.

The approach to assessing functional connectivity changes in PD using XAI methods was fairly successful. The objectives of the dissertation were fulfilled, with hopes for contribution to novel PD diagnosis techniques.

Keywords: explainable artificial intelligence, Parkinson's disease, functional connectivity, resting-state networks, convolutional neural network

Resumo

A doença de Parkinson (DP) é uma patologia neurogenerativa caracterizada pela perda de neurónios dopaminérgicos, em particular nos gânglios da base, e acumulação da proteína α-sinucleína. A DP é caracterizada por quatro sinais cardinais motores: tremores, bradicinesia, rigidez muscular e instabilidade postural. A doença é também manifestada por sintomas não motores como perda do olfato, doenças neuropsiquiátricas como depressão e ansiedade, e distúrbios do sono.

Esta doença progressiva não tem cura, sendo que os tratamentos procuram a melhoria da qualidade de vida dos pacientes atenuando os sintomas. Relativamente ao diagnóstico, este é ainda principalmente baseado na análise da apresentação clínica dos sintomas. Entidades como a Sociedade de Doenças do Movimento apresentam uma série de critérios clínicos para aferir o diagnóstico da DP. Não existindo qualquer exame de imagiologia ou teste analítico que confirme um diagnóstico, as técnicas de neuroimagem surgem como ferramentas complementares com o fim de detetar alterações neuroquímicas relacionadas com a DP. O exame imagiológico mais comum é o DatScan, um tipo de aquisição de tomografia computorizada de emissão de fotão único que visa a deteção do transportador de dopamina, um biomarcador da degeneração dos neurónios dopaminérgicos. Dada a precisão e confiança insuficiente nos critérios clínicos de diagnóstico, bem como a falta de consistência do DaTScan, métodos de neuroimagem alternativos têm sido considerados para averiguar alterações cerebrais funcionais relacionadas com a DP, como por exemplo, a ressonância magnética (RM). Em particular, o fluxo sanguíneo cerebral e a conectividade do cérebro são analisadas através de RM funcional (RMf), uma técnica de RM que determina a atividade cerebral, em repouso ou perante uma tarefa, através da deteção de alterações no fluxo sanguíneo.

Deste modo, vários estudos têm apontado como uma potencial e inovadora abordagem a utilização de aprendizagem profunda (AP) para auxiliar e automatizar o diagnóstico de doenças neurológicas como a doença de Parkinson, baseando em dados de neuroimagem como a RMf. Não obstante, estas investigações ao nível da DP, AP e RMf não incluem, até ao momento e à luz do nosso conhecimento, estudos em larga escala: os números de sujeitos são ainda consideravelmente reduzidos, na ordem das dezenas. Ademais, os modelos de AP apresentam uma natureza de "caixa negra", ou seja, não é possível aferir de que forma o algoritmo chegou às decisões que levaram à classificação efetuada. Assim, a inteligência artificial explicável (IAE), um conjunto de métodos que pretende explicar e interpretar as decisões tomadas por modelos de inteligência artificial, surge como uma ferramenta apropriada para ultrapassar a falta de transparência dos modelos de AP.

Posto isto, o trabalho que surge no âmbito desta dissertação tem como objetivo o desenvolvimento de métodos para estudar e detetar alterações ao nível da conectividade funcional (CF) do cérebro relacionadas com a DP, recorrendo a um modelo de classificação baseado na arquitetura de redes neuronais convolucionais (RNC), e a métodos de IAE. Adicionalmente, pretende-se identificar potenciais biomarcadores funcionais da DP.

Para este fim, utilizaram-se aquisições de RMf do conjunto de dados do PPMI, que inclui 120 *scans*

de doentes com DP, e 22 de controlos saudáveis. Como este conjunto apresentava um desequilíbrio devido ao reduzido número de dados de controlos, recorreu-se ao conjunto de dados ADNI para recolher mais 131 *scans* de controlos. Este ajustamento foi efetuado considerando que a diferença entre os parâmetros de aquisição de RMf entre os dois consórcios, em particular o tempo de repetição, não leva a alterações significativas na avaliação da CF.

Os dados de RMf foram pré-processados de acordo com uma sequência de métodos que incluíram: realinhamento funcional e distorção, correção temporal, identificação de *outliers*, segmentação e normalização, e atenuação funcional. Foi ainda removido ruído dos dados, através da regressão de potenciais efeitos perturbadores e da aplicação de um filtro passa-banda entre os 0,008 Hz e os 0,09 Hz. Os dados foram segmentados de acordo com um atlas que inclui 14 redes neuronais de repouso. A conetividade funcional de cada sujeito foi aferida através do cálculo das matrizes de CF, que correspondem a matrizes de correlação entre as 14 redes funcionais de repouso. Para tal, foi aplicado o cálculo do coeficiente de correlação de Pearson e a transformada de Fisher.

As matrizes de conetividade foram inseridas numa RNC denominada de ExtendedConnectomeCNN, uma rede inspirada na ConnectomeCNN. Esta é composta por três camadas convolucionais e uma camada totalmente conectada. O tamanho da janela dos filtros é de 3 por 3 e o passo igual a 2. O número de filtros diminui ao longo das camadas convolucionais, de 256 para 128, e para 64. Em termos de parâmetros de treino, foram selecionados um número de épocas igual a 200 e um tamanho de grupo igual a 16. Como hiperparâmetros a otimizar, foram selecionados: a taxa de *dropout*, a taxa de aprendizagem, e a presença de uma camada de normalização em lote em cada camada convolucional. O processo de otimização dos hiperparâmetros foi efetuado através de validação cruzada com 10 *folds* (ou subconjuntos). Neste processo foi utilizado o conjunto de desenvolvimento dos dados, que corresponde a 90% do conjunto total das matrizes de CF. Da otimização de hiperparâmetros, foi selecionado o conjunto de hiperparâmetros que apresentou a melhor *performance*, isto é, com valores de médias das métricas de avaliação satisfatórios e balanceados. O conjunto com melhor *performance* apresentava uma taxa de *dropout* de 0,1 nas camadas convolucionais e de 0,4 na camada totalmente conectada, uma taxa de aprendizagem de 0,00001, e não tinha inseridas camadas de normalização em lote. Destacam-se os valores de exatidão de treino, 0,8814, de exatidão de validação, 0,7760, e de área sob a curva de característica de operação do receptor (AUC ROC) de 0,7496. Estes valores refletem modelos generalizáveis que detetam tanto as classes positiva (DP) como negativa (controlo).

Foi, de seguida, desenvolvido um modelo final com os melhores hiperparâmetros, treinado no conjunto de desenvolvimento e testado no conjunto de teste reservado à parte. Foram obtidas: uma extaidão de treino de 0,8776, exatidão de teste de 0,8214, e uma AUC ROC de 0,8230. Logo, o modelo construído apresenta valores de *performance* satisfatórios e balanceados, e potencial de interpretabilidade, o que permite a aplicação de métodos de IAE.

Ao modelo final foram aplicados três métodos de IAE: propagação de relevância camada a camada (do inglês LRP, *layer-wise relevance propagation*), rede de deconvolução, e retropropagação direcionada. Para cada método foi calculada a área de curva de perturbação do *mais relevante primeiro*, ou AOPC do inglês *area over the MoRF perturbation curve*, que avalia o quão relevantes são as explicações fornecidas pelos métodos de IAE. Considerando que o método LRP produziu mapas de explicação mais específicos e não dispersos, e que apresentou ainda valores de AOPC maiores e melhor distribuídos, considerou-se esse método como o que melhor explica a classificação de DP.

A partir das explicações fornecidas pelo método LRP foram extraídas as redes funcionais de repouso que mais relevância têm na classificação de DP. Não foram identificadas quaisquer alterações referentes à rede dos gânglios da base, apesar de tal ser esperado. No entanto, identificaram-se como potenciais

biomarcadores funcionais da DP as redes de modo padrão dorsal, de modo padrão ventral, e de saliência posterior, essencialmente envolvidas em manifestações não-motoras da doença.

Considerando que (1) o pré-processamento dos dados de RMf seguiu métodos adequados e produziu resultados satisfatórios, (2) o modelo de RNC para classificação de DP demonstrou ser suficientemente generalizável, com métricas de avaliação satisfatórias e equilibradas, e (3) a análise de IAE aparenta ser fidedigna e concordante com a literatura referente às alterações de redes funcionais de repouso perante a DP, conclui-se que a abordagem tomada para o estudo da CF relacionada com a DP utilizando métodos de IAE foi bem sucedida. Assim, os objetivos da dissertação foram cumpridos, com a expetativa de que este estudo resultará num progresso no desenvolvimento de técnicas inovadoras de diagnóstico de DP assistido por métodos de inteligência artificial.

Palavras chave: inteligência artificial explicável, doença de Parkinson, conectividade functional, redes funcionais de repouso, redes neuronais convolucionais

Table of Contents

List of Figures

List of Tables

Acronyms

AAL Automated Anatomical Labelling. 17, 52, 73, 76

ABIDE Autism Brain Imaging Data Exchange. iii, x, xii, 51–53, 73–76

AD Alzheimer's disease. 4

ADHD attention deficit hyperactivity disorder. 27

ADNI Alzheimer's Disease Neuroimaging Initiative. iii, v, ix, x, xii, 41, 42, 52, 53, 55, 59–61, 73, 74, 76, 77

AI artificial intelligence. iii, xii, 2, 3, 19, 27, 28, 31, 33, 34, 39–41, 51, 56, 60, 62, 75, 77

ALFF amplitude of low-frequency fluctuations. 17

AOPC area over the MoRF curve. iii, v, xi, 49, 50, 63, 67–69, 76

ART artifact detection tools. 43

ASD autism spectrum disorder. xi–xiii, 27, 38, 39, 51, 52, 73–75

ATP adenosine triphosphate. 12

AUC area under the ROC curve. v, ix, x, 25, 47, 56–59, 61, 73, 74

BOLD blood oxygenation level dependent. ix, 13, 15–17, 43–45, 72

CAD computer-aided diagnosis. 2

CAM class activation mapping. 38, 39

CGP cartesian genetic programming. 33, 34

CNN convolutional neural network. iii, x, 25–27, 29–32, 34, 38, 39, 45–48, 52, 53, 60–63, 68, 72, 75, 77

CNS central nervous system. 4

CompCor component based noise correction. 44

CSF cerebrospinal fluid. 15, 43, 44

CT computed tomography. 38

dACC dorsal anterior cingulate cortex. 19

DaT dopamine transporter. 9

DBS deep brain stimulation. 8

DCAE deep convolutional autoencoder. 31

DCM dynamic causal modelling. 33

DeconvNet deconvolutional network. iii, ix, 29, 30, 48–50, 62, 67–69

DI demographic information. xii, 34

DL deep learning. iii, 2, 19–21, 23, 25, 27, 32–34, 37–40, 47, 50, 75, 78

DLPFC dorsolateral prefrontal cortex. 18

DMN default mode network. 13, 33, 36, 37

DNN deep neural network. 2, 3, 25, 27, 28, 34, 42, 77

DTI diffusion tensor imaging. 32–34

EA evolucionary algorithm. 33, 34

EC Eigenvector centrality. 35

EEG electroencephalograhy. 14, 31, 61

EMG electromyography. 31

EPI echo-planar imaging. ix, 43

fALFF fractional amplitude of low-frequency fluctuations. 17

FC functional connectivity. iii, x, 1–3, 15, 16, 33–37, 45, 46, 48, 49, 52–55, 63, 68, 73, 76, 77

FEF frontal eye field. 19

FID free-induction decay. 11

fMRI functional magnetic resonance imaging. iii, x, xii, 1, 2, 10, 12, 13, 15, 16, 18, 19, 27, 31–35, 38–42, 44, 52–55, 59, 61, 73, 75–77

FWHM full-width half maximum. 44

GABA gamma-aminobutyric acid. 5

GAT graph attention network. 32–34

GCN graph convolutional network. 34

GLM general linear model. 44

Grad-CAM gradient-weighted class activation mapping. 38

GRE gradient recalled echo. 13

ICA independent component analysis. 17, 36

IPS intraparietal sulcus. 19

L-DOPA L-3,4-dihydroxyphenylalanine. 8

LB Lewy bodies. 6, 7

LogReg logistic regression. xii, 34

LRP layer-wise relevance propagation. iii, v, 28, 29, 38, 39, 48–50, 62, 67–69, 73, 76

LSTM long-short-term memory. 33, 34, 39

MDS Movement Disorders Society. 8

MDS-UPDRS Movement Disorders Society - Unified Parkinson's Disease Rating Scale. 33

MEG magnetoencephalography. 15

MIBG metaiodobenzylguanidine. 9

ML machine learning. 2, 19, 31–33, 38

MNI Montreal Neurological Institute. 15

MoRF most relevant first. v, xi, xiv, 49, 50, 67, 68

MPFC medial prefrontal cortex. 18

MRI magnetic resonance imaging. 1, 10–13, 31–33, 38, 39, 41, 51, 61, 77, 78

MTL medial temporal lobe. 19

NMR nuclear magnetic resonance. 10, 11

NN neural network. 20, 22

PCC posterior cingulate cortex. 18

PD Parkinson's disease. iii, xii, 1–10, 19, 31–42, 45, 47, 49, 51, 56, 57, 60–63, 68–73, 76–78

PET positron emission tomography. 38, 41

PNS peripherical nervous system. 4

PPMI Parkinson's Progression Markers Initiative. iii, iv, ix, x, xii, 32, 33, 41, 42, 51–53, 55, 59, 61, 73, 74, 76

PReLU Parametrized Rectified Linear Unit. 21, 47

ReLU Rectified Linear Unit. ix, 20, 21, 23, 30, 46, 47

REM rapid-eye movement. 7

RF radio-frequency. 11

RNN recurrent neural network. 33

ROC receiver operating characteristic. v, x, 25, 47, 56–59, 61, 73, 74

ROI region of interest. x, 16–19, 33, 39, 44, 53–55

rs-fMRI resting-state functional magnetic resonance imaging. 2, 13, 14, 32–36

RSC retrospenial cortex. 19

RSN resting-state networks. ix, 13, 14, 34, 35, 39

SCA seed-based correlational analysis. 16

SE spin echo. 13

sMRI structural magnetic resonance imaging. 32, 34, 41, 42

SNpc substantia nigra pars compacta. 6, 7

SPECT single-photon emission computed tomography. 9, 31, 76

SPM Statistical Parametric Mapping. 42, 43, 45

SVM support vector machine. 32–34, 38, 39

TE echo time. ix, 11, 12, 42

TL transfer learning. xii, 34, 76

TR repetition time. ix, 11, 12, 41, 42, 73

V1 primary visual cortex. xi, xiii, 18, 97, 99

V2 secondary visual cortex. xi, xiii, 18, 97, 98

XAI explainable artificial intelligence. iii, ix–xi, 2, 3, 19, 27, 28, 30, 31, 34, 37–41, 46–50, 53, 56, 57, 60–69, 72, 73, 76, 77

Chapter 1

Introduction

This dissertation project proposes a method for the study of brain functional connectivity changes in Parkinson's disease, using explainable artificial intelligence methods to identify those alterations. The context and motivation, the objectives, and the workflow structure will be presented in this chapter.

1.1 Context and Motivation

The first time I came across the role of a biomedical engineer was in 2019 when I was attending a deep brain stimulation procedure on a patient with Parkinson's disease (PD). I remember observing that patient with severe tremors in the lower limbs and reflecting on how difficult would be to live with such impairment and how this issue could be tackled. Afterwards, when the biomedical engineer in the operating room began to input electric current to the electrodes placed on the patient's brain, particularly on the left hemisphere, her right leg immediately stopped shaking. This was the moment when I realised that a biomedical engineer has a fundamental role in medicine, by finding methods to enhance diagnosis and treatment approaches to several pathologies like PD.

PD has an incidence of 1 to 2 per 1,000 people worldwide, and the estimated total number of cases in Portugal is 180 per 100,000 people, in 2017 [1, 2]. This movement disorder is common among the elderly, affecting approximately 1 % of the population over 60 years old [3]. Since PD affects mostly elderly people and considering that the world population is ageing, namely in high-income countries, it is expected to observe an increasing incidence and prevalence of PD [4]. In fact, according to the World Health Organisation, for the last 25 years, the number of PD cases has doubled and PD-related disability and mortality has been growing as well [5]. A comprehensive study indicates that PD affected 6.1 million people worldwide in 2016, up from 2.5 million in 1990, turning PD into the neurologic disease with the fastest-growing cause of disability [6].

As we notice, Parkinson's disease is a pathology with increasing prevalence and incidence due to population ageing. Considering this, it is important to enlarge research on this disease and to explore new and innovative approaches to improve existing diagnosis and treatment strategies.

Nowadays, the diagnosis of neurological disorders reveals itself as a great challenge for modern medicine [7]. The acquisition of biomedical images, such as magnetic resonance imaging (MRI), has been useful in aiding the detection, monitoring and treatment of neurological diseases like PD. In particular, there has been a growing interest in analysing functional magnetic resonance imaging (fMRI) in order to assess how the functional connectivity (FC) of the brain changes upon those pathologies, and to find new pathological biomarkers. However, the amount of produced data is significantly large and it is not always trivial to interpret and analyse them manually. Taking this issue into account, computer-

aided diagnosis (CAD) techniques emerged in the early 1980s [7]. These methods detect neurological abnormalities through three main steps, pre-processing, feature extraction and classification, aiming to enhance diagnosis consistency and increase the efficacy of treatment [7].

Machine learning (ML) has been playing an important part in CAD systems. ML uses previously extracted features from the medical imaging data and describes patterns used for interpretation. Despite being an useful strategy, ML implies feature selection, which is not a straightforward task in the case of medical images due to the large quantity of data information, especially for non-experts [8]. Deep learning (DL) emerges then as a suitable alternative since it does not require professional knowledge to select features. DL automatically finds useful representations for the task in consideration [8]. Because of that, and also considering the advances in the availability of data and improvement of computer technology, DL has been chosen as the preferred method of medical imaging analysis for numerous tasks such as classification, detection, registration, and segmentation of those images [8, 9, 10, 11]. In particular, DL has been applied to break new ground in the classification of different neuroimages to detect neurological diseases such as Alzheimer's disease, mild cognitive impairment, attention deficit hyperactivity disorder, autism spectrum disease, schizophrenia, and, of course, Parkinson's disease [8, 12]

Notwithstanding the growth of DL in the context of medical image analysis, this approach comes with a problem: the black-box nature. Explaining the decisions that deep neural networks present is not immediate and requires high knowledge of their internal operation, and the end-users of those tools, such as healthcare professionals, lack that understanding [13]. Briefly, explainability, transparency and accountability are key issues in AI nowadays not only in the medical field but in all areas. To overcome this challenge, another field of artificial intelligence (AI), explainable artificial intelligence (XAI), has been used to produce interpretable, intuitive, and human-understandable explanations of AI decisions [13].

Considering the above, it appears to exist an opportunity to use XAI algorithms on DL-based CAD techniques of neurological diseases, by extracting the explanations of the decisions given by the neural network applied to neuroimaging data. In fact, researchers in medical imaging started to apply XAI more frequently to interpret their results [14]. And why should not this be applied to Parkinson's disease, to better understand the pathological mechanisms behind this neurological disorder, and to break new ground on diagnosis methods?

1.2 Objectives

The main objective of this dissertation project is to study functional connectivity (FC) changes due to Parkinson's disease based on resting-state functional magnetic resonance imaging (rs-fMRI) data, by applying explainable artificial intelligence methods. As additional goals, it is proposed the following:

- Review of the existing approaches on AI-based PD diagnosis, the study of FC changes in PD, and the use of XAI on neuroimaging;
- Application of adequate pre-processing methods on functional magnetic resonance imaging (fMRI) data;
- Development of a deep neural network (DNN) for classification task between PD patients and healthy subjects;

- Use of multiple XAI methods to extract the explanations of the DNN decisions, to compare them and assess brain FC alterations.

1.3 Structure

The dissertation is divided into six chapters. The first chapter introduces the topic to study, by presenting the context and motivation for the development of this work. Moreover, the objectives and structure of the dissertation are mentioned. Chapter 2 presents the scientific background behind the dissertation topic. Chapter 3 includes the state-of-the-art regarding AI-based PD diagnosis, the study of FC on PD and the use of XAI on neuroimaging. In Chapter 4, the materials and methods used in this project are described, while Chapter 5 presents the obtained results and respective analyses and discussions. Finally, a conclusion on the most relevant findings of this project and suggested future work are given in the sixth and last chapter of the dissertation.

Chapter 2

Scientific Background

2.1 Parkinson's Disease

Parkinson's disease is an idiopathic movement disorder and the second most common neurodegenerative disease, following Alzheimer's disease (AD) [15]. Described by James Parkinson in 1817, PD is caused by the progressive death of dopaminergic neurons in the brain and by the accumulation of the protein α-synuclein. This pathology is expressed by four cardinal signs: tremors, bradykinesia, rigidity and postural instability. Although the majority of Parkinson's disease cases are sporadic, there may be particular types of susceptibility genes that raise the likelihood of contracting the illness [16].

To give further context on the involved physiologic mechanisms and manifestations of PD, this section will provide some basic concepts on the pathological process and subsequent clinical symptoms and signs. Moreover, the current diagnosis approaches will be presented, and the related challenges and perspectives as well.

2.1.1 Pathophysiology and Clinical Presentation

Before referring to the pathophysiological mechanisms of Parkinson's disease, it is crucial to know which brain regions and systems are involved in those phenomena.

The nervous system is divided into two major groups: the central nervous system (CNS) and the peripherical nervous system (PNS), which includes the sensory and motor components. The CNS is defined as the brain (pons, cerebellum, midbrain, diencephalon, medulla and cerebral hemispheres) and the spinal cord (see Figure 2.1). In general terms, the functions of CNS are the detection, analysis, and transmission of information [16].

The brain and spinal cord coordinate the spatial and temporal patterns of muscular contractions to produce movements, whether they are voluntary or involuntary. The lower motor neurons in the spinal cord and brainstem directly innervate skeletal muscles causing all motions to be produced by the musculature, while the innervation of visceral smooth muscles is organised separately by the autonomic divisions of the visceral motor system [16]. Lower motor neuron activity is coordinated by the activity of upper motor neurons, which affects the local circuits in the brainstem and spinal cord that control movement. Additionally, the basal ganglia and cerebellum also play a relevant role in motor control: they influence movement by regulating the activity of upper motor neurons.

Parkinson's disease is one of the deficits that result from damage to the component nuclei of the basal ganglia. The basal ganglia are located in the cerebrum and include nuclei embedded deep in the

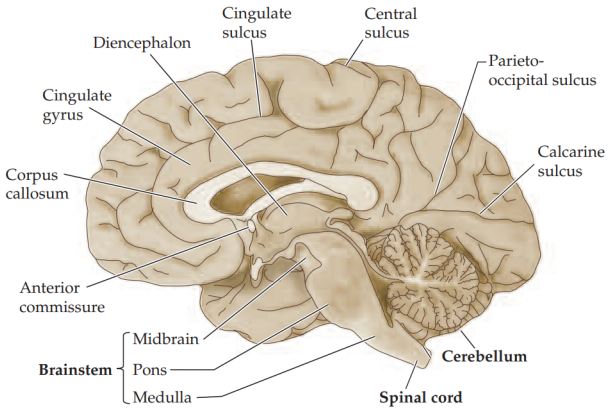

Figure 2.1: Structures of the brain, including the brainstem (midbrain, pons and medulla) and the cerebellum (from [16]).

brain hemispheres (striatum - caudate nucleus and putamen - and globus pallidus). Additionally, the related nuclei consist of structures located in the diencephalon (subthalamic nucleus), mesencephalon (substantia nigra), and pons (pedunculopontine nucleus) [17]. Figure 2.2 displays a coronal plane cut of the brain, highlighting the structures related to the basal ganglia.

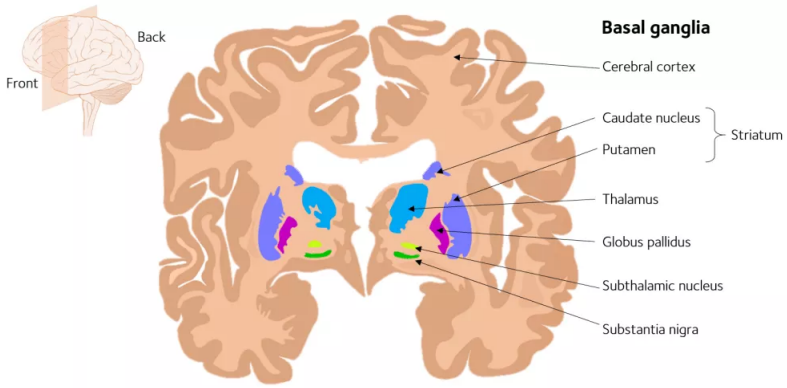

Figure 2.2: Basal ganglia related brain structures, shown in a coronal plane cut (from [18]).

The basal ganglia include three major pathways: the basal ganglia circuits [1]. The cortical–basal–ganglionic–thalamic–cortical loop has inputs from the premotor, primary motor and primary sensory cortices that project to the corpus striatum, which sends fibres to the globus pallidus. Moreover, the substantia nigra sends dopaminergic fibres to the corpus striatum, forming a second loop. Finally, the third loop includes reciprocal connections between the globus pallidus and the subthalamic nucleus.

Basal ganglia circuits present an important role in the onset, amplitude and speed of movements. Thus diseases related to the malfunctioning of the basal ganglia lead to aberrant motor system activation, which causes stiffness, tremors, involuntary movements, or motor deficiencies (e.g. bradykinesia and postural instability) [1].

The most significant neurotransmitters that are involved in the basal ganglia circuits are acetylcholine and dopamine. Acetylcholine is present in high concentrations in the corpus striatum and acts as an excitatory transmitter at striatal neurons that synthesise and release the inhibitory neurotransmitter gamma-aminobutyric acid (GABA). Dopamine is a catecholamine neurotransmitter synthesised in substantia nigra, whose axons form the nigrostriatal pathway [19]. Dopamine, on the other hand, inhibits striatal GABAergic neurons. This neurotransmitter system is affected in PD.

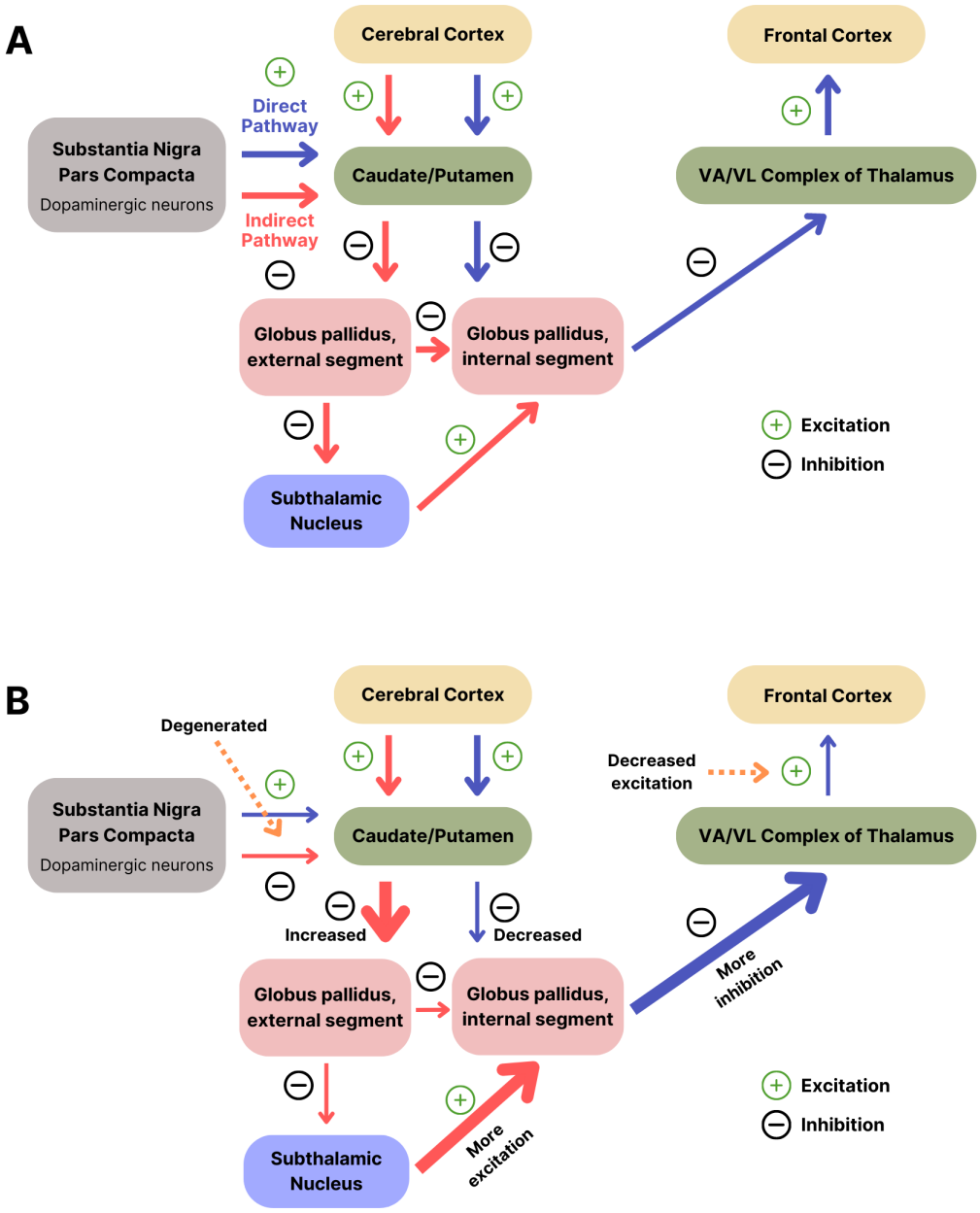

Figure 2.3: A. Normal basal ganglia dopaminergic circuit; the dopaminergic neurons in the direct pathway present an excitatory behaviour in the caudate/putamen, while in the indirect pathway, there is an inhibitory relation. B. Basal ganglia dopaminergic circuit in Parkinson's disease. VA: ventral anterior; VL: ventral lateral (based on [16]).

The pathological phenomena involved in PD are still unclear and complex to comprehend. In simple terms, PD has two pathological hallmarks: (1) the degeneration of neurons in the substantia nigra pars compacta (SNpc); (2) the aggregation of the protein α-synuclein protein in Lewy bodies (LB).

The degeneration of SNpc neurons is related to the decrease of dopamine release in the striatum [20]. It is estimated that PD patients with motor symptoms have 50 to 80 % of nigral dopaminergic neurons degenerated [21]. Normally, the dopaminergic neurons of the SNpc input to the striatum leads to excitation of the medium spiny neurons that project directly to the internal globus pallidus and inhibition of the spiny neurons that project to the external globus pallidus cells in the indirect pathway of the basal ganglia (see Figure 2.3). Consequently, both dopaminergic effects lead to a decrease of the inhibitory

outflow of the basal ganglia and thus to increase the excitability of the upper motor neurons. However, as the dopaminergic neurons degenerate, the inhibitory outflow of the basal ganglia is abnormally high, and then the thalamic activation of upper motor neurons in the motor cortex is less likely to occur [16]. Due to this, the cardinal manifestation of the disease (tremors, rigidity, bradykinesia and postural instability) emerge [1, 20]. Aggregation of Lewy bodies is identified inside and outside of the SNpc. Studies report that α-synuclein accumulation begins in the enteric nervous system and the vagus nerve and then spreads to the dorsal motor nucleus of the vagus in the brain stem, and the rest of the brain [20]. Moreover, LB aggregation is often linked to genetic mutations and to exposure to substances like pesticides [16, 22, 23].

PD pathology is linked to a staging system introduced in 2003 by Heiko Braak [24]. This system proposes an assessment of PD according to LB disease evolution based on autopsy findings. Braak staging of PD includes six stages [25]:

- Stages 1 and 2: accumulation of LB mainly in the dorsal motor nucleus. reticular formation and anterior olfactory nucleus; asymptomatic or presymptomatic motor phase and non-motor signs (autonomic, olfactory and sleep disorders);
- Stage 3: LB disease and loss of neurons in SNpc, and extension to locus coeruleus and amygdala; beginning of motor signs onset;
- Stage 4: LB disease extends to the temporal limbic cortex; motor symptoms emerge;
- Stages 5 and 6: LB involvement on the prefrontal cortex and primary sensory and motor areas; severe parkinsonism (gait impairment and dementia).

The Braak staging infers that the evolution of the disease manifestation is related to the progressive accumulation of Lewy bodies [25]. This hypothesis is controversial, yet significant since it introduces the prodromal manifestation related to PD.

Notwithstanding that, the clinical presentation of Parkinson's disease is widely variable among patients. Nonetheless, as aforementioned, PD is marked by four cardinal symptoms: tremor, bradykinesia, rigidity and postural instability, although the last one is often not considered a cardinal symptom [26]. This clinical syndrome is usually referred to as parkinsonism. PD is the most common cause of parkinsonism [1].

A tremor in PD is not triggered by any purposely engaged physical activity: the trembling body part is only supported by gravity (resting tremor). When the affected body part is actively moving, a rest tremor frequently dampens or completely disappears [27]. The rest tremor can be identified on a resting limb and disappears when the limb is held outstretched [28]. Bradykinesia (slowness of movements) is usually detected through the repetition of a task like tapping index finger and thumb or opening and closing fist [28]. Muscular rigidity is described as an unconscious, velocity-independent resistance to an examiner's passively moving a joint, and postural instability corresponds to impairment of balance that disturbs posture control during, for example, standing or walking positions [28].

Besides the parkinsonism cardinal signs, other motor symptoms may emerge: gait disturbance, impaired handwriting, precision grip impairment, and speech deficits [29].

PD is mostly acknowledged due to the characteristic motor symptoms. Nevertheless, there is a set of non-motor symptoms that should not be depreciated. PD is also expressed by anosmia (partial or complete loss of the sense of smell), impaired colour vision, neuropsychiatric disorders, such as depression and anxiety, constipation and urinary dysfunctions, and sleep disturbances such as rapid-eye-movement (REM) sleep behaviour disorder [30, 31]. Parkinson's disease can lead to early

cognitive dysfunction and eventually to dementia [30]. These autonomic, olfactory, psychiatric and sleep disturbances often appear during the pre-symptomatic stages of PD and are considered important prodromal markers [32]. These symptoms are not normally associated with Parkinson's disease, which may hamper the early diagnosis of this pathology.

Parkinson's disease is a progressive neurological disorder. As such, symptoms progressively get worse over time. PD symptoms affect immensely the daily basis of patients. Therefore, and also because there is no cure for this pathology, treatment approaches aim to improve patient's quality of life by easing symptoms [19].

The most used pharmacological approach for PD treatment is the dopamine precursor L-3,4-dihydroxyphenylalanine (L-DOPA), or levodopa [33]. This drug aims to increase the amount of dopamine in the striatum. Dopamine can not be administered directly, since it does not cross the brain-blood barrier [34]. Levodopa is often combined with other drugs that minimise its metabolism and the degradation of dopamine [19]. However, as PD progresses, the efficacy of levodopa decreases, since it requires a minimum number of dopaminergic neurons to have an effect. Alternative drugs, like pramipexole and ropinirole, activate postsynaptic dopamine receptors [19, 35, 36].

There are additional medications that target the cholinergic system (such as cholinesterase inhibitor rivastigmine) and the noradrenergic system (norepinephrine transporter ligand amantadine), which are also involved in the pathophysiology of PD [19].

Moreover, a commonly used surgical strategy is deep brain stimulation (DBS), normally eligible for PD patients suffering from motor fluctuations refractory to optimal medical treatment [37]. DBS resorts to electrode implantation on the brain to induce neuromodulation, synaptic plasticity and neuroprotection, leading to motor function improvement [19]. DBS, in the case of PD treatment, usually targets the subthalamic nucleus or globus pallidum internus in order to improve motor functions in patients, however, it may cause a decline in specific cognitive domains [38]. A study concluded that DBS treatment shows approximately 32% motor improvement in patients with a 10-year-old PD diagnosis [39].

As previously referred, each PD patient displays a distinct manifestation of the disease. Hence, each treatment plan must be designed according to their motor and non-motor symptoms. Additionally, we cannot jeopardise the possible differential diagnosis of parkinsonism syndrome. If the treatment approaches are adequate for the actual pathology, symptom improvement will be seen. Therefore, addressing the patient's symptoms and optimising diagnostic tools and criteria of PD is crucial not only to minimise misdiagnosis cases but also to enhance the patient's treatment and to allow an early diagnosis of the pathology.

2.1.2 Diagnosis

Nowadays, the diagnosis of Parkinson's disease still mainly relies on the assessment of clinical features. The Movement Disorders Society (MDS) presented clinical diagnostic criteria for PD for research and clinical use [40].

The most important feature to consider is the manifestation of parkinsonism, characterised by the presence of bradykinesia and the presence of tremor or muscle rigidity [40]. Parkinsonism is a necessary criterion to take into account in a clinical diagnosis of PD.

Clinically established PD is also marked by at least two supportive criteria. These supportive criteria include the observation of response to dopaminergic therapy, with marked improvement of symptoms with an increase in dose. Furthermore, the presence of a rest tremor of a limb, often unilateral, is

considered a supportive criterion. Other supportive criteria are olfactory loss or cardiac sympathetic denervation detected through metaiodobenzylguanidine (MIBG) scintigraphy and levadopa-induced dyskinesia [40].

To clinically diagnose PD, no absolute exclusion criteria nor red flags must be identified. These exclusion criteria are referred to in Figure 2.4. Figure 2.4 also contains part of the main supportive criteria.

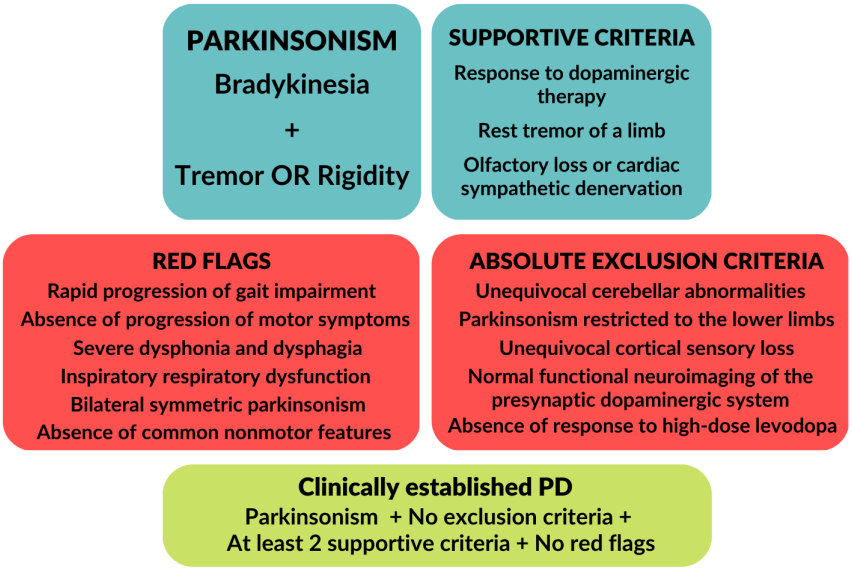

Figure 2.4: MDS-PD Diagnostic Criteria (based on [40]).

Clinical evaluation of symptoms is the main approach to diagnosing PD. There are no discriminating imaging exams or blood tests that confirm a diagnosis based on the clinical history and neurological examination. Currently, neuroimaging techniques are used as complementary tools to detect neurochemical changes in the brain related to Parkinson's [19].

The most popular biomedical imaging technique used for the detection of PD-related brain alterations is single-photon emission computed tomography (SPECT), a method of nuclear imaging that uses radionuclides emitting single gamma photons.

In the context of PD, there are specific targets of interest to study through neuroimaging. Regarding the dopaminergic system, which is impaired in PD patients, as we have seen, the most common targets are dopamine synthesis (related to the loss of nigrostriatal dopaminergic nerve endings), dopamine transporter, vesicular monoamine transporter and dopamine receptors [19]. Targets on the serotonergic system (serotonin receptors and transporters) and the cholinergic system (as vesicular acetylcholine transporter and norepinephrine transporter) are also relevant to detect brain changes related to PD.

The most popular imaging exam used to support the diagnosis of PD and other parkinsonism-related diseases is a SPECT scan with ioflupane injection named DaTScan^{TM} [41]. A biomarker for nigrostriatal dopaminergic degeneration, the dopamine transporter (DaT), is largely found in presynaptic dopamine nerve terminals. Within 3 to 6 hours of injection, SPECT imaging is possible due to the radioligand ioflupane's distribution to the striatum, where it binds to the presynaptic DaT protein with great affinity [41]. Since the current clinical approach to diagnosing PD lacks accuracy, European centres have been resorting to DaTScan, more frequently [42]. Nonetheless, there is still room for improving the validation of a PD diagnosis based on DaTScan. In clinical practice, DaTscan image interpretation is based on

visual judgment, however, quantitative approaches are on the rise [41].

However, these diagnosis tools come with a problem. PD is the most common cause of parkinsonism, but other pathologies present this set of clinical symptoms. It is estimated that one in four patients who receive a PD diagnosis during life suffer from an alternate pathology, diagnosed at post-mortem [43]. Other parkinsonism causes include dementia with Lewy bodies (patients present early dementia [44]), multiple system atrophy (marked by autonomic failure, including prodromal autonomic symptoms, which mainly affects urogenital and cardiovascular systems [45]), progressive supranuclear palsy (caused by the accumulation of neurofibrillary tangles, this pathology is characterised by the presence of lurching gait and falls without loss of consciousness [46]) and corticobasal syndrome (characterised by progressive aphasia or frontotemporal dementia, with motor and cognitive manifestations [47]).

Considering the above, the clinical evaluation for Parkinson's diagnosis must take into account other possible causes of parkinsonism, in order to avoid misdiagnosis. Nevertheless, due to the inherent subjectivity of symptom evaluation, it is not expected to observe a satisfactory rate of correctly diagnosed PD cases. According to a 2016 meta-analysis that included 11 different studies between 1988 and August 2014, the pooled diagnostic accuracy of PD using standard examination was nearly 81 % [48]. This study also reflects that there was no improvement in the accuracy of PD diagnosis in 25 years.

Moreover, DaTScan imaging cannot consistently differentiate between PD and other parkinsonism diseases, such as multiple system atrophy or progressive supranuclear palsy [42]. Additionally, there is a concern regarding the reliability of DatScan. An investigation inferred that in approximately 50 % of the cases when DaTScan was repeated, there were disparities between the first and second readings [49].

In light of the presented challenges, alternative neuroimaging techniques such as magnetic resonance imaging and respective advanced methods have been explored to also assess brain function alterations linked to PD. In particular, cerebral blood flow and brain connectivity are analysed through fMRI, a MRI technique that assesses brain activity by detecting blood flow changes [19].

2.2 Brain Connectivity and Functional Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) emerged in the 1970s, but it is based on the nuclear magnetic resonance (NMR) process, first described in 1946 by Bloch and Purcell, whose discoveries resulted in a Nobel Prize in 1952 [50, 51]. It has become an important imaging method, since it is considered to be safe and provides an image with high contrast between different tissues, compared with other techniques such as radiography or computed tomography.

The use of MRI in the neuroimaging field has been showing important relevance in understanding brain circuit changes related to the aetiology, pathophysiology and treatment response of neurological diseases [52]. In particular, functional magnetic resonance imaging, a sub-technique that appeared in the 1990s, detects blood flow alterations related to specific brain activity. In light of this, it seems to be possible to study the abnormal functional activity of brain networks, leading to aberrant brain connectivity, associated with different neuropathologies [53].

In this section, the basic principles of MRI and fMRI will be presented, in order to summarise the theoretical background behind these imaging techniques. Furthermore, it will be clarified the concept of brain connectivity and respective methods of computation and interpretation.

2.2.1 Magnetic Resonance Imaging: Basic Principles

As previously mentioned, MRI is the result of experiments and manipulation of the NMR phenomenon. MRI targets the hydrogen proton, which is a nucleus abundantly present in human tissues.

As commonly known, a nucleus is characterised by mass and charge. But there is one more property that a hydrogen nucleus (hereby referred to as protons) carries: spin. Spin can be mathematically perceived as a rotation on the proton's own axis, although the spinning does not happen in the classical meaning [51]. Consequently, a proton acts as a small magnet, having a magnetic moment and an angular momentum. Due to these properties, when a nuclear spin interacts with a strong magnetic field B_0, the proton will tend to adopt one of two possible orientations: parallel or antiparallel with B_0 [54]. The slightly larger number of protons in the parallel direction will give rise to a net magnetisation. The proton axis, on the other hand, will start precessing around B_0. The frequency of precession is given by the Larmor equation [50, 51, 54], which describes that the frequency of precession ω_0 is proportional to the magnetic field of magnitude B_0, by a factor of λ (the gyromagnetic ratio, a constant fixed for each nucleus):

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

A proton with spin within the magnetic field B_0 is excited when a radio-frequency (RF) magnetic field (B_1) is applied in short pulses (with a duration of microseconds), in a transversal direction to B_0 [50]. The application of a RF pulse with the frequency of Larmor results in the transition of protons from the parallel to the (more energetic) anti-parallel orientation. Simultaneously, the net magnetisation starts precessing along the direction of the field associated with the RF pulse (B_1), hence giving rise to a transverse component of the magnetisation. This precessing transverse component is the source of the magnetic resonance signal: the ensuing variable magnetic flux will be picked up by the coil as a time-varying signal called free induction decay (FID). When the RF pulse is switched off, protons will tend to return to the parallel orientation (regenerating the longitudinal magnetisation) and, independently, the transverse component will start decaying due to interactions between individual protons. A MRI image is obtained by the application of multiple RF pulses, resulting in multiple FIDs that can be resolved by the application of the Fourier transformation [51].

Protons eventually stop precessing around B_1, since they lose the energy supplied to precess on the transversal plane. Hence, protons start to realign with the external magnetic field, while axial magnetisation reappears. The time that approximately 63 % of the protons take to realign with B_0 is T1. This is referred to as T1 relaxation (or spin-lattice relaxation) and it influences the obtained contrast in a T1-weighted MRI image [51, 54]. On the other hand, the rate of reduction in transverse magnetisation is characterised by the time constant T2, corresponding to the T2 relaxation (or spin-spin relaxation). However, in practice, the dephasing of transverse magnetisation is altered by inhomogeneities in the B_0 field. Considering this, we have now the constant $T2^*$, which takes into account the internal and external relaxation rates (introduced by the inhomogeneities) through the following equation [50]:

$$\frac{1}{T2^*} = \frac{1}{T2} + \frac{1}{T2'} \tag{2.2}$$

Additionally, there are other two time parameters, defined by the MRI technician, to adjust the desired image contrast, depending on the tissues supposed to be targeted: the echo time (TE) and the repetition time (TR). TE includes the time between the RF excitation and the respective echo, and TR is the time duration between successive RF pulses [51, 55]. TE and TR are manipulated to set the weight of T1, T2,

or proton-density weighting (see Figure 2.5) [55]. T1-weighted images can be acquired using relatively short TE and short TR, while T2-weighted images correspond to long TE and long TR. Proton-density-weighted images, related to the number of nuclei in the image region, are obtained using short TE and long TR.

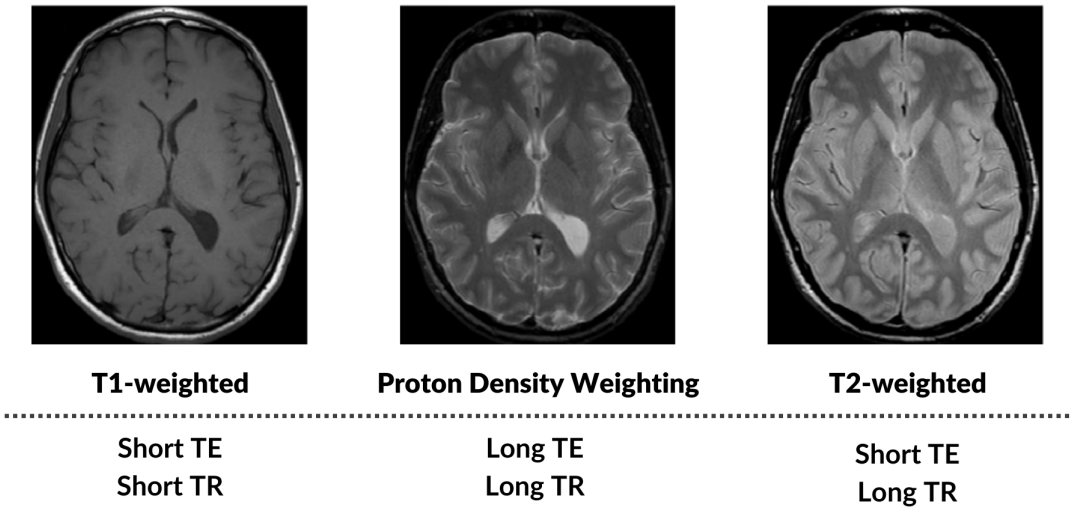

Figure 2.5: Echo time (TE) and repetition time (TR) settings for T1-weighted, proton density weighting, and T2-weighted images (based on [55] and adapted from [56]).

To fully obtain a MRI image, it is crucial to localise spatially the resulting signal. To perform this, it is needed to resort to gradients, which are additional fields that vary in B_0 according to their spatial location. They are applied in the three orthogonal directions using the three sets of gradient coils, so we have G_x, G_y, and G_z. Consequently, the precession frequency measurements can be used to distinguish obtained signals at different spatial positions, allowing the 3D image reconstruction [51, 54].

From a MRI scanner, we can obtain two types of acquisition: structural MRI and functional MRI. With structural MRI we receive structural information related to the brain, in particular the shape, size and integrity of brain regions. With functional MRI, or fMRI, we detect changes in brain activity by alterations of blood circulation. In light of this dissertation's goals, it is crucial to better understand what is behind the acquisition of fMRI data.

2.2.2 Functional Magnetic Resonance Imaging

As with any other human cells, the processes involved in the neural signalling of brain cells demand the use of energy in the form of adenosine triphosphate (ATP), obtained through the process of oxygenation of glucose in the mitochondria [57]. Neuron firing follows any cognitive task performed by the human, therefore, it is expected to observe a higher demand of energy on the correspondent brain region. Due to this requirement for energy, brain cells need more oxygen from blood: levels of deoxygenated haemoglobin (or deoxyhaemoglobin) increase, while levels of oxygenated haemoglobin (or oxyhaemoglobin) decrease [57, 58, 59]. As a compensatory mechanism, cerebral blood flow will increase in the respective brain region to reestablish blood oxygen levels. However, for not-so-understood reasons, the response to reverse the decrease of oxygenated haemoglobin surpasses the baseline levels.

So, the decrease of deoxygenated haemoglobin and increase of oxygenated haemoglobin will transcend the resting levels [57, 58, 59, 60].

As a consequence of these processes involved in brain activation, we will observe the blood oxygenation level-dependent (BOLD) effect as brain activity increases [61]. This effect is based on the magnetic properties of oxyhaemoglobin and deoxyhaemoglobin. Oxyhaemoglobin is a diamagnetic molecule, meaning that it is not distinguishable from brain tissue when exposed to a magnetic field. On the contrary, deoxyhaemoglobin is paramagnetic, resulting in haemoglobin-concentration-dependent gradients on the surrounding magnetic field [62]. From these gradients, two types of BOLD contrast can be obtained: $T2^*$-weighted signal, which occurs due to intravoxel dephasing and can be observed through gradient recalled echo (GRE) imaging, and T2-weighted signal, due to the diffusion of spins through the gradients [57, 60] and detected by spin echo (SE) MRI. The weighting of T2 contrast increases relative to $T2^*$ weighted contrast as the magnetic field strength rises. Nowadays, GRE sequences with $T2^*$ contrast, especially fast echo-planar imaging sequences, are usually preferred for fMRI acquisitions [57, 60, 63]. Figure 2.6 displays a graphical representation of the hemodynamic response function which is the BOLD response to an instantaneous stimulus. During the first 1 to 2 seconds we observe a small and short negative deflection of the BOLD signal (the initial dip). It should indicate an increase in the cerebral metabolic rate of oxygen before the cerebral blood flow response. In between 3 and 6 seconds, there is a larger peak corresponding to the cerebral blood flow increase: the main BOLD response. Finally, we observe the post-stimulus undershoot, after 12 to 30 seconds [64].

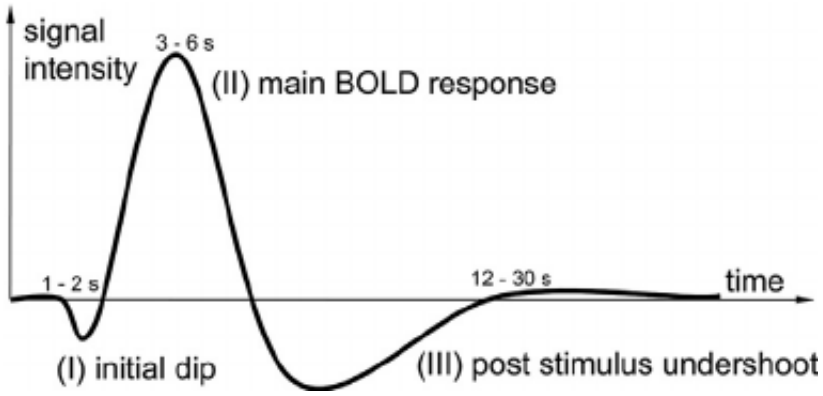

Figure 2.6: Typical representation of a BOLD hemodynamic response (from [64]).

fMRI experiments can be conducted in two distinct plans: task-based fMRI and resting-state fMRI. In task-based fMRI, the purpose is to compare two different states in order to assess the differences in neural function upon the execution of a specific cognitive task. Upon a sensory stimulus (visual, auditory, or other), the participant executes the correspondent task, while the BOLD images are being acquired [60]. In resting-state fMRI, or rs-fMRI, the objective is to measure the spontaneous and low-frequency neuronal oscillations, corresponding to resting-state networks (RSN). [65]. The participant usually remains awake, avoiding targeted mentation, and maintains the eyes open or closed during the scan, depending on the protocol [60].

Many RSN have already been identified, the most relevant being the default mode network (DMN). The brain regions corresponding to this network have increased activation at rest, while it decreases upon the execution of a cognitive task. Other RSN include somatomotor network, visual network, language network, dorsal attention network, ventral attention network, frontoparietal control network (see Figure 2.7) [66].

The study of RSN through rs-fMRI imaging has been crucial to developing a variety of clinical

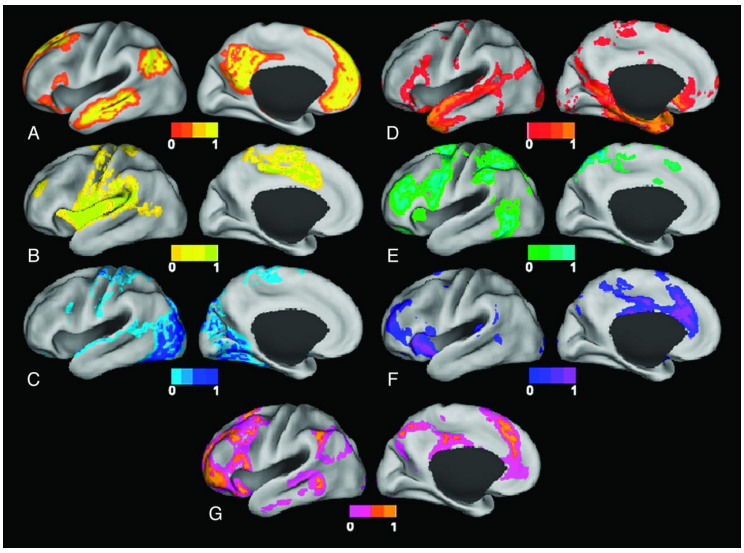

Figure 2.7: Plots of resting-state networks (RSN): A. Default mode network; B. Somatomotor network; C. Visual network; D. Language network; E. Dorsal attention network; F. ventral attention network. G. Frontoparietal control networks (from [66]).

applications. rs-fMRI can be applied in pre-surgical planning in the neurosurgical context, allowing an experimental design with minimal demands on the clinician and patient [65]. Research on rs-fMRI has also been revealing interesting findings regarding the study on RSN changes between a healthy brain and in a disease condition brain such as Alzheimer disease, depression, and other psychiatric disorders [57, 66].

To further study the mentioned RSN alterations between healthy and disease states, it is necessary to resort to methods to estimate and characterise brain functional connectivity. This matter will be discussed below, in the next subsection.

2.2.3 Assessment of Brain Connectivity

The brain is a widely complex and intriguing structure, capable of integrating and processing information. To achieve this, the brain contains several anatomical and functional connections and interactions which are optimally organised and represented in connectomes [67, 68]. The term "brain connectivity" refers to the anatomical and functional linkages of the brain that allow the execution of its functions. Brain connectivity can be characterised in three different types [67]:

- Anatomical or structural connectivity: describes the connections made by synapses between neighbouring neurons or fibre tracks (white matter) connecting neuron pools in spatially distant brain regions;
- Functional connectivity: corresponds to the temporal dependency of neuronal activation patterns of anatomically distinct brain regions, which means that there are statistical dependencies between several regions of information-processing neuronal populations;
- Effective connectivity: characterises the effect that one neuronal system displays upon another, which results in casual interaction between activated brain regions.

Each type of brain connectivity is usually represented in a connectivity matrix that encodes the relations between every considered brain region [67]. To measure brain connectivity, it is possible to resort to biomedical signals and image acquisition techniques such as electroencephalography (EEG),

magnetoencephalography (MEG) and fMRI [69, 70]. In the scope of this dissertation's purpose, we will bring our focus to fMRI, from which we can calculate functional connectivity.

As previously noted, FC is based on the idea that brain areas can be linked in the form of a network when they are correlated functionally with each other, reflecting a direct measure of functional relationships [71]. Considering the meaning behind functional connectivity, the interest related to the clinical application on neurological pathologies has been growing: the potential for the identification of clinical biomarkers, patients' stratification into biologically-based subgroups, pinpointing treatment targets, the tracking of disease progression, and the prediction of future disease development, progression, and treatment outcome had been arising in the research community [72, 73]. As such, it becomes clear that the continuous study of FC measures, analysis and interpretation, is essential to breaking new ground in the diagnosis and treatment of neurodegenerative diseases, like Parkinson's disease. For that, it is important to understand how the analysis of fMRI data is processed in order to get FC.

The first step to extract FC measures from fMRI is the image pre-processing. The steps to take on in pre-processing will be further described in Chapter 4, but essentially they consist of [1] [74]:

1. **Functional realignment and unwarp**: Scans are co-registered and resampled to a reference image, which is the first scan of the first session considered, using b-spline interpolation. This step is necessary mainly due to the head movement from the subject, thus scans need to be aligned with each other. Potential susceptibility distortion-by-motion interactions are identified in this step by estimating the derivatives of the deformation field related to head motion, which represents a sparse array of b-spline coefficients that quantifies the similarity between the warped moving image and the fixed, and resampling the functional data to match the deformation field of the reference image.

2. **Slice-timing correction**: Functional data is time-shifted and resampled using sinc-interpolation to match the time in the middle of each acquisition time, so that temporal misalignment between different slices of the functional data is corrected.

3. **Outlier identification**: Potential outlier scans are identified from the observed global BOLD signal and the amount of subject motion in the scanner. For example, criteria to consider outliers are acquisitions with framewise displacement above 0.9 mm or global BOLD signal changes above 5 standard deviations. Framewise displacement is computed at each time point by considering a bounding box around the brain and estimating the largest displacement among control points placed at the centre of the bounding box faces. Global BOLD signal change is computed at each timepoint as the change in average BOLD signal.

4. **Direct segmentation and normalisation**: Functional and anatomical data are normalised into standard MNI space and segmented into grey matter, white matter, and CSF tissues. Normalisation is important to define a one-to-one correspondence between the brains of different subjects. For the tissue classification, the estimation of the tissue probability maps from the intensity values of the reference functional/anatomical image is computed, and then the registration is based on the estimation of non-linear spatial transformation best approximating the posterior and prior tissue probability maps, until convergence.

[1] The mentioned steps of pre-processing correspond to the default pre-processing pipeline proposed by the CONN toolbox for fMRI connectivity analysis.

5. **Functional smoothing**: Functional data is smoothed using spatial convolution with a Gaussian kernel, to increase BOLD signal-to-noise ratio and reduce the influence of residual variability in functional and gyral anatomy across subjects.

After completing the pipeline of fMRI data, we can move on to the functional connectivity analysis. For that, it is crucial to perform the functional mapping of the brain in order to define which brain regions should be analysed in terms of functional connectivity [75]. To achieve this, it is needed to resort to statistical methods to analyse fMRI data. Nieto-Castaton enumerated the applied statistical methods for the analysis of fMRI data and measurement of functional connectivity in six groups [74]: seed-based connectivity methods, ROI-to-ROI connectivity methods, graph measures, network measures, dynamic connectivity and other measures.

Seed-based connectivity measures take into account a pre-defined region of interest (ROI), which will constitute the seed region and will be compared with other regions [74]. They require an *a priori* selection of the brain region, depending on the objective of the study. After that, statistical methods are applied to assess the correlation between the seed region activity and the remaining regions. From the corresponding results, it is possible to obtain a connectivity map that reflects the measures of FC. The most common method is the seed-based connectivity maps, or seed-based correlational analysis (SCA) [76], which correlates the seed region with every other location of the brain through the Fisher-transformed bivariate correlation coefficients [74]. So, considering the integral formula for the Pearson correlation coefficient [77]:

$$r(x) = \frac{\int (X-\bar{X})(Y-\bar{Y})dt}{(\int (X-\bar{Y})^2(t)dt \int (Y-\bar{Y})^2(x,t)dt)^{1/2}} \tag{2.3}$$

Taking into account that the timeseries are considered centred to zero mean, the formula for r will assume the following:

$$r(x) = \frac{\int S(x,t)R(t)dt}{(\int R^2(t)dt \int S^2(x,t)dt)^{1/2}} \tag{2.4}$$

$$Z(x) = tanh^{-1}(r(x)) \tag{2.5}$$

where S is the BOLD timeseries at each voxel, R is the average BOLD timeseries within an ROI, r is the spatial map of Pearson correlation coefficients, and Z is the seed-based connectivity map of Fisher-transformed correlation coefficients for this ROI. The Fisher transform is useful since it turns the Pearson correlation coefficient approximately normally distributed, so it stabilises the sampling distribution and facilitates the application of statistical techniques [78].

ROI-to-ROI connectivity measures refer to all the correspondences between all considered brain ROIs. Organisation and properties are similar to seed-based connectivity measures, but in this case, multiple networks of connections are taken into account [74]. Hereupon, the equations that define the Fisher-transformed bivariate correlation coefficient between a pair of ROI BOLD timeseries, to obtain the ROI-to-ROI connectivity matrices, follow [74]:

$$r(i,j) = \frac{\int R_i(x,t)R_j(t)dt}{(\int R_i^2(t)dt \int R_j^2(x,t)dt)^{1/2}} \tag{2.6}$$

$$Z(x) = tanh^{-1}(r(i,j)) \tag{2.7}$$

where R is the ROI timeseries within each ROI, r is the matrix of correlation coefficients, and Z is the ROI-to-ROI connectivity matrix of Fisher-transformed correlation coefficients.

Graph measures consider nondirectional graphs with nodes corresponding to ROIs and edges to suprathreshold connections. A graph adjacency matrix A is computed for each subject. This matrix is obtained through the thresholding of the correspondent ROI-to-ROI connectivity matrix r by an absolute or relative threshold. With the obtained graphs, it is possible to get various measures considering the topological properties of each ROI within the graph as of the entire network [74]. There are various ROI-to-ROI graph measures that can be considered.

On the other hand, the goal of network measures is to summarise voxel-to-voxel functional connections in the brain into a series of understandable measures at each individual voxel. These comprise both data-driven measures that are guided by individual-level properties, such as intrinsic connectivity, global connectivity and local correlation, and also measures that initially express group-level properties and then attempt to ascertain how those observed properties are expressed in each individual subject, like independent component analysis (ICA), principal component analysis [79] and multivariate pattern analyses [74].

Additionally, to identify and investigate the causes of temporal variability in functional connection patterns, dynamic connectivity metrics are used. In order to assess potential changes in these measures over time, some of the most popular dynamic connectivity techniques are based on sliding-window approaches, where functional connectivity measures of relevance are estimated over relatively short time windows. Alternatively, we can use the dynamic ICA, a method that aims to identify groups of connections that exhibit comparable temporal fluctuations in functional connectivity [74].

Finally, Nieto-Castanon groups two other connectivity measures: amplitude of low-frequency fluctuations (ALFF) and fractional amplitude of low-frequency fluctuations (fALFF) [74] [2]. ALFF corresponds to a measure of the BOLD signal power within the frequency band of interest. $ALFF$ is defined as the root mean square of BOLD signal (S) at each individual voxel after low or band-pass filtering h [80]:

$$ALFF(x) = \frac{1}{N}\sum_{t}(h(t) * S(x,t))^2 \tag{2.8}$$

These connectivity measures are a group of the many options that we can take an approach on. As mentioned above, the obtained connectivity values are usually represented in a connectivity matrix or map. Whatever measure it is opted to compute, it is important to have the regions of interest defined, taking into account the objective of the study.

There are various human brain atlases that one can choose from. Human brain atlases are ROI-based, meaning that they correspond to a parcellation of the cortex of the brain. Two examples of ROI-based atlases are the Automated Anatomical Labelling (AAL) and the Harvard-Oxford Cortical Structural Atlas, very popular in functional neuroimaging research. The most recent version of the AAL, AAL3, includes 170 brain atlas regions. This version includes subdivisions of the anterior cingulate cortex, the thalamus the nucleus accumbens, substantia nigra, ventral tegmental area, red nucleus, locus coeruleus and raphe nuclei [81]. The Harvard-Oxford Atlas, on the other hand, includes 48 cortical and 21 subcortical structural areas, derived from structural data of T1-weighted images.

Brain atlases can also be network-based. As previously mentioned, some of the main identified resting-state networks include the default mode network, somatomotor network, visual network,

[2]The mentioned ALFF metrics technically are measures of amplitude of the signal. However, the author considers them as connectivity measures.

language network, dorsal and ventral attention networks and frontoparietal networks (refer to Figure 2.7). Most of these networks are included in the atlas defined by Yeo et al., in a study that aimed to study the organisation of the functional networks in the human brain [82]. From this work, seven networks (visual, somatomotor, dorsal and ventral attention, limbic, frontoparietal and default mode networks) were identified. However, in the context of this dissertation, we intend to study connectivity alteration in Parkinson's disease, it would be interesting to choose an atlas that includes a resting network corresponding to the activity related to the basal ganglia. Considering this, the set of identified networks by the group Shirer et al. seems to be a proper option in light of the aim of this work [83].

Shirer et al. aimed to build a classifier to identify specific patterns of the whole brain connectivity, by defining resting networks. For this, the group acquired fMRI scans from 27 healthy subjects, from which 15 were used to define the functional networks, and the remaining 12 were used for later validation of the produced classifier [83]. The first group completed four different tasks: a rest, an episodic memory task, a music lyrics task and a subtraction task. The latter group completed an additional task, beyond the previous, which consisted of imagining walking through the rooms of their home. The classifier would consider these four tasks and then identify the involved functional networks.

The results of the group comprised 14 functional networks, which included 91 ROIs [83]:

- Auditory network: enables interaction with the world, including the processing of auditory information and other sensory and cognitive processes [84];
- Basal ganglia network: involved in several functional processes, as motor, cognitive and limbic [85];
- Dorsal default mode network: located in the posterior cingulate cortex (PCC) and medial prefrontal cortex (MPFC), is thought to be involved in the evaluation of imagined future events [86];
- Higher visual network: located in the secondary visual cortex (V2), which receives visual stimuli from the primary visual cortex [87];
- Language network: allows language processing, comprehension and production [88];
- Left executive control network: located in the left dorsolateral prefrontal cortex (DLPFC) and left parietal lobe, is responsible for executive function (which includes working memory, flexible thinking, and self-control) [89];
- Sensorimotor network: is thought to be activated upon preparation of the brain in performing and co-ordinating a motor task [90];
- Posterior salience network: located in the posterior insula, engaged in the interoception of the reward-related emotions [91]
- Precuneous network: located in the precuneus cortex, its functions include recollection and memory, integration of information relating to the perception of the environment, cue reactivity, mental imagery strategies, episodic memory retrieval, and affective responses to pain [92];
- Medial/primary visual network, located in the primary visual cortex (V1), which receives and processes visual information;
- Right executive control network, located in the right dorsolateral prefrontal cortex (DLPFC), is responsible for executive function [89];

- Anterior salience network: located in the insula and dorsal anterior cingulate cortex (dACC), engaged in the interoception of the reward-related emotions [91];
- Ventral default mode network: located in the retrospenial cortex (RSC) and medial temporal lobe (MTL), is thought to be involved in the construction of imagined future events [86];
- Visuospatial network: located in the intraparietal sulcus (IPS) and frontal eye field (FEF), involved essentially in visual perception [93].

Figure A.1 (Appendix A) represents the brain region that each network comprises. The networks identified by Shirer et al. are described according to their anatomic location in Appendix A, Tables A.1, A.2, A.3. We raise attention to the basal ganglia network, which may be particularly relevant in the scope of this dissertation.

Throughout this section, the basic concepts regarding functional magnetic resonance imaging were explained, especially how the scans are obtained and what information we can obtain from them. In particular, resting-state fMRI has been raising attention in the scope of alterations in resting-state networks related to neurologic and psychiatric diseases [57, 66]. To analyse those fluctuations, we resort to the computing of brain functional connectivity, which refers to the functional connections between different ROIs of the brain [67].

In this project, we want to compare the functional connectivity between healthy subjects and patients with PD and provide a diagnostic tool through this analysis. Since we aim to perform such a classification task, artificial intelligence emerges as a powerful tool to solve this problem.

2.3 Artificial Intelligence

Artificial intelligence (AI) has been growing and breaking new ground in several areas, and medicine and healthcare are no exceptions. AI is a multidisciplinary set of technologies that aims to reproduce human-like actions such as decision-making, learning and judgment, by using computers through simulation and training [94]. Some of the technologies that comprise AI and have been far-reaching in the development of medicine are natural language processing, rule-based expert systems, robotic process automation, physical robots, and the most popular and common one, machine learning (ML) [95].

ML is defined by Peter Flach as "the systematic study of algorithms and systems that improve their knowledge or performance with experience" [96]. It creates a model from large and complex data sets without explicitly programming the model. ML problems can be based on labelled data (supervised learning), unlabeled data (unsupervised learning), or they can include a mixture of both (semi-supervised learning) [94]. One of the main "ingredients" of a ML task is features, but it is possible to resort to complex neural networks to compute hidden features and use them to solve the given problem [95]. On this scope, a sub-field of ML, deep learning (DL) emerges.

Throughout this section, the DL concept will be explored, so its applications and limitations are presented. Furthermore, explainable artificial intelligence (XAI), an AI branch that shows promise in tackling the black-box problem of DL, will be introduced.

2.3.1 Deep Learning

Deep learning is considered one of the major developments in the history of computer science. DL is already showing promise in several tasks in numerous fields, such as cancer diagnosis, precision

medicine, self-driving cars, predictive forecasting, and speech recognition [97]. Deep learning technologies are based on neural networks (NN), which comprise a set of neural nodes. In a way, DL works on multiple functions stacked together, based on linear regressions and activation functions [98]. In DL, algorithms are constantly being learned as the amount of data increases, so that the training improves [99]. We could say that the DL mimics the human brain's perception and learning of multiple information while capturing large amounts of data.

The first acknowledged model that led to the development of artificial neural networks resulted from the work by McCulloch and Pitts, in 1943 [100]: the McCulloch-Pitts neuron. They aimed to simulate the human brain by understanding how the brain used interconnected neurons to create complex patterns involved in the perception processes [101]. This inspired the creation of the first prototype of a neural network, the perceptron, a work published by Frank Rosenblatt in 1958 [102]. This network, based on the firing mechanism of a biological neuron, is composed of two layers of processing units that could recognise simple patterns [97]. To study and apply far more complex and recent neural networks, with multiple hidden layers, it is crucial to understand the simple architecture of the perceptron.

Figure 2.8 represents a scheme of the performed calculations in the perceptron. This network takes as inputs x_k, $k = 1,2,3,...,n$, which are corresponded to the respective weights, w_k. The first operation of the perceptron is a weighted sum (the linear function): each input is multiplied by its weights and those values are summed with each other:

$$z = (\sum_{k=1}^{n} w_k x_k) + b \tag{2.9}$$

where b stands for the bias term. Afterwards, it is applied a non-linear function, the activation function $a(z)$. From this operation, we then obtain the output of the perceptron:

$$y = a(z) = a((\sum_{k=1}^{n} w_k x_k) + b) \tag{2.10}$$

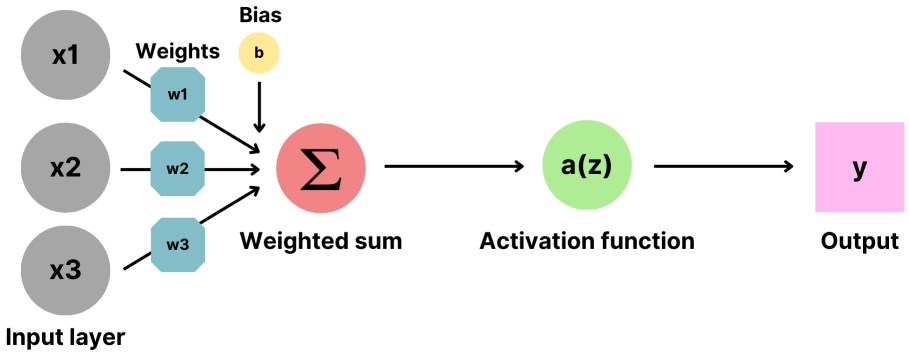

Figure 2.8: Example of a Perceptron architecture.

As for the activation functions, that allow us to interpret non-linear transformations in data, some of the most commonly used are the sigmoid function, the hyperbolic tangent and the rectified linear unit (ReLU) [97]. The sigmoid function narrows the output to a range of 0 to 1, so, it is often opted in binary classification cases. The same goes for the tanh function, which ranges from -1 to 1, meaning that its outputs are zero-centred. As observed in Figure 2.9, both sigmoid and tanh saturate on the binary outputs, hence their gradients will be close to zero for very large or very small values of z. Differently from these functions, ReLU, while saturating for small values, it does not saturate for larger values of z. The vanishing problem previously identified is not verified in ReLU, so it is expected for the algorithm to

learn faster [97]. However, we still have a zero gradient for negative values of z. To surpass this issue, a modified version of ReLU was created: the Leaky ReLU. This function has a small slope m for negative values, defined prior to the training process, so the gradient for negative values is different from zero:

$$LeakyReLU = \begin{cases} mx, \text{ if } x \leq 0 \\ x, \text{ if } x > 0 \end{cases} \tag{2.11}$$

Additionally, we may opt for the Parametrized ReLU (PReLU), which is similar to Leaky ReLU, but the parameter m is learnable throughout the training process.

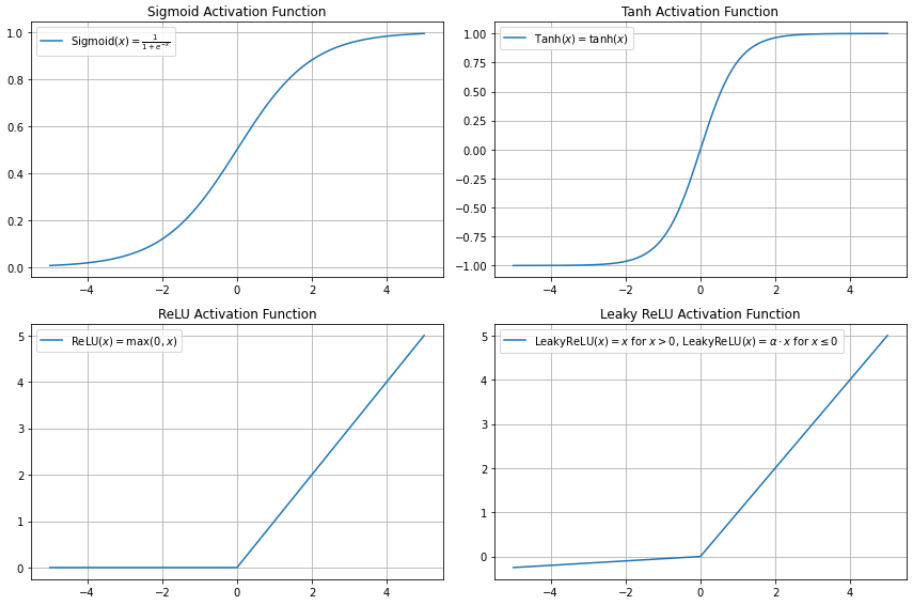

Figure 2.9: Plots of the activation functions sigmoid, hyperbolic tangent (tanh), rectified linear unit (ReLU), and Leaky ReLU.

But how did they reach deep neural networks from a simple model like the perceptron? The perceptron showed the capability of learning any continuous problem with a single hidden unit, although, something else was missing [97]. How could we automatically extract features, and solve even more complex problems? And here is where backpropagation learning algorithms arise. These algorithms aimed to indicate how a machine should adjust the model's parameters (weights), which are used to compute the representation in each layer from the representation in the previous layer [9]. The backpropagation algorithm is applied to the multilayer perceptron, an artificial neural network that includes an input and an output layer, and one or more hidden layers (refer to Figure 2.10). While a perceptron neuron must have an activation function that restricts to a threshold, neurons in a multilayer perceptron can use any activation function. The multilayer perceptron is the basis of the modelling of deep neural networks.

Hereupon, to create deep neural networks, we "simply" add more hidden layers to our model. Between the input and output, a DL architecture is composed of multiple layers (the hidden layers), with the neural nodes being the hidden nodes. Each module in the stack alters its input in order to improve the representation's selectivity and invariance. With multiple non-linear layers, a model can implement functions of its inputs that are sensitive to precise and specific details, and yet insensitive to large and irrelevant variations [9].

Now that we understand the basis of neural networks and what inspired deep learning architecture, we are proceeding to the training process of a deep neural network.

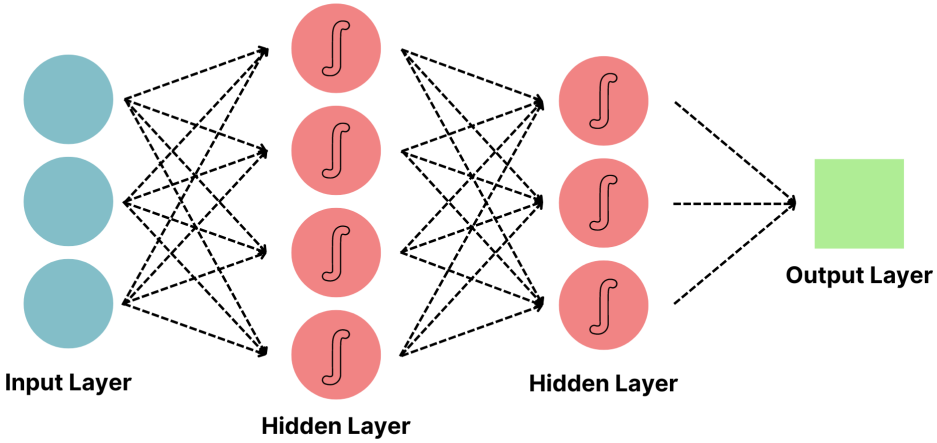

Figure 2.10: Example of an architecture of a multilayer perceptron with two hidden layers. Hidden layers reflect the complexity and performance of a deep neural network, and define what data transformations the correspondent functions will apply.

2.3.1.1 Training and Evaluating a Neural Network

The process of training a NN involves forward and backward calculations throughout each layer, while minimising a loss function. Hence, the three main steps of training are forward propagation, calculation of loss function, and backward propagation.

As the name suggests, during forward propagation, the inputs in the initial layers will move forward through the hidden layers, where on each one a different feature or aspect is identified. Then, the output layer is activated and the corresponding results are retrieved. Afterwards, the loss (or cost) function is used to calculate the difference between the predicted value from the actual value, and the performance of the model is evaluated [9]. Ideally, the loss function is equal to zero, hence, we aim to minimise it. Some of the most commonly applied loss functions are mean squared error (often applied to regression tasks) [103], binary cross-entropy (applied to binary classification tasks), and categorical cross-entropy (multi-class classification tasks) [104].

Lastly, the training process proceeds to the backward propagation step, where computation is made from the output layer to the input layer. With the value of the cost function, the objective is to adjust the learnable weights in order to optimise the algorithm. Keeping in mind that we want to minimise the cost function, it is natural that the optimisation will be based on the gradient [9, 97]. In fact, the most common optimisation algorithm is gradient descent, which looks for the minimum value of a function by searching for the (close to) zero derivative.

Another aspect to take into account during the algorithm optimisation is the hyperparameters of the model. For example, the number of epochs, which corresponds to the number of training iterations over the entire training data set, must be large enough to allow the loss function to minimise, while enabling an optimal training time. Then we also have the learning rate, which refers to the size of each step to take to the minimum derivative. The learning rate is an important hyperparameter to optimise since it needs to be enough to reach the function minimum, avoiding gradient explosion, although, again, the training time must not be too long. Thus, optimisation methods to adapt the learning rate were developed. One of the most popular is the Adam (adaptive moment estimation) optimiser. This algorithm is a modern juxtaposition of two other optimisers, the AdaGrad and RMSprop, combining their respective advantages: working well with sparse gradients and working well in online and non-stationary settings.

Additionally, Adam optimiser implies that magnitudes of parameter updates are invariant to the rescaling of the gradient. Furthermore, its stepsizes are constrained by the stepsize hyperparameter, it does not require a stationary goal, works with sparse gradients, and employs a type of step size annealing [105].

The initialisation of the weights and bias parameters was also refined to appropriate algorithms. For example, He initialisation optimises the initialisation of neural networks with layers that use ReLU as the activation function. This technique enables the neural network to start with a consistent level of variation between inputs and outputs during both forward and backward passes. This has been observed to provide substantial enhancements in the stability and speed of training [106].

During the training process, it is important to ensure that the model is generalising, by verifying the test performance on a test data set. Hence, to obtain an appropriately fitting model, it needs to perform well on both training and test sets, so the respective prediction errors must be minimised. Of course, this is the ideal scenario, which does not often happen. The model may be too simple, and it does not obtain a sufficiently low error value on the training set. The model is then unable to capture the underlying pattern of the data, which means that makes more assumptions about the data (high bias). This means that the model is underfitting [107]. On the other hand, the model could almost perfectly fit the training set, although on the test set the performance is not appropriate, and consequently, the training error and the test error are too distinct (high variance). In this case, the model is overfitting [107]. Figure 2.11 graphically represents the phenomena of underfitting and overfitting.

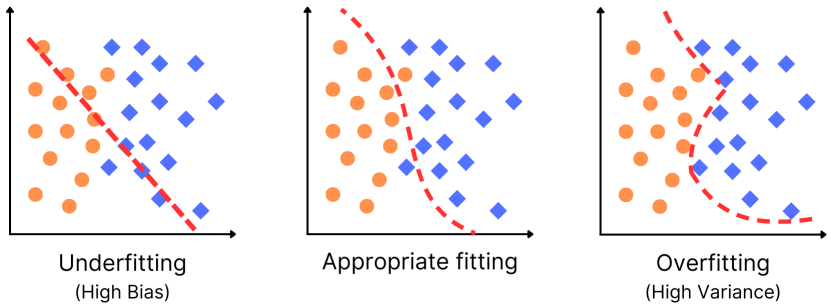

Figure 2.11: Graphic representations of underfitting, appropriate fitting and overfitting of a model to training data.

To solve underfitting, it might be needed a more complex model that captures the discriminative features of the training data. In the light of DL models, the solution might be to create a deeper model, with more hidden layers. To tackle overfitting, getting more data, if possible, will allow the training data to become more robust, hence the model may become more general. Nonetheless, there are additional techniques that aim to avoid overfitting: regularisation. It involves any alteration that allows a learning algorithm to reduce its test error (or generalisation error), but not the training error [107]. Most regularisation approaches are based on adding a penalty term Ω to the loss function J:

$$\bar{J} = J + \alpha\Omega \tag{2.12}$$

where α is a hyperparameter that weights the relative contribution of the penalty term relative to the loss or cost function. Two regularisation methods that are the most applied are the L1 and L2 regularisation. L1 regularisation uses as the penalty term Ω the norm of the model weights, $||\omega||$. This method leads to sparsity since it drives some weights to zero, and this property can be used as a feature selection mechanism. The penalty term of L2 regularisation is the squared magnitude of the weights, $\frac{1}{2}||w||^2$. L2 regularisation allows the detection of high variance in the training data, which causes it to reduce the weights on features whose correlation with the output target is low in comparison to the additional

variance [107].

Another regularisation technique to avoid overfitting is the dropout, where the neuron units are randomly chosen and their weights and outputs are nullified so that they do not influence the forward or the backward propagation, and the model can be less specific on the training set [97].

As we have seen throughout the process of training, a considerable amount of hyperparameters of the model arise, and it is not an easy task to tune them. The hyperparameters must correspond to a model that fits satisfactorily into the data and that is generalisable to other sets of data, ideally. There is a method that is often used while fine-tuning the model hyperparameters: k-fold cross-validation, or simply cross-validation [107]. The available data set is partitioned into k distinct subsets of equal size. This division is done by randomly picking examples from the learning set and not replacing them. The model is trained using $k-1$ subsets that compose the training set. The model is then applied to the remaining subset (the validation set) and its performance is evaluated. This process is continued until all k subsets have served as validation sets. The cross-validated performance is the average of the k performance measurements across the k validation sets.

As previously mentioned, when a model is trained, it is needed to evaluate its performance. For that, we resort to specific metrics that reflect how the model is fitting into the data. One of the most used metrics is accuracy, which in the case of a binary classification problem, is given by:

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{2.13}$$

TP being is the number of true positives (positive class instances correctly classified), TN the number of true negatives (negative class instances correctly classified), FP the number of false positives (negative class instances incorrectly classified as positive) and FN the number of false negatives (positive class instances incorrectly classified as negative). Accuracy reflects the percentage of correctly classified instances. This metric is a good choice when the number of samples in each class is similar [108]. Although, in classification problems, especially in the medical field, the data sets are often imbalanced. This is why when evaluating the performance of a model, it is important to not only look at accuracy but also at other important metrics.

To give a visual representation of the performance of the model, we have the confusion matrix. The confusion matrix displays the distribution of true positives, true negatives, false positives and false negatives, by comparing the predicted classes *versus* the actual classes in a table. The table below displays an example of a confusion matrix for a binary classification problem.

Table 2.1: Confusion matrix for a binary classification task.

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

There are several metrics that aim to describe how the model is classifying each one of the classes considered in the problem. Sensitivity, or recall, or true positive rate (Equation 2.14), corresponds to the rate of positive class instances that were correctly identified, comparing with the number of actual positive instances:

$$\text{sensitivity} = \frac{TP}{TP+FN} \tag{2.14}$$

On the other hand, specificity, or true negative rate (Equation 2.15), is the number of negative class

instances that were correctly classified divided by the total number of actual negative instances:

$$\text{specificity} = \frac{TN}{TN+FP} \tag{2.15}$$

Precision (Equation 2.16) is the fraction between true positives and the sum of true positives with false positives. It represents the rate of instances classified as positive that are actually positive instances:

$$\text{precision} = \frac{TP}{TP+FP} \tag{2.16}$$

F1-score (Equation 2.17), a harmonic mean between precision and recall, which ranges from 0 to 1. The closer to 1 is the F1-score, the more precise and robust the model.

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.17}$$

A ROC (receiver operating characteristic) curve is a graph that plots the true positive rate and false positive rate ($\frac{FP}{FP+TN}$) to reflect the performance of a classification model at all classification thresholds [109]. Lowering the classification threshold causes more items to be classified as positive, which increases both false positives and true positives. There is a classification metric that is relevant for the evaluation of the performance of the model and is related to the ROC curve: the AUC ROC (area under the ROC curve). AUC ROC may be interpreted as the likelihood that the model rates a random positive case higher than a random negative example [109]. The closer is AUC to one, meaning the ROC curve is closer to a true positive rate equal to 1 and a false positive rate equal to 0, the better the performance of the model. Figure 2.12 shows the graphical representation of the ROC curve and respective AUC.

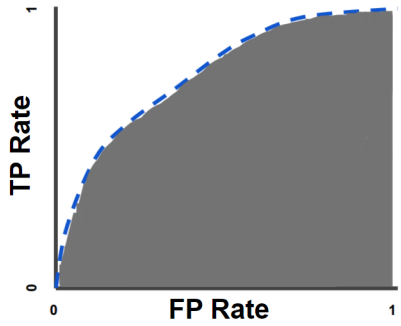

Figure 2.12: Graphic representation of AUC (from [109]).

We have looked through some good practices while training a model, how to tune it and how to evaluate its performance. A good DL model must be accurate (have a low error rate and be able to generalize well to unseen data), efficient (achieve high performance with minimal computational resources and time), and robust (capable of performing well across different scenarios, data sets, and conditions). To achieve this, we must take into account all the best practices through all the steps of training and evaluating a model.

2.3.1.2 Convolutional Neural Networks

Depending on the type of problem that will use a DL model to be solved, several architectures of DNN can be chosen. The most popular choice in computer vision tasks, which comprise most of the problems involving medical imaging analysis, is the convolutional neural network (CNN). CNN are

neural networks that apply the convolution operation instead of a general matrix multiplication in at least one of their layers [107].

The (discrete) convolution operation, in the context of machine learning, is given by:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n) \tag{2.18}$$

with I being the input data in the convolution operator (for example, a two-dimensional image) and K the kernel function. The result S is often referred to as the features map [107].

A CNN is usually composed of three types of layers: the convolutional layer, the pooling layer, and the fully connected layer [110].

The role of convolutional layers is to extract features from the previous layer [9]. Each kernel, or filter, slides across the input data, performing element-wise multiplications and summing the results to produce feature maps, as we have seen in Equation 2.18. Convolutional layers play an important role in feature extracting by recurring to three concepts that contribute to an improvement of a machine learning mode: sparse interactions, parameter sharing, and equivariant representations [107]. Sparse interactions refer to the capacity of small kernels, for instance, with the size of hundreds of pixels, identifying local and meaningful information through feature extraction in images with thousands or even millions of pixels. Hence, fewer parameters need to be stored, reducing the memory required for the model and improving its statistical efficiency. Parameter sharing corresponds to using the same parameter for more than one function in a model, which means that for getting output, weights applied to one input are the same as the weight applied to another input [107]. Finally, equivariant representations refer to equivariance inherent to the convolution operation and respective parameter sharing: if the input changes, the output changes in the same way.

The pooling layer is often referred to as subsampling or downsampling layer. This layer aims to reduce the spatial dimensions of the input feature maps while retaining important information. Common pooling operations include max pooling and average pooling, where the maximum or average value within a pooling region is taken to form the output [97]. Activation layers are applied to the output of the convolutional or pooling layer. A dropout layer and/or a batch normalisation layer may also be added, for regularisation purposes. Figure 2.13 shows a pipeline of a simple CNN, representing its fundamental layers.

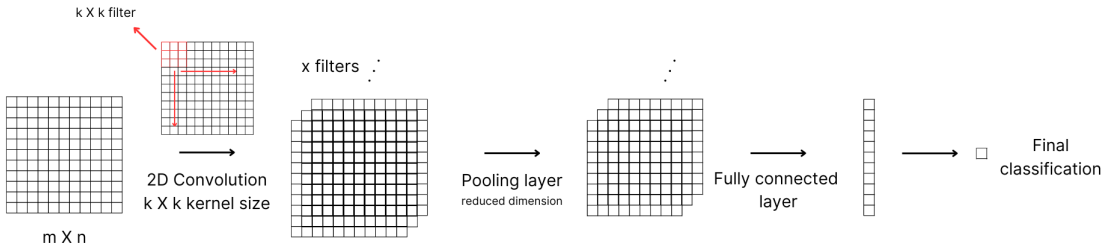

Figure 2.13: Simple architecture of a convolutional neural network, representing the convolutional layer, the pooling layer, and the fully connected layer.

The fully connected layer is responsible for the final classifications. A flatten layer arranges the previous feature map into a one-dimensional vector, which is fed to the fully connected layer. At the end of a CNN, we have the final prediction vector accordingly to the number of classes considered in the problem.

As mentioned above, CNN are mostly used with two-dimensional inputs, such as images, hence,

this DNN architecture is a powerful choice for computer vision tasks. In the scope of this dissertation, we will be working with connectivity matrices, which constitute 2D data that can be fed to a CNN. For instance, there are several groups that attempted to use CNN to classify different diseases with connectivity matrices, such as attention deficit hyperactivity disorder (ADHD) [111] and autism spectrum disease (ASD) [112].

2.3.2 Challenges in Deep Learning and Neuroimaging

It is true that deep learning algorithms reveal themselves as promising and powerful tools to enhance various tasks in several fields, in particular in medicine and neuroimaging. DL has been explored as a potential method to predict clinical outcomes through neuroimaging data such as fMRI. However, DL is still a relatively recent resource in terms of applications in neuroimaging. Therefore, there are challenges to take into account and to be addressed. The most evident hurdle that arises in DL is the need for huge amounts of data, which is often not feasible in the medical context, due to the time and monetary expense required to perform neuroimaging studies [113]. Additionally, feature selection is not an easy task upon high dimensional data such as fMRI.

To overcome the issues of small sample sizes and high dimensional data, two possible solutions arise: transfer learning and data augmentation. Transfer learning refers to the transfer of knowledge from a source domain model, to the target domain model. The source domain model usually solves one problem that is distinct from but related to the target problem and uses larger amounts of training data [114]. There have been several studies that evidence that transfer learning is a useful technique when the availability of medical images is limited [115, 116, 117]. On the other hand, data augmentation techniques aim to obtain more samples by applying transformation techniques. In the case of neuroimaging, the mixup technique arises as a data augmentation technique. Mixup is a type of self-supervised learning in which the learner creates virtually labelled instances as a result of combining separate data pieces that help smooth decision boundaries and, consequently, help prevent overfitting [118].

There is one more challenge that developers face while building DL models: the black-box nature. This means that DL algorithms provide predictions, but the features and/or feature combinations that were used to make the decisions that lead to the output are not known [113]. The black-box issue in DL models is particularly a concern in applications to neuroimaging since it is not possible to assess how the model identified a specific pathology, therefore, this approach is unable to determine the targetable biomarkers. If DL models apparently lack explainability and transparency, it will become challenging for healthcare professionals to rely on these automated solutions for detecting pathologies. Furthermore, explaining the obtained decisions of neural networks is not an easy task, since it requires an internal understanding of the model that the end-users (namely healthcare professionals) do not present [13]. Considering this, it becomes clear that a solution to build confidence and interpretability into the DL models, especially for neuroimaging, is needed so that the implementation of these algorithms in clinical practice becomes possible in the near future. Upon this, explainable artificial intelligence (XAI), emerges as a tool to produce interpretable, intuitive, and human-understandable explanations of AI decisions [13].

2.3.3 Explainable Artificial Intelligence

There is no doubt regarding the rapid growth of artificial intelligence during the last decade. In fact, AI models have been becoming more and more complex and larger in order to improve their predictive power. However this comes with a drawback: we have left black-box models that lack transparency and explainability, and because of this, users are not able to interpret the decisions made by the models.

Explainable artificial intelligence (XAI) emerges as a means to turn AI models and the respective results understandable for their human user [119]. Explainable artificial intelligence corresponds to a group of methods that provide black-box models explainability and interpretability [120]. In [13] three main reasons why XAI should be an important topic on research are pointed out: (1) it improves transparency, since XAI provides human-readable explanations of the model decisions, (2) it improves trust, built by the scientific explanations created by XAI, (3) it helps mitigate introduced biases and create fairer models.

XAI methods can be described in terms of scope (where the method is focusing on), methodology (what the algorithmic approach is) and usage (how the method is developed) [13]. According to scope, XAI models can be local or global. Local XAI methods present the individual contribution of each feature of a single instance x from the data set X. Global XAI methods, on the other hand, deduce the insight of the whole model, providing an explanation about the model. In terms of methodology, models are categorised as backpropagation-based and perturbation-based. In backpropagation-based methods, during the backpropagation step, the XAI algorithm uses partial derivatives of the activations to generate attributions after one or more forward passes through the neural network. Differently, in perturbation-based methods, the algorithm perturbs the feature set of an input instance by using methods such as occlusion or masking. As regarding usage, XAI models can be described as intrinsic (the algorithm is embedded in the neural network, hence it depends on the model architecture and it cannot be transferred to different models) or post-hoc (the algorithm is applied to *a priori* trained neural networks, so it does not depend on the model architecture). Additionally, another distinction between XAI strategies is whether they are model agnostic, which means they may be applied to any form of model architecture, or model specific, that are only applicable to a single type of architecture [121].

Each XAI method or algorithm has different features in terms of scope, methodology and usage. Taking into account those characteristics, the XAI method should be carefully chosen according to the explanations and interpretations intended to be captured. In the case of this work, it is preferable to opt for a post-hoc method, so it can be applied to a previously trained DNN. Furthermore, since we aim to specifically identify the brain regions that are significant in the classification of Parkinson's disease, a local explanation method should be the most suitable approach, to identify the correspondent features. Considering these characteristics, three XAI models will be introduced: layer-wise relevance propagation, guided backpropagation, and deconvolution network.

2.3.3.1 Layer-wise Relevance Propagation

Layer-wise relevance propagation (LRP) is a XAI method based on pixel-wise decomposition: it is based on the understanding of the contribution of a single pixel of an image to its prediction. LRP aims to infer quantitatively which pixels contribute to a positive or negative classification result (in case of a binary classification problem) [122]. This is an algorithm that is applied to images, i.e., it is often applied to computer vision problems, where to each pixel of the image a relevance score is attributed. LRP is a backpropagation-based and model-agnostic method. For instance, when applied to a convolutional neural network, LRP backpropagates relevance information from the output class to the input layer [13].

The relevance propagation process proceeds as described in Figure 2.14. The obtained output result from the network is backpropagated throughout each layer, using specific propagation rules. This structure is characterised by a relevance conservation property similar to Kirchhoff's conservation laws in electrical circuits [123]. Considering j and k indices for the neurons of two successive layers, R_j and R_k their respective relevance score for a prediction $f(x)$, and $R_{j\leftarrow k}$ the share of R_k that is redistributed to

neuron j. The conservation rule for the neuron j in the lower layer is defined as [124]:

$$\sum_j R_{j \leftarrow k} = R_k \tag{2.19}$$

Considering the rule for the neuron k in the higher neuron:

$$\sum_k R_{j \leftarrow k} = R_j \tag{2.20}$$

and combining both Equations 2.19 and 2.20, we obtain the global conservation of relevance through the following chain of equalities [124]:

$$\sum_{i=1}^{d} R_i = \ldots = \sum_j R_j = \sum_k R_k = \ldots = f(x) \tag{2.21}$$

Regarding the existent propagation rules, the most elemental one is the basic rule (LRP-0), which redistributes in proportion to the contributions of each input to the neuron activation as [123]:

$$R_j = \sum_k \frac{a_j w_{jk}}{\sum_{0,j} a_j w_{jk}} R_k \tag{2.22}$$

with a_j being to the neuron activation at the corresponding layer, and w_{jk} the weight of the connection between neuron j and neuron k in the higher layer. This rule seems straightforward, however, it can be demonstrated that applying it uniformly throughout the entire neural network yields an explanation comparable to the gradient of a deep neural network, which is often noisy [123]. Considering this, the development of more robust propagation rules is seen as needed. One example is the epsilon rule (LRP-ε), which adds a small positive term in the denominator of the LRP-0, to absorb some relevance when the contributions to the activation of neuron k are weak. This way, explanations become sparser in terms of input features and less noisy [123]. The equation of the LRP-ε follows:

$$R_j = \sum_k \frac{a_j w_{jk}}{\varepsilon + \sum_{0,j} a_j w_{jk}} R_k \tag{2.23}$$

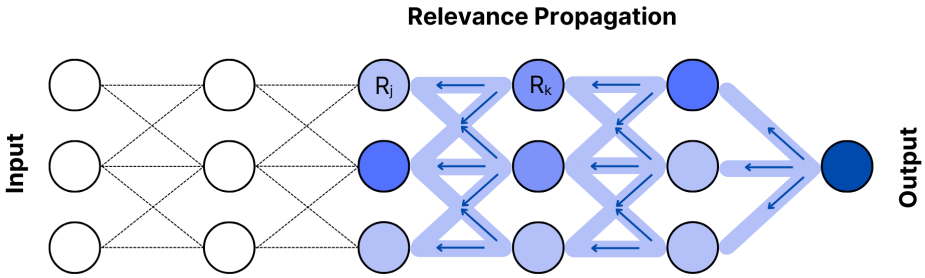

Figure 2.14: Graphic representation of the layer-wise relevance propagation (based on [123]).

2.3.3.2 Deconvolutional Network and Guided Backpropagation

The deconvolutional network, or DeconvNet, is a convolutional neural network designed with filters and unpooling operations to reconstruct the original image part of the input image that is most strongly activating the corresponding neurons [13, 125]. A DeconvNet is coupled to each one of a CNN layers to provide a continuous path back to picture pixels. First, an input image is sent to the CNN, and the

respective features are computed. To investigate a specific CNN activation, other activations in the layer are set to zero and then the feature maps are fed into the connected DeconvNet layer. To recreate the activity in the layer corresponding to the selected activation, there are three steps to take: unpooling (uses switch variables to place reconstructions of the max pooling from the layer above), rectifying (passing the reconstructed signal through a ReLU non-linearity) and filtering (uses transposed versions of the convolutional filters to the rectified maps) [126].

Springenberg et al. evaluated the DeconvNet method and pointed out that the applied switch variables in the unpooling layer, which correspond to the positions of maximum values of each pooling layer, will not directly reflect the previously learned features [125]. Furthermore, the authors highlight that backpropagation, an alternative way to assess the most significant features of an image which is based on the backward computing of the gradient of the activation referent to the image, is partially conditioned on an image through the activation functions of the network. Considering the above, the group proposes an XAI method, guided backpropagation, that combines both DeconvNet and backpropagation methods, by adding an additional guidance signal from the higher layers (a characteristic from the DeconvNet method) to the usual backpropagation [125]. Guided backpropagation shifts the negative gradients to zero so that the positive gradients are highlighted, allowing the assessment of the features that activate the neurons [127].

Figure 2.15 visually describes how each XAI method, DeconvNet, backpropagation, and guided backpropagation, performs on the backward pass. In the forward pass, i.e., the application of the ReLU activation, the negative values shift to zero. For the backward pass in backpropagation, it passes the gradient to a previous layer only if the original input was positive. As for DeconvNet, it simply sets to zero the negative entries. In guided backpropagation, we have a combination of the previous two methods.

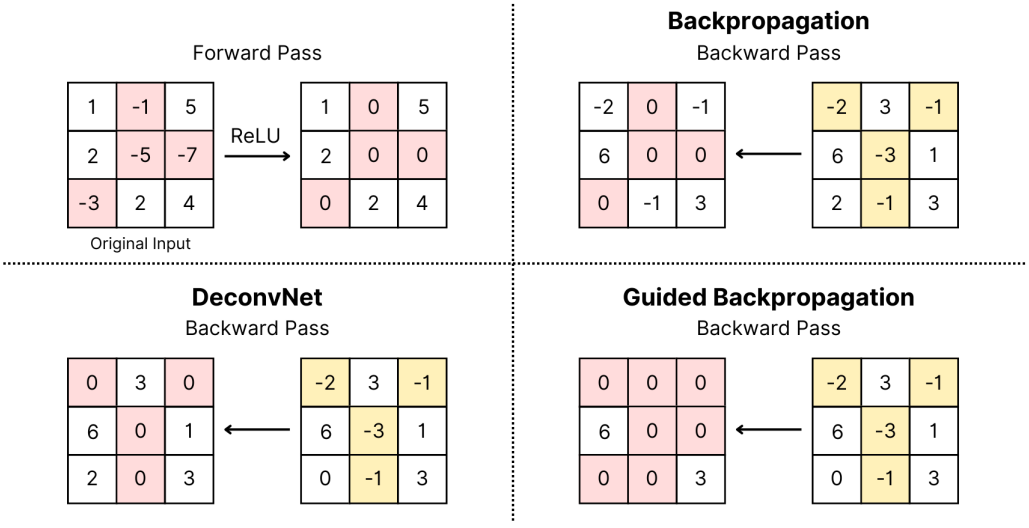

Figure 2.15: Graphic representation of the the forward pass on ReLU activation, and backward pass in the XAI methods backpropagation, DeconvNet, and guided backpropagation (based on [125]).

Chapter 3

State-of-the-art

As we have seen throughout Chapter 2, there are three pillars that comprise the basis of this dissertation: Parkinson's disease, brain functional connectivity and artificial intelligence. This chapter has three parts that cover the mentioned three concepts. First, we will see how artificial intelligence has been valuable in the diagnosis of PD. Afterwards, the state-of-the-art of study on brain functional connectivity assessment in PD will be presented. Finally, we will investigate how the application of XAI on neuroimaging has aided in the interpretation of medical images of the human brain.

3.1 Use of Artificial Intelligence for Parkinson's Disease Diagnosis

The use of AI for the diagnosis of pathologies has been growing over the years, and PD is no exception. As aforementioned in Chapter 2, PD diagnosis is often based on medical history and clinical presentation. Nonetheless, the use of physiological signals acquisition methods such as speech and voice detection, electroencephalography (EEG), electromyography (EMG), and biomedical imaging techniques such as SPECT, MRI and fMRI, have been raising attention in the development of PD diagnosis and in the search of biomarkers of the pathology [128, 129]. Resorting to AI, especially ML techniques, allows the automatisation of the analysis of these data, in order to optimise the diagnosis of PD and other pathologies. Considering the scope of these dissertation objectives, the state-of-the-art on utilisation of AI in the diagnosis of PD will focus on the use of biomedical images. Furthermore, it will focus on the binary classification task (PD vs healthy controls).

As previously mentioned, the most commonly used imaging technique for PD detection is DaTScan, a SPECT method. Kim et al. (2018) propose the use of transfer learning to classify DaTScan images (obtained from the Royal Devon and Exeter Hospital UK) in order to detect PD [130]. The training set was augmented using geometrical alterations to produce a 44-fold amplification, resulting in 2376 normal and 2376 abnormal images. This set was split into train, validation and test sets with a ratio of 80:10:10. The augmented training data was used to retrain the top layer of the Inception v3 CNN [131] (this is the basis of transfer learning). The group obtained a sensitivity of 96% and a specificity of 66.7%, which reflects promise on the use of convolutional neural networks for medical image classification in the context of PD. Furthermore, this study also infers that transfer learning of CNNs trained with millions of non-medical images may be a useful tool in order to obtain a robust model for biomedical image classification, since one of the known limitations is the lack of acquired medical data, compared with other data [130]. Martinez-Murcia et. al also resort to SPECT scans to build ML models for PD classification, first using a 3D-CNN in 2017 [132], and later in 2018 a deep convolutional autoencoder (DCAE) [133] (both studies are mentioned in the review on deep learning approaches on psychiatric

diseases by Zhang et al. [8]). Sivaranjini and Sujatha (also referred to in [8]) opted for a transfer learned Alex-Net model (with a 2D-CNN architecture) [134].

Moreover, there are several studies that propose the use of ML methods on both structural MRI (sMRI) and fMRI scans, and additionally, diffusion tensor imaging (DTI) [8, 128, 135, 136, 137, 138]. Most of the approaches on sMRI and fMRI, according to the 2019 review on application of ML on PD diagnosis and MRI by Xu et al. [135], opt to use a support vector machine (SVM) classifier [139, 140, 141, 142, 143, 144]. There is also a report on the use of linear discriminant analysis [145]. The ML approaches described and summarized by Xu et al. [135] reached an accuracy of approximately 70 to 100 %. Xu et al. point out that most studies are based on small data sets (which is a recurrent problem in many other studies in the field of PD). They also recommend adopting multimodal approaches, the inclusion of patients in prodromal states, and deep learning applications [135].

Rubbert et al. (2019) also propose a machine learning model to identify PD patients resorting to rs-fMRI data [146]. The study included 42 patients diagnosed with idiopathic PD and 47 healthy controls. In this case, the group trained a boosted logistic regression model on the correlation matrices. They obtained a mean accuracy of 76.2% (mean sensitivity = 81% and mean specificity = 72.7%). These results reveal evidence of the possibility of identifying PD patients using a ML model based on resting-state between-network functional connectivity [146].

Shi et al. (2022) opted as well for a ML approach on rs-fMRI, using a multicenter radiomics analysis [138]. The primary set included 59 patients with PD and 41 healthy controls, while the external data set used for validation had 27 patients with PD and 16 controls. For feature selection, t-test and least absolute shrinkage and selection operator were applied. As the classifier, the SVM model with a linear kernel was chosen. The mean accuracy was 81.45%, leading to the conclusion that the proposed method can effectively classify patients with PD [138]. Shi et al. also identified the discriminant features included the bilateral superior frontal gyrus, precentral gyrus, right paracentral lobule, precuneus, left inferior temporal gyrus, and superior parietal lobule (all located in the frontal lobe) [138].

DL methods have also been applied in the context of the diagnosis of PD. Some of the studies that resort to DL are referred to in the review article from [8]. The group pointed out the use of 3D-CNN by Esmaeilzadeh et al., with sMRI scans and demographic information from 452 PD patients and 152 healthy controls, which resulted in an accuracy of 100% [147]. Additionally, two studies using graphic convolutional networks, both using sMRI and DTI, are mentioned [148, 149]. Zhang et al. data set was composed of 596 PD patients and 158 controls and their model reached an accuracy of 95.37% [148]. On the other hand, McDaniel and Quinn's data set was smaller, with 117 PD patients and 30 normal controls, and resulted in 92.14% accuracy [149]. In a general scope, these examples of the application of deep learning to the diagnosis of PD use imbalanced data sets, despite presenting high values of accuracy that may be not so reliable. These challenges are important to keep in mind throughout the development of techniques for PD diagnosis using DL and neuroimaging. On an additional note, the mentioned three groups resorted to neuroimaging data from the Parkinson's Progression Markers Initiative (PPMI), a collaborative open data and biosample library.

To overcome the lack of explanation and interpretability of the obtained results, a novel graph attention network (GAT) is proposed by Safai et al [150]. In simple terms, a GAT is a neural network design that uses masked self-attentional layers, in which specific nodes in the neighbourhood are given more attention than others, focusing on the most important area of the graph. The group opted to use a multimodal approach, with brain connectivity information from sMRI, fMRI and DTI. The model from [150] demonstrated a F1-score of 86% and a test accuracy of 73%. These performance metrics scores appear to surpass the referred comparative models. In terms of interpretability, Safai and colleagues

point out "the structural and functional topological influence of motor network and cortico-subcortical brain regions, among which structural features were correlated with the onset of PD". This study shows evidence of the advantage of using multimodal brain connectomic markers and a deep GAT architecture not only to automatise PD diagnosis but also to study how relations between brain regions are affected by this pathology through an attention mechanism-based interpretability framework [150].

From the previously presented study, it is noticed that the study of brain connectivity in PD might reveal an interesting approach. In the scope of PD, there are some studies that opt to focus on functional connectivity, using solely fMRI data. A novel approach to develop automatic procedures to identify PD brain imaging biomarkers, and to classify participants diagnosed with PD, prodromal PD participants, and healthy controls using rs-fMRI data is proposed by Dehsarvi and Smith [151]. This group opts to use an evolutionary algorithm, or EA: the cartesian genetic programming (CGP) to classify dynamic causal modelling (DCM) data obtained from rs-fMRI data (from the PPMI database), focusing on DMN regions. A CGP algorithm benefit is the obtained "white box solution giving more information on the inputs used and enhanced knowledge concerning the final solution obtained in classification" [151]. The findings were compared with the application of an artificial neural network and support vector machine and were validated using k-fold cross-validation. Since the data was highly imbalanced (102 PD participants, 18 prodromal PD participants, and 8 controls), adaptive synthetic sampling was applied. The following performance scores were obtained: PD vs control participants were classified with a maximum accuracy of 92%, PD vs prodromal PD participants with a maximum accuracy of 86%, and prodromal PD vs control participants with a maximum accuracy of 75%. These results revealed that the applied EA algorithm leads to equivalent accuracy values compared to the other classification methods [151].

Guo et. al also opted to use solely fMRI data to build a deep neural network, in this case, to distinguish early stages of PD [152]. The group applied a long-short-term memory (LSTM) network in fMRI mean ROI time-series of 56 subjects at stage 2 and 28 at stage 1 (according to the Hoehn and Yahr scaling [153], with five stages of the progression). A LSTM is a type of recurrent neural network (RNN) which takes the previous information and the current data input to update the cell state and hidden state [152]. They compared the obtained results with a ML approach with a SVM, using FC matrices. The LSTM model presented a higher value of accuracy, F1-score and recall, which reflects in better performance: the LSTM resulted in approximately 75 % of accuracy, while the SVM model presented approximately 58 % of accuracy. Furthermore, it was observed with the LSTM model a higher correlation between the respective results and the results from the motor scores assessment through the questionnaire MDS-UPDRS-III (Movement Disorders Society-Unified Parkinson's Disease Rating Scale) [152]. Additionally, the group studied the interpretability of the LSTM model, from which they concluded that there were identified significant FC changes and disease severity between the stages 1 and 2 of PD. The model's interpretation was assessed by analysing the learnable input hidden weights of the LSTM cell: the ROIs with the magnitudes of the associated weights above mean and standard deviation were considered the most significant for the classification task [152].

As we have seen throughout this review of AI approaches to the diagnosis of PD through neuroimaging, the majority of the studies adopt supervised machine learning classifiers, using statistical or ML algorithms to extract the features [8, 128, 135, 136, 137, 138]. Nonetheless, DL approaches have been emerging, especially with DaTScan images [130, 132, 133, 134] and structural MRI and DTI scans [147, 148, 149]. Regarding rs-fMRI applications, it was found an approach with an EA, which was found not to be so beneficial as expected, compared with an artificial neural network or a SVM [151]. Moreover, the approach of Guo et al. also included the use of rs-fMRI time-series on a LSTM network and FC matrices on a SVM model, which provided an alternative direction to take when

analysing brain functional connectivity in different stages of PD, by using directly the rs-fMRI series instead of computing the FC [152].

In spite of the already achieved advances in the research in the fields of AI and medicine, it is possible to infer that the research still lacks deeper investigation on approaches using deep neural networks to detect PD. Most of the presented studies use small data sets, jeopardising the obtained models' generality. This may be related to the lack of available PD public sets of medical images, which is a challenging limitation for the researchers who work on this pathology. Additionally, the absence of explainability and interpretability from deep models is a restraint that must not be ignored. The study by Safai et al. adopts a complex multimodal approach with an interpretable graph neural network that aims to tackle some of the mentioned issues [150]. It presents robust performance scores and information on the affected brain regions by PD. This study in particular lets us infer that explainable networks may be an interesting and suitable path to take.

Table 3.1: Summary of the approaches to the diagnosis of Parkinson's disease (PD) using artificial intelligence (AI). TL: transfer learning. LogReg: logistic regression. DI: demographic information.

Reference	Year	Image Technique	ML/DL model	Results
[130]	2018	DatScan	Retrained Inception v3 CNN (TL)	sensitivity = 96% specificity = 66.7%
[146]	2019	fMRI	Boosted LogReg	accuracy = 76.2%
[138]	2022	fMRI	SVM	accuracy = 81.45%
[147]	2018	sMRI (and DI)	3D-CNN	accuracy = 100%
[148]	2018	sMRI and DTI	GCN	accuracy = 95.37%
[149]	2019	sMRI and DTI	GCN	accuracy = 92.14%
[151]	2019	fMRI	EA (CGP)	accuracy = 92%
[150]	2022	sMRI, fMRI and DTI	GAT	accuracy = 73%
[152]	2022	fMRI	LSTM (1) and SVM (2)	accuracy (1) = 72% accuracy (2) = 58%

In summary, this state-of-the-art reflects the need to broaden the investigation on the use of DL algorithms for the diagnosis of PD: there is the need to use larger data sets and to improve the training performance of the algorithms. In light of this, using transfer learning (as Kim et al. [130] did with the Inception v3 CNN, a neural network trained with millions of examples of non-medical images) may be an intriguing solution. Furthermore, it is crucial to consider the inclusion of an explainable component (like XAI algorithms, as we will explore later in this chapter), not only to tackle the black-box nature of DL algorithms but also to assess if the model is considering the appropriate features to detect markers of Parkinson's disease. In the scope of the purposes of this dissertation, explainability of the trained DNN on rs-fMRI data is a central step to assess brain functional connectivity in PD. In this way, it is essential to determine which functional connectivity paths influenced by PD have been identified in the literature.

3.2 Study on Brain Functional Connectivity in Parkinson's Disease

Brain functional connectivity alterations related to Parkinson's disease have been explored through several types of metrics and statistical analysis. To assess these FC changes, most studies correlate the identified RSN with motor and non-motor assessments, in order to establish relations between the brain connectivity modification with the cognitive and motor impairments that affect PD patients. In this review, some of the publications that aim to study these correlations will be referred to and analysed.

In 2019, Tessitore et al. published a review "to summarise the most recent rs-fMRI findings highlighting the role of brain networks re-organisation and adaptation in the course of PD" [154]. The group evaluated studies on the diagnosis of PD and motor symptoms correlations [73, 155, 156, 157, 158]; premotor phase and non-motor symptoms correlations [159, 160, 161, 162, 163]; Levodopa-induced dyskinesias [164]. In general, the identified studies showed evidence that abnormal functional interplay within corticostriatal loops correlates with PD identification. Furthermore, this aberrant connectivity seems to be proposed as "early biomarkers to track PD related neurodegeneration pathway" [154]. Despite the many carried studies regarding functional connectivity, Tessitore and colleagues conclude that more research on the identification of rs-fMRI biomarkers of Parkinson's disease is needed since there are still inconsistencies in the literature to be disclosed.

Manza et al. (2015) focused on the study of the resting-state functional connectivity patterns of caudate and putamen subdivisions (parts of the striatum, structures of the basal ganglia) in relation to cognitive and motor impairments related to PD [165]. This study evaluated 62 early-stage PD patients and opted for a seed-based functional connectivity analysis (for each of the seed regions, the fMRI signal time courses were averaged over all voxels). Manza and colleagues found that functional connectivity of the basal ganglia may be related to motor dysfunction and cognitive decline (the major impairments related to PD). While higher motor dysfunction seems to be associated with decreased functional connectivity between the anterior putamen and the midbrain, cognitive decline is related to higher functional connectivity between the dorsal caudate and the rostral anterior cingulate cortex (positioned between limbic and cortical structure) [165]. These results seem to demonstrate that the degree of alteration of functional connectivity, in particular in the basal ganglia, is associated with the symptomology of each PD patient.

A study on the altered functional interactions between brain networks in Parkinson's disease, with 50 patients on dopaminergic medication, was performed by Boon and colleagues [166]. To analyse correlations between the brain regions, the group opted for the Pearson correlation coefficients between the time series of 196 regions. Those values were represented in a connectivity matrix, for each participant. Both static and dynamic functional connectivity were studied. For static FC, they calculated the average connectivity of each considered RSNs with the remaining brain regions. As for dynamic FC, Pearson correlation coefficients were calculated per window per subject, with a window length of 48.6 seconds. Their results show there are relative increases in FC of deep grey matter, the default mode network and the ventral attention network, with the rest of the brain, while the visual network showed decreased FC. Additionally, it was found that executive impairment in PD patients on dopaminergic medication is associated with higher static, but lower dynamic functional connectivity, in particular between deep grey matter areas and the fronto-parietal network. These are intriguing findings since it is concluded that attention networks may also be involved in executive impairment related to PD, in addition to the dopaminergic mechanisms [166].

Cao et al. focused on the evaluation of cognitive functions in PD patients and their correlations with functional alterations in the brain [167]. This group opted for the Eigenvector centrality (EC) mapping, which attributes a value for each voxel in the whole brain, detecting the regions with greater connections with others. Their analysis shows evidence that values in the left inferior occipital gyrus (it communicates with the superior temporal sulcus, where changeable aspects of faces are processed [168]) and the lingual gyrus (access to the hippocampus and other brain regions involved in explicit memory [169]) "were positively correlated with the scores of the Letter Number Sequencing test" (a working memory test) [167]. Furthermore, it was found that the left superior frontal gyrus was associated with immediate recall scores for memory in early PD patients. Results from Cao and colleagues suggest that,

in the early stage of PD, the verified cognitive impairment is correlated with alterations in functional connectivity.

To explore and evaluate the implementation of ICA and graph-theory methods to study the functional connectivity of different pathologies, Dipasquale et al. performed a review of some applications on several neurodegenerative conditions, including Parkinson's disease [170]. Regarding PD, the group included six different studies that describe changes in functional connectivity. Szewczyk-Krolikowski et al. focused their study on the basal ganglia network (networks were isolated through ICA), in a group of early PD patients (on and off medication) compared to a group of healthy controls [171]. The group showed evidence that subjects with PD presented reduced functional connectivity with the basal ganglia network in a wide range of areas, and medicated participants seemed to present significantly improved connectivity. Putcha et al., on the other hand, evaluated primarily three networks that are usually involved in the pathophysiology of many neurological disorders: the default-mode network, the salience network, and the central executive network [172]. Putcha and colleagues analysed 20 subjects with PD and 20 healthy controls, concluding that there is significantly less coupling between the salience network and executive network related to PD, while greater coupling between DMN and the executive network was verified. Karunanayaka et al. opted to focus their study on three groups: healthy controls (24 subjects), PD patients with primary akinetic/rigidity (17 subjects) and PD patients with tremor-predominant symptoms (15 subjects) [173]. The group found that it was verified decreased activity in the left inferior cortex and the left posterior cingulate cortex within the DMN when comparing PD patients with primary akinetic/rigidity with controls and patients with tremor-predominant symptoms. However, these findings were not seen between controls and tremor-predominant patients. This may be related to the hypothesis that PD patients with primary akinetic/rigidity are more likely to develop cognitive deficits, which may be reflected in changes in the DMN [173]. Wei et al. applied the graph-theory method to identify changes in functional connectivity between healthy subjects and PD patients [174]. This group found that the PD subjects presented "decreased efficiency" in the cortico-basal-ganglia motor network, particularly in the rostral and caudal supplementary motor areas primary motor cortex, primary somatosensory cortex, thalamus, globus pallidus and putamen [174]. Zhang et al. recurred to graph theory and a combination of voxel-wise centrality, seed-based functional connectivity, and network efficiency analyses to evaluate the brain connectivity in PD patients with tremor, comparing with healthy controls [175]. In this study, it was found that the PD patients presented increased centrality in the frontal, parietal, and occipital regions, and decreased centrality in the cerebellum anterior lobe and thalamus. It was inferred then that the findings regarding connectivity changes "exhibited a clinically relevant power in distinguishing the patients from healthy controls" [175]. Finally, the study performed by Berman et al. analysed the differences in connectivity between healthy controls and PD patients on and off medication [176]. They found that patients off medication showed increased local efficiency in executive and salience networks, while patients on medication presented a decreased local efficiency in those areas (except within the subcortical network) [176].

As we have seen throughout the analysis of the mentioned studies, functional connectivity changes upon the motor impairment or cognitive dysfunction are triggered by the pathology [154, 165, 166, 167, 170, 171, 172, 173, 174, 175, 176] (see Table 3.2). Several resting-state networks seem to be linked to the many different PD manifestations. In spite of the availability of several studies regarding FC changes in PD in the literature, the inconsistencies verified in the correspondent results, the inherent complexity of the pathological mechanism of PD, which are still not fully understood, the protocol variability in the rs-fMRI data acquisition, pre-processing and in the chose metrics for the calculation of FC correlations, still do not allow a robust and reliable definition of pathological biomarkers of PD related to functional

Table 3.2: Summary of the identified functional connectivity changes related to Parkinson's disease, using resting-state functional magnetic resonance imaging.

Reference	Year	Identified connectivity changes
[154]	2019	Abnormal functional interplay within corticostriatal loops correlates with identification of PD
[165]	2015	(1) Higher motor dysfunction associated with decreased functional connectivity between the anterior putamen and the midbrain, (2) Cognitive decline related to higher functional connectivity between the dorsal caudate and the rostral anterior cingulate cortex
[166]	2020	Relative increases in FC of deep grey matter, the default mode network and the ventral attention network, with the rest of the brain, while the visual network showed decreased FC
[167]	2020	Cognitive impairment is correlated with alterations in FC (left inferior occipital gyrus, lingual gyrus, left superior frontal gyrus)
[171]	2014	Reduced functional connectivity with the basal ganglia network in PD patients
[172]	2015	(1) Less coupling between the salience network and executive network related to PD, (2) greater coupling between DMN and the executive network.
[173]	2016	Decreased activity in the left inferior cortex and the left posterior cingulate cortex within the DMNin PD patients with primary akinetic/rigidity.
[174]	2014	PD subjects presented "decreased efficiency" in the cortico-basal-ganglia motor network.
[175]	2015	(1) Increased centrality in the frontal, parietal, and occipital regions, (2) decreased centrality in the cerebellum anterior lobe and thalamus, in PD patients
[176]	2016	PD patients off medication showed increased local efficiency in executive and salience networks, patients on medication presented a decreased local efficiency in those areas (except within the subcortical network)

connectivity alterations [154, 177]. Notwithstanding that, the need to study how Parkinson's disease affects brain functional connectivity, in order to further comprehend the phenomena behind the onset of this pathology, remains a crucial and relevant objective.

Now that the applications of artificial intelligence on the automated diagnosis of Parkinson's disease have been reviewed, and the resting-state network alterations involved in the pathology have been identified, we can now investigate how explainable artificial intelligence has been benefitting the interpretation of neuroimaging, and how this approach can be translated to the goal of this dissertation.

3.3 Explainable Artificial Intelligence and Neuroimaging

Explainable artificial intelligence (XAI) has been increasingly applied in medical imaging in order to overcome the black-box nature of deep learning models. The lack of explanation given by DL models is a concern among the medical community since they are unable to disclosure which are the biomarkers that lead to the decision made by the model [14, 113, 178]. In addition, there are legal and ethical aspects to be taken into account. Hence, it is crucial to understand the decision and results obtained regarding a patient's condition [179]. The neuroimaging field is no exception to this need: several researchers have

been considering XAI as a tool to explain and interpret decisions of the ML and DL models applied on different types of neuroimages, to evaluate different conditions and pathologies. Throughout this section, we will disclose which are the most commonly used XAI algorithms and to which neuroimaging modalities they have been applied. In particular, the application of XAI on fMRI scans analysis will be explored.

Van der Velden et al. [14] performed a review of the XAI approaches on several anatomical locations and types of medical images. For images of the brain, the group included mostly MRI (for example, [180, 181, 182]), but there are reports of resorting to PET [183], CT [184], ultrasound [185], and histology [186] images. The XAI methods applied to brain images mentioned in the review of van de Velden and colleagues are included in the group of visual explanation approaches (saliency maps). Most of them provide a post-hoc explanation, as class activation mapping (CAM), layer-wise relevance propagation (LRP), (guided) backpropagation, occlusion sensitivity, deconvolution, gradient-weighted class activation mapping (Grad-CAM), and prediction difference analysis, but a model-based approach, trainable attention, has been also applied [14].

An investigation that raises attention aims to study MRI scans using a 3D-CNN network for the detection of PD and 3D class activation maps to add interpretability to the model [181]. Chakraborty et al. concluded that the CAM technique considered the substantia nigra region as the most significant MRI feature in the prediction of the scans, regarding the detection of PD. This is concordant with the pathophysiological phenomena involved in the disease since it is known that one of the hallmarks of PD is the degeneration of dopaminergic neurons in the substantia nigra, which leads to the motor symptomology pattern. This investigation supports the idea and the purpose of XAI: the class activation maps showed that the model paid attention to the adequate brain regions and provided a proper interpretation [181]. Forasmuch the obtained results side with the main objective of XAI and with the pathological mechanisms of PD, the interest of this dissertation will focus on the applications on fMRI data.

Smucny et al. pointed out the current challenges of deep learning applications on neuroimaging, in particular, on fMRI data [113]. They identified the black-box problem as one of the most critical, which can be tackled with XAI. This way, it is not only possible to identify which brain regions and connectivity patterns are given the most attention during the prediction but also to gather diagnostic biomarkers [113]. Farahani et al, who review the XAI applications to neuroimaging until the date (end of 2022), refer to some examples of XAI approaches to fMRI [178].

The first application goes back to 2005, in work performed by Mourão-Miranda et al. [187]. This study aimed to assess multiple brain states from whole fMRI volumes without prior selection of spatial features. For classification, they opted to apply the SVM and Fisher linear discriminant algorithms and compared them, while for the identification of the significant and discriminant regions, through XAI, they applied spatial maps. Their results showed that regions selected as significant were located in the expected brain areas [187]. Wang et al. (2007) also opted to resort to spatial maps for explainability [188], in order to analyse multiple fMRI acquisitions between different subjects through a SVM model. The authors aimed to study the multivariate nature of fMRI, considering sensorimotor or cognitive tasks.

Li and colleagues (2018) focused on the prediction of autism spectrum disorder (ASD) from fMRI using deep learning [189]. Moreover, the authors pretended to address the problem of interpreting reliable biomarkers associated with identifying ASD by resorting to saliency features, extracted through a corrupting strategy (based on "corrupting" sections of the original photos and evaluating how the model's prediction changes after being trained on the original, uncorrupted data). For the classification, Li et al. opted to apply a simple CNN, and they resorted to both synthetic and real fMRI data. The

proposed method for biomarkers identification and ROI interpretation is said to present efficiency on the explainability of DL models applied to neuroimaging [189].

Thomas et al. (2019) developed DeepLight, a framework that includes a long-short-term memory model (LSTM) to analyse the brain networks from fMRI data, resorting to the Human Connectome Project [190]. The data set included data from 100 subjects viewing images of body parts, faces, places, and tools. To explain the decisions produced by the DL model, the group opted to apply a LRP technique. The group demonstrated that, in decoding the cognitive states and identifying the physiologically relevant brain regions relative to each considered state, the taken approach with a LSTM model and LRP technique outperforms traditional methods of fMRI analysis (such as general linear model, whole-brain Lasso, searchlight analysis), in particular in terms of decoding accuracy and F1-score, in both group-level and subject-level analysis. In addition, the group states that the proposed framework for the analysis and interpretation of neuroimaging data scales well to large data sets, and preserves the interpretability of the data [190].

The mentioned studies regarding the use of XAI in the context of PD [181] and the application on fMRI scans [187, 188, 189, 190] are gathered in Table 3.3.

Table 3.3: Summary of the used methods of explainable artificial intelligence in Parkinson's disease and functional magnetic resonance.

Reference	Year	Pathology	Neuroimage technique	AI model	XAI method
[181]	2020	PD	MRI	3D-CNN	CAM
[187]	2005	N/A	fMRI	SVM	spatial maps
[188]	2007	N/A	fMRI	SVM	spatial maps
[189]	2018	ASD	fMRI	CNN	saliency features
[190]	2019	N/A	fMRI	LSTM	LRP

As observed throughout this analysis, the applications of XAI on neuroimaging have been growing and revealing themselves as useful and valuable tools for the interpretability of various models and the search for biomarkers [14, 113, 178]. It seems that XAI may become a fundamental aid for physicians in the diagnosis of many neurological pathologies, like Parkinson's disease, aiming to find what are the features in the medical image that discriminate the condition of the given patient [14].

In light of this dissertation's objectives, and considering what has been analysed throughout the state-of-the-art, there are two main questions that are aimed to be answered with the use of XAI:

1. Which brain connectivity patterns and RSN are involved in the pathophysiology of Parkinson's disease, and how do they change relative to healthy controls?

2. Is the developed deep learning model considering adequate features on the classification between healthy controls and patients, considering the findings in the literature regarding the pathology in question?

3.4 Overview

This review aimed to present the current state of the use of AI for PD diagnosis, what has been discovered regarding the brain functional connectivity changes related to PD, and the recent advances on XAI applied to neuroimaging.

As discussed earlier, the main limitations regarding the current investigation on the use of DL models for PD diagnosis were identified. Most of the studies displayed small and imbalanced data sets but,

overall, they showed high values of accuracy, which are possibly not so reliable. Moreover, in the medical context, the shortage of explainability of DL is seen as an adversity, considering that understanding and interpreting the clinical outcomes given by those models is crucial. To overcome this setback, XAI emerges as an innovative solution, aiming to add interpretability to the diagnosis DL models, and also to identify and develop pathological biomarkers.

Furthermore, the study on functional connectivity changes in PD, in particular in resting-state networks, has become an interesting topic to broadly explore, in regards to the pathological phenomena of the disease and their association with the symptoms and signs.

Considering the above, an opportunity to use different AI approaches to study the brain functional connectivity changes in Parkinson's diseases arises, which will be investigated throughout this dissertation project. Hence, one of the purposes of this work is to build an optimised deep learning model for fMRI data classification, to discriminate PD patients and healthy controls. Different from other studies, in this work, we will explore the hyperparameter optimisation of the network in order to obtain an optimal performance. In addition, XAI techniques will be applied to the trained model in order to obtain the features (which will correspond to connectivity networks) that were the most significant to the classification, and then assess how the brain connectivity from PD patients is distinct from healthy subjects. Different XAI methods will be compared to assess which technique produces the most reliable explanations. This workflow will allow the construction of a roadmap to build a more systematic application of XAI methods in this field.

Chapter 4

Materials and Methods

Based on the scientific background presented in Chapter 2, and considering the state-of-the-art on the use of AI for Parkinson's disease diagnosis, brain functional connectivity in Parkinson's disease and XAI and neuroimaging, the following methods were followed to fulfil the objectives of this dissertation.

4.1 Data

We used the Parkinson's Progression Markers Initiative database to obtain the fMRI scans from PD patients and healthy controls. PPMI is a project cooperating with partners worldwide to establish a strong open-access data collection and biosample library to accelerate scientific discoveries and novel therapies [191]. This study included approximately 1000 enrolled people (500 PD patients, 100 healthy controls and 400 with prodromal PD) between 2018 and 2020, and 3000 subjects between 2020 and 2023 (900 PD patients, 100 healthy controls and 2000 with prodromal PD) [192]. PPMI divide their study into two groups: the clinical data collection and the biological collection. In the clinical data collection, besides the biomedical images collection (which includes DatScan and MRI scans), there are motor assessments, neurobehavioral/neuropsychiatric testing, autonomic, olfaction and sleep testing and online patient-reported outcomes.

From the PPMI we need two biomedical imaging data: fMRI scans and the corresponding sMRI scans (3D T2 FLAIR), from the same visit/date. The first step was to verify which subjects had fMRI and sMRI images scanned on the same day. It was selected as the date, the baseline visit, i.e., the first visit from the subjects to the PPMI acquisitions. From this data filtering resulted in 120 PD patients (mean age 63.4 years, minimum 32, maximum 84) and 22 healthy controls (mean age 61.2 years, minimum 32, maximum 85) fMRI and sMRI scans.

As noticed, there is a substantial difference in the PPMI data between PD patients and healthy controls. Hence, not only there is a limitation in the small number of available data, but we also face an imbalanced data problem. To overcome this hurdle, we thought about the possibility of including healthy control data from a different cohort, but with similar subjects' age range and MRI acquisition parameters. As a matter of fact, studies show that conventional resting-state metrics, such as functional connectivity, are not significantly affected by different TRs [193]. Hence, a study that checks out these characteristics, with a similar age range (with a mean age of the selected group of healthy controls of approximately 66.5 years, minimum 56 and maximum 88), and similar MRI acquisition parameters to PPMI, is the Alzheimer's Disease Neuroimaging Initiative (ADNI) [194]. ADNI aims to better understand the progression of Alzheimer's disease using data such as MRI and PET imaging, genetics, cognitive assessments and blood biomarkers. To obtain a balanced data set, we selected 131 healthy

controls. The distribution of the number of subjects used in this work is displayed in Figure 4.1.

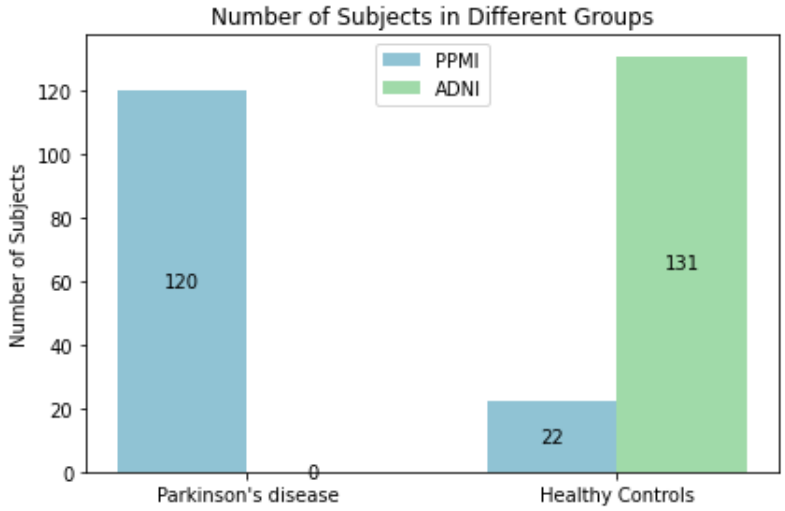

Figure 4.1: Distribution of Parkinson's disease subjects and healthy controls, from PPMI and ADNI data sets.

The following tables present the acquisition parameters for PPMI and ADNI data sets.

Table 4.1: fMRI acquisition parameters from PPMI and ADNI.

Data set	PPMI	ADNI
Manufacturer	SIEMENS	SIEMENS
Field Strength (T)	3.0	3.0
Repetition time (TR) (ms)	2500	3000
Echo time (TE) (ms)	30	30
Slice thickness (mm)	3.5	3.4
Number of slices	240	197
Matrix X × Y (pixels)	448 × 448	448 × 448
Field of View (mm)	224 × 224	220 × 220
Flip angle (°)	80	90
Time of acquisition (min)	10	10

4.2 Functional Magnetic Resonance Imaging Pre-processing

After retrieving the fMRI and sMRI data, the next step is to pre-process them, as both PPMI and ADNI only have the raw data available. This is a very important step, that takes a long time to complete, since from the pre-processed data we will be able to compute the connectivity matrices to input in our DNN and, consequently, to evaluate how the functional connectivity in the brain changes between PD patients and healthy subjects.

The selected pre-processing pipeline for this work followed the default pre-processing pipeline from the toolbox CONN [195]. CONN is an open-source SPM-based toolbox for the computation, display, and analysis of fMRI. CONN is used to analyse resting-state or task-related data. SPM (Statistical Parametric Mapping) is an fMRI analysis package that is run in MATLAB. In this work, during the fMRI pre-processing, the MATLAB version R2022b was used [196].

The default flexible preprocessing pipeline includes realignment with correction susceptibility distortion interactions, slice timing correction, outlier detection, direct segmentation and MNI-space

normalisation, and smoothing on the functional and anatomical data [197]. Functional data were realigned using SPM *realign & unwarp* procedure [198] (refer to Figure 4.2), where all scans were coregistered to a reference image using a least squares approach and a 6-parameter (rigid body) transformation [199], and resampled using b-spline interpolation to correct for motion and magnetic susceptibility interactions.

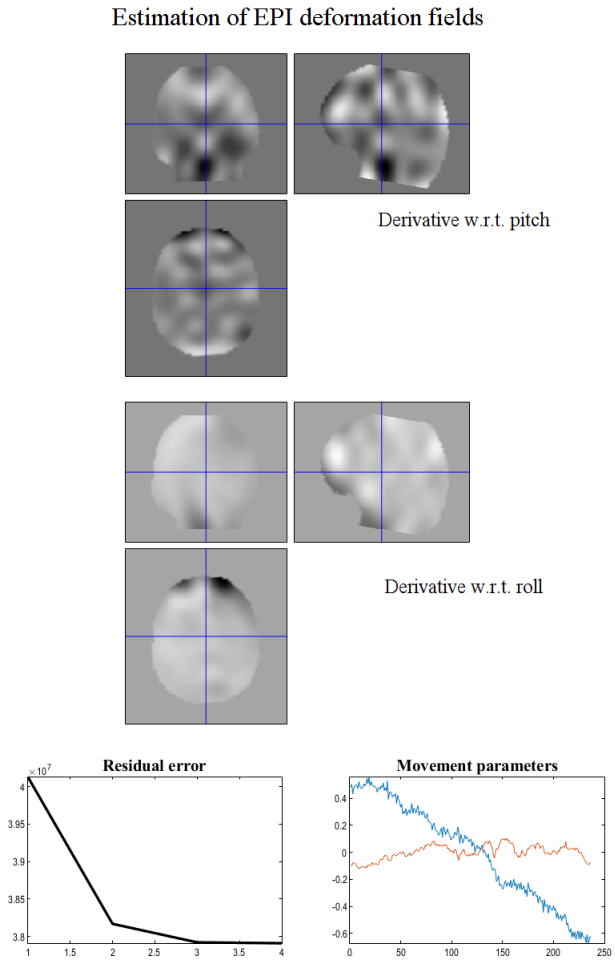

Figure 4.2: Representation of the echo-planar imaging (EPI) deformation fields, obtained from the functional realignment and unwarp step of the pre-processing of one single subject.

Temporal misalignment between different slices of the functional data (acquired in ascending order) was corrected following SPM slice-timing correction procedure [200, 201], using sinc temporal interpolation to resample each slice BOLD timeseries to a common mid-acquisition time.

Potential outlier scans were searched using artefact detection tools (ART) [202] as acquisitions with framewise displacement above 0.9 mm or global BOLD signal change above 5 standard deviations [203, 204], and a reference BOLD image was computed for each subject by averaging all scans excluding outliers. None of the subjects was identified as outliers, hence all scans were considered in the analysis.

Functional and anatomical data were normalised into standard MNI space, segmented into grey matter, white matter, and CSF tissue classes, and resampled to 2 mm isotropic voxels following a direct normalisation procedure [204, 205] using SPM unified segmentation and normalisation algorithm [206, 207] with the default IXI-549 tissue probability map template. In addition, the functional and anatomical data were segmented into the 14 resting-state networks defined by Shirer et al. [83]. Lastly, functional data were smoothed using spatial convolution with a Gaussian kernel of 8 mm full-width half

maximum (FWHM).

Furthermore, functional data were denoised using a standard denoising pipeline [197]. This pipeline included the regression of potential confounding effects characterised by white matter timeseries (5 component-based noise correction - CompCor - noise components), CSF timeseries (5 CompCor noise components), motion parameters and their first-order derivatives (12 factors) [208], outlier scans (below 0 factors) [203], session effects and their first order derivatives (2 factors), and linear trends (2 factors) within each functional run. This was followed by bandpass frequency filtering of the BOLD timeseries [209] between 0.008 Hz and 0.09 Hz. CompCor [210, 211] noise components within white matter and CSF were estimated by computing the average BOLD signal as well as the largest principal components orthogonal to the BOLD average, motion parameters, and outlier scans within each subject's eroded segmentation masks.

To ensure the correctness of all the pre-processing steps of the fMRI scans, quality assurance plots were assessed. In particular, the registration of the grey matter ROI outline on the functional data (Figure 4.3) and on the structural data (Figure 4.4) were plotted. Additionally, the quality assurance of the normalisation of the functional data (Figure 4.5) and the structural data (Figure 4.6) on the MNI coordinate space was also assessed.

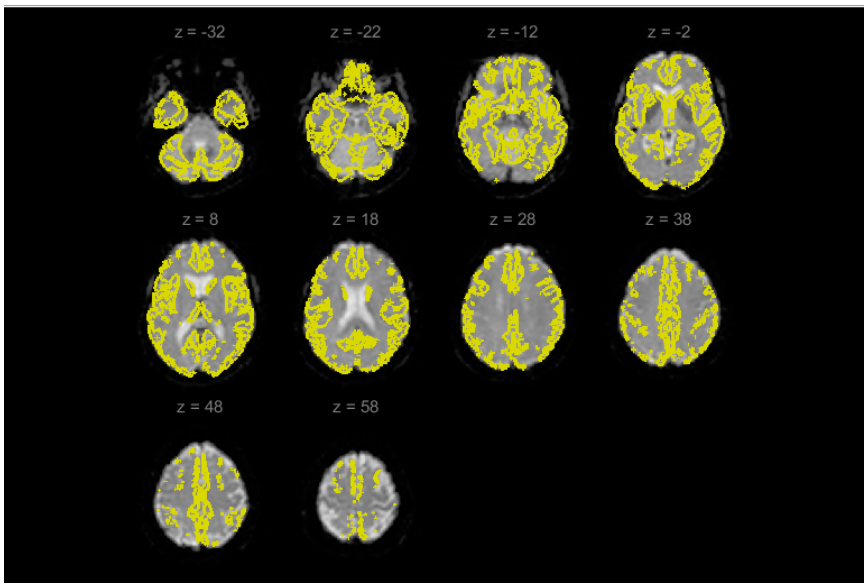

Figure 4.3: Quality assurance of the registration of the functional data on the outline of the grey matter (correspondent to one subject).

4.2.1 Computation of Functional Connectivity Matrices

Following the pre-processing and denoising of the fMRI scans, we proceeded to the calculation of the functional connectivity matrices. For the correlation metric, we opted for the ROI-to-ROI connectivity measures, calculated as the Pearson correlation coefficients between each pair of ROIs (refer to Equation 2.6) and the respective Fisher-transform (Equation 2.7). For more information on the scientific background regarding the assessment of brain connectivity, refer to subsection 2.2.3.

ROI-to-ROI connectivity matrices were estimated characterising the functional connectivity between each pair of regions among the considered 14 resting-state networks defined by Shirer et al. [83]. Functional connectivity strength was represented by Fisher-transformed bivariate correlation coefficients from a general linear model (weighted-GLM [197]), estimated separately for each pair of ROIs,

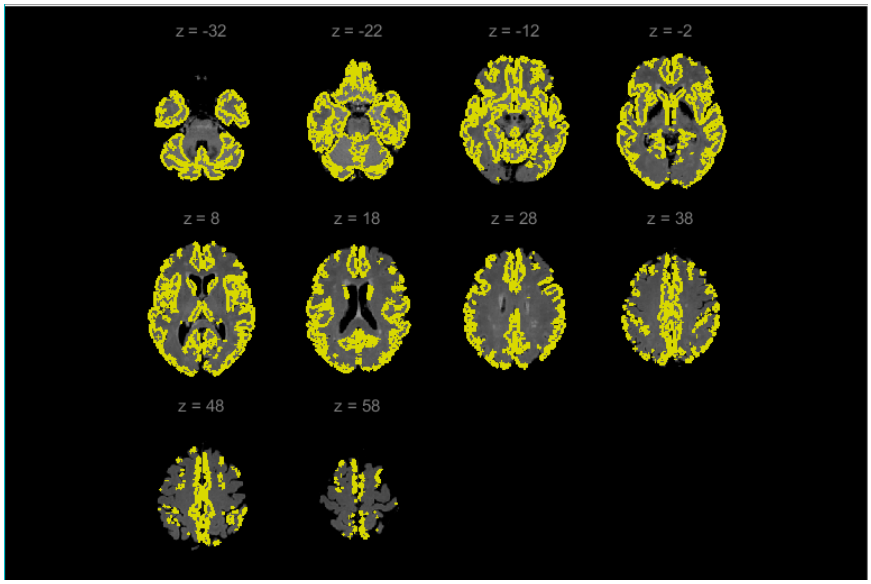

Figure 4.4: Quality assurance of the registration of the structural data on the outline of the grey matter (correspondent to one subject).

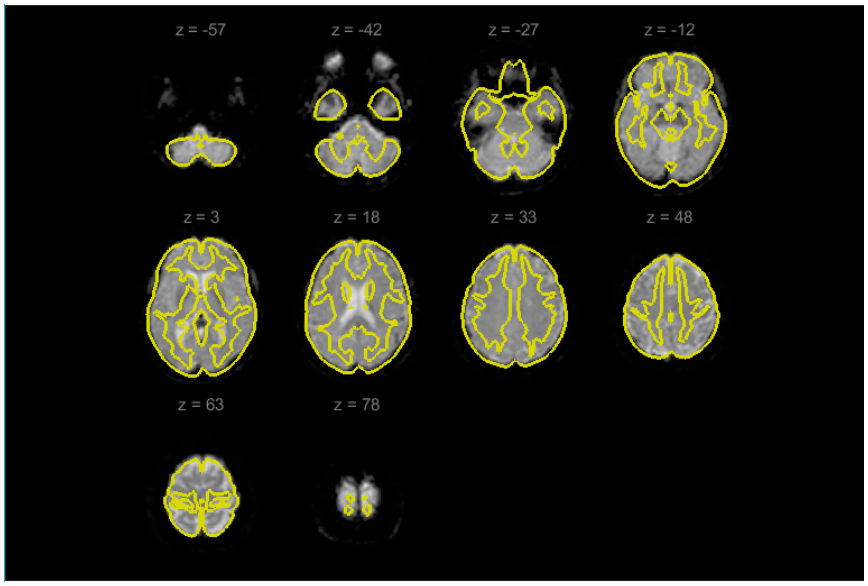

Figure 4.5: Quality assurance of the normalisation of the functional data on the MNI template (correspondent to one subject).

characterising the association between their BOLD signal timeseries. In order to compensate for possible transient magnetisation effects at the beginning of each run, individual scans were weighted by a step function convolved with an SPM canonical hemodynamic response function and rectified.

4.3 Building the Classification Deep Learning Model

After the computation of the functional connectivity matrices, we can proceed to the development, training and evaluation of a deep neural network for the classification of PD and healthy controls. As previously mentioned, in the scope of this project, the FC matrices will be treated as images and fed to a convolutional neural network (CNN). To develop the CNN, the TensorFlow-based Keras API [212] was used.

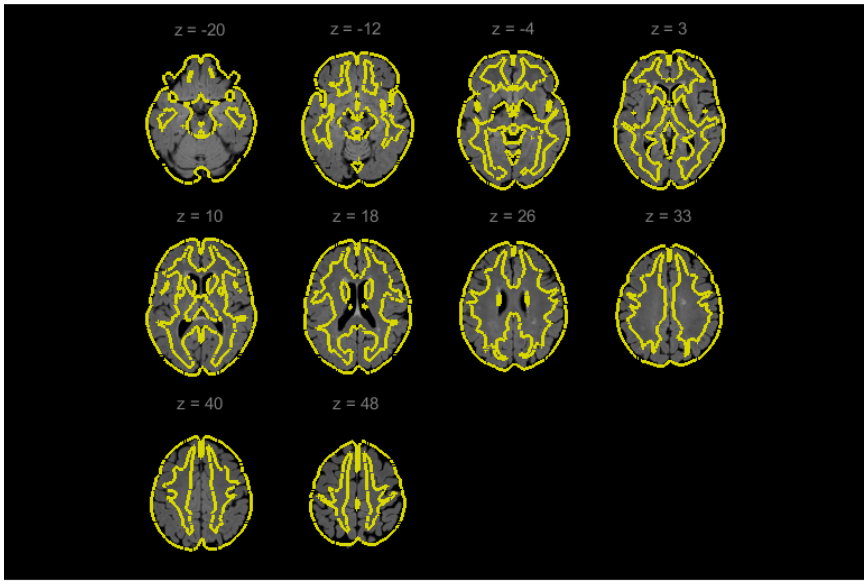

Figure 4.6: Quality assurance of the normalisation of the structural data on the MNI template (correspondent to one subject).

Since the correlation values were already normalised through the Fisher transform, the matrices are ready to be directly inputted into the model. The values in the diagonal were set to zero since they give no information on the functional connectivity of the brain (the diagonal values correspond to correlation with pairs of the same network, which does not have significance in this context). Furthermore, although the symmetry of the FC matrices can induce redundant information, we opted to use the complete matrices (i.e., not use solely triangular matrices), since eliminating information from the matrices could affect the convolutional operations on them.

The architecture of the used CNN was inspired in the ConnectomeCNN, a CNN architecture created to evaluate brain connectivity matrices and classify subjects according to the connectivity patterns of brain regions [213]. The ConnectomeCNN is composed of two convolutional layers, with squared kernel filters with a stride, and two fully connected layers, with the output layer with two output neurons that correspond to the two classes, in which is applied a softmax function to calculate the probabilities of each subject belonging to each class.

To our CNN, which we named ExtendedConnectomeCNN, we added one convolutional layer to the whole architecture. The ExtendedConnectomeCNN includes three convolutional layers (Conv2D), with kernel filters with size 3 by 3 on all three layers, with a stride equal to 2. It was verified that a stride of 2, instead of 1, produced better performance values when handling our data. Our network receives an input size of $[14 \times 14 \times 1]$ for each one of the subjects. Contrarily the tendency of the original ConnectomeCNN, the number of filters on the convolutional layers is decreasing by a factor of 2: the first convolutional layer has 256 filters, the second has 128 filters, and the last one has 64 filters. This architecture with a decreasing number of filters is inspired by the Keras Convolutional Autoencoder [214], and it seems to produce better results on performance in our case. To each convolutional layer, a max pooling layer was added, with a pool size of 2 by 2. In addition, as overfitting avoidance techniques, a batch normalisation layer and a dropout layer were added to each convolutional layer. The rate of dropout was one of the hyperparameters selected to be optimised, as well as the presence of the batch normalisation layer. As activation functions, we opted for the ReLU. The fully connected layer, preceded by a flatten layer, includes 128 units. This layer also includes a dropout layer and ReLU activation. The library for application of XAI methods only accepts as input models with ReLU

activation, and not other variations such as Leaky ReLU and PReLU. Finally, the output layer of the ExtendedConnectomeCNN is similar to the ConnectomeCNN: it includes the two output units that are obtained through the softmax activation, and correspond to the probabilities of a subject being associated with one of the classes 'Healthy' ([1 0], or 'Parkinson's' ([0 1]). Figure 4.7 shows the architecture of the proposed ExtendedConnectomeCNN.

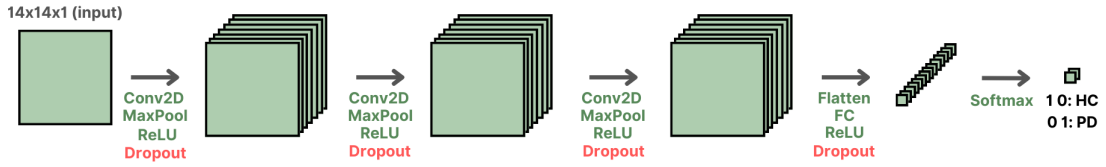

Figure 4.7: Architecture of the ExtendedConnectomeCNN, a convolutional neural network based on ConnectomeCNN.

As for the training parameters, we opted for a number of epochs of 200 and a batch size of 16. We chose the He initialisation of the weights and bias parameters, and the L2 regularisation as the penalty term of the loss function. For learning rate optimisation, we selected the Adam optimiser. The initial learning rate was selected as a hyperparameter to be optimised. The loss function is obtained through binary cross entropy. For more details on the theoretical background of the training parameters of a DL model, refer to the subsubsection 2.3.1.1.

The first step is to divide the data set into a development set [1] (90 %) and a test set (10 %). For this, the function *train_test_split* with the option *stratify* to ensure that classes are evenly distributed in both subsets was applied.

To construct the model and optimise the hyperparameters, a development part will take part, where we will use the development set. On the development data set, stratified 10-fold cross-validation will be applied. Stratified cross-validation ensures that the proportion of samples for each class is preserved during the process of cross-validation. Several model architectures with different combinations of hyperparameters were tested during this step so that the model with the most suitable performance on the cross-validation, i.e., the most generalisable model, is chosen. To evaluate the performance, the average values of training and validation accuracy, AUC ROC, sensitivity, specificity and F1-score of the obtained 10 models from the 10-fold cross-validation are calculated. The equations of the used evaluation metrics are in the subsubsection 2.3.1.1. The table below displays the tested hyperparameters during the cross-validation part [2].

After assessing the hyperparameters corresponding to the most generalisable model on the development data, we can proceed to the training of the final model and evaluate its performance on the separate test data, which corresponds to 10 % of the whole data set. The final model will have the same hyperparameters as the architecture with the best results on cross-validation, although this final model will be trained on the whole development set, which will be referred to as the training set from now on. After the training step of the final model, we move to the evaluation part, which corresponds to the testing step on the test set. The results of the separate test set will be used to study the XAI methods and, consequently, the functional connectivity alterations related to Parkinson's disease. The workflow of the development of the CNN classification model for functional connectivity matrices corresponding to healthy controls and PD patients is presented in Figure 4.8.

[1] The denominated development set is usually referred to as training set, but in this case, we named it development set since this data set is used to fine-tune the model in cross-validation.

[2] More combinations than the presented were tested. However, not all the combinations are presented in the dissertation, since they produced redundant results.

Table 4.2: Tested models during the hyperparameters' optimisation step. The hyperparameters to be optimised are the learning rate, the dropout rate and the presence of a batch normalisation layer on convolutional layers.

ID	Learning Rate	With Batch Normalisation	Dropout Rate
1	0.001	Yes	0.3 in all layers
2	0.0001	Yes	0.3 in all layers
3	0.00001	Yes	0.3 in all layers
4	0.0001	Yes	0.5 in all layers
5	0.0001	Yes	0.6 in all layers
6	0.00001	Yes	0.3 in conv layers, 0.5 in FC layer
7	0.00001	No	0.1 in conv layers, 0.4 in FC layer
8	0.00001	No	0.5 in all layers

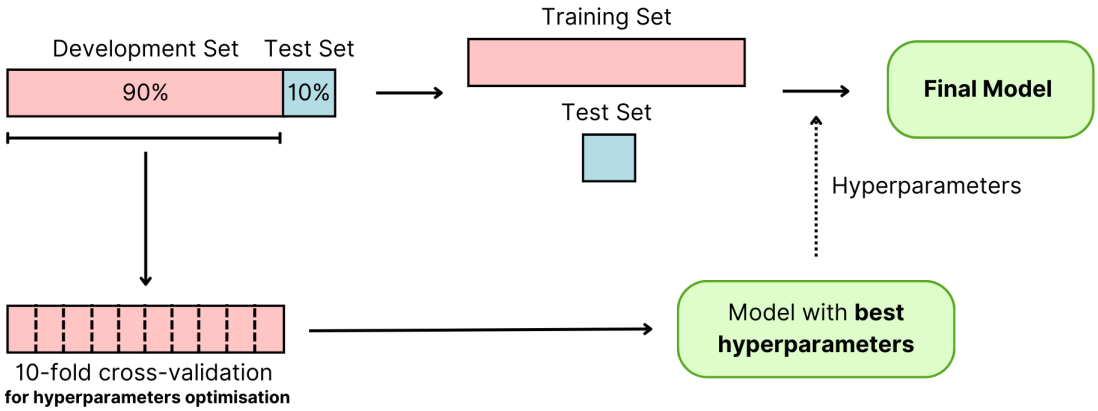

Figure 4.8: Pipeline for construction of the classification CNN model.

4.4 Interpreting the Model: Explainable AI

With the final model built, we may now proceed to its interpretation and explainability. As aforementioned, there are three main objectives in this part of the work:

1. To compare and evaluate different XAI methods, in particular layer-wise relevance propagation, deconvolutional network and guided backpropagation, which are methods commonly applied to CNN;

2. To gain an understanding of how the model is handling the FC matrices globally and locally;

3. To compare the results from XAI analysis with the literature regarding functional connectivity analysis in Parkinson's disease.

To accomplish the mentioned goals, the Keras and TensorFlow 2-based toolbox iNNvestigate was used. This library was developed by Alber et al. and includes various XAI methods, including LRP, DeconvNet and guided backpropagation [215].

The XAI methods were applied to the classified test data (corresponding to 10 %) of the data. Since this classification corresponds to a set considered as "external" to the optimisation and training process of the final model, it is considered appropriate to evaluate the explainability of the test set, instead of the called development (or training) set.

Before applying any of the XAI methods, the first step to take is to remove the last layer of the final model, after its training and evaluation. The final layer corresponds to the output activation layer, the softmax function layer. It is a common practice to remove the activation layer from the model before applying an XAI method so that it will focus on the weights from the layers that select and extract features that will be discriminative in the classification of Parkinson's disease through 14×14 FC matrices.

Using the function *innvestigate.create_analyzer()*, the three models that we applied were created: LRP with the epsilon rule, LRP-ε, DeconvNet and guided backpropagation. The three methods give a local explanation for each input and are backpropagation-based. While LRP uses a rule equation (Equation 2.22) to backpropagate from the output class to the input class, deconvolutional network and guided backpropagation are based on deconvolution operations (see Figure 2.15). Refer to subsubsections 2.3.3.1 and 2.3.3.2 to get to know more about the theoretical background of the selected XAI methods.

Then, the three methods are applied to the model without softmax activation layer, on the test set inputs. When applying each one of the XAI methods, it is possible to select which one of the output neurons (0 for healthy controls and 1 for PD patients) the method should focus on. Since this study aims to assess functional connectivity changes associated with Parkinson's disease, it makes sense to choose the output neuron '1', which corresponds to the disease class. The results of each explanation are represented in heatmaps, for each subject individually. A heatmap assigns to each pixel (in this case, one pixel corresponds to one element of the FC matrix) an explanation value given by the XAI analyser function.

Afterwards, for each XAI method, a global analysis was performed by computing the mean heatmaps of the true positive instances. The true positive instances are the most significant in this analysis since the explanations that are the most interesting correspond to the ones that represent the significant features of correct classification of the subjects with Parkinson's disease. From the mean heatmaps of the XAI methods, the 10 top explanation values were retrieved, to assess the most significant pairs of resting-state networks for the classification of PD. Those lists of networks will be further analysed and compared with the current literature on the alterations of functional connectivity related to Parkinson's disease.

As expected, the three applied XAI methods will result in different explanations' heatmaps. To assess which method is the most suitable and accurate, a comparison method was applied. The metric proposed by Samek et al. was created keeping in mind the evaluation of the heatmaps that result from an XAI analysis. The authors aimed to create objective criteria for heatmaps' evaluation to overcome the inherent human prior knowledge and experience, which more often than not, leads to biased and inaccurate decisions [216]. The group proposes an iterative procedure to assess the quality of a heatmap, which involves measuring how the class attributed to one image vanishes while gradually applying perturbations to the image, i.e., removing information from the image at specific locations. The first step of this process is to order the defined regions of the images with descending relevance. Then, starting with the most relevant locations, those will be iteratively removed through a specific function. This method is referred to as *most relevant first*: MoRF [216]. Then, a metric of interest is calculated: the area over the MoRF perturbation curve (AOPC), that is given by:

$$AOPC = \frac{1}{L+1}\langle \sum_{k=0}^{L} f(x_0^{MoRF}) - f(x_k^{MoRF}) \rangle_{p(x)} \tag{4.1}$$

with L being the number of iterations of the MoRF method, f the function correspondent to the classification DL model, x_0^{MoRF} the original image and x_k^{MoRF} the perturbed image after k iterations. $\langle \rangle_{p(x)}$ denotes the average over all images. A region order in which the most sensitive areas are ranked first suggests a rapid decline in the MoRF graph, and hence a bigger AOPC [216]. Then, larger values of AOPC correspond to an algorithm that better explains the classification.

Hence, to evaluate and compare the LRP, DeconvNet and guided backpropagation methods applied to this work, the computation of the AOPC reveals itself as a suitable metric. The first step is to get the coordinates that correspond to the L most relevant features (i.e. pairs of resting-state networks with the highest relevance score). Then, iteratively, the algorithm adds a perturbation to each one of the most relevant elements of the matrices (remind that we are only considering the true positive matrices): in the first iteration, the most relevant element is perturbated, in the second iteration, the two most relevant elements, and so on. In our case, we opted for a number of iterations L equal to 100. As for the perturbation, we switched the most relevant elements to 0.

After computing the x_k^{MoRF} matrices, we computed the prediction using the classification model, our ExtendedConnectomeCNN, on each one of the matrices, $f(x_k^{MoRF})$, and applied the Equation 4.1 to compute the AOPC. The values of the AOPC on each iteration were plotted in a graphic of a function of the number of iterations L.

The complete pipeline of the application of the XAI methods and the respective interpretation and analysis in the scope of this dissertation are presented in Figure 4.9.

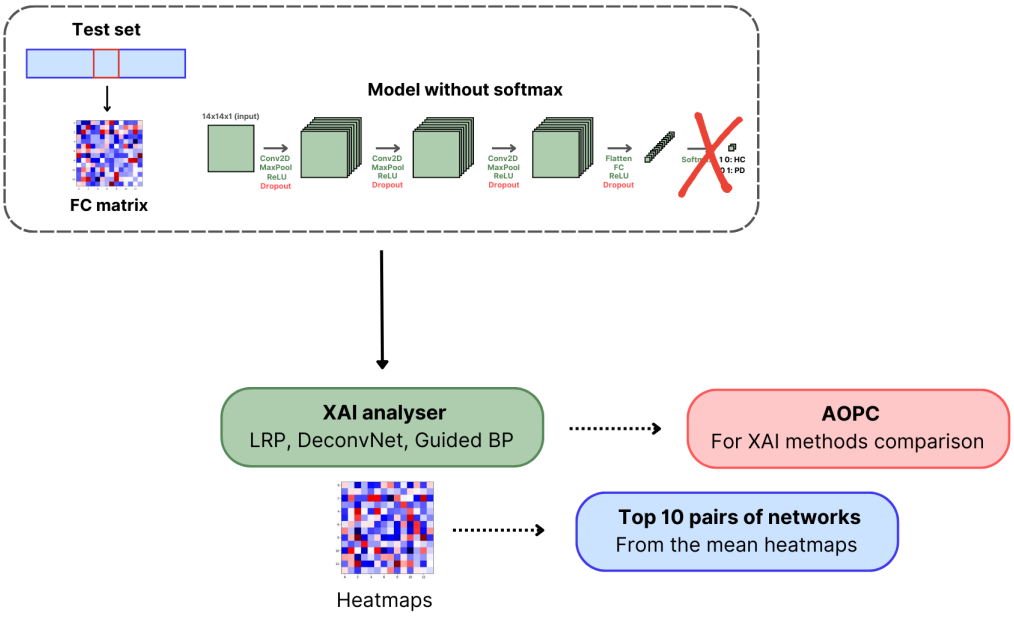

Figure 4.9: Pipeline for the application of the XAI methods and respective interpretation.

4.5 Graphical Abstract

To sum up the steps taken in the methods of this project, a graphical abstract of the dissertation is presented in Figure 4.10.

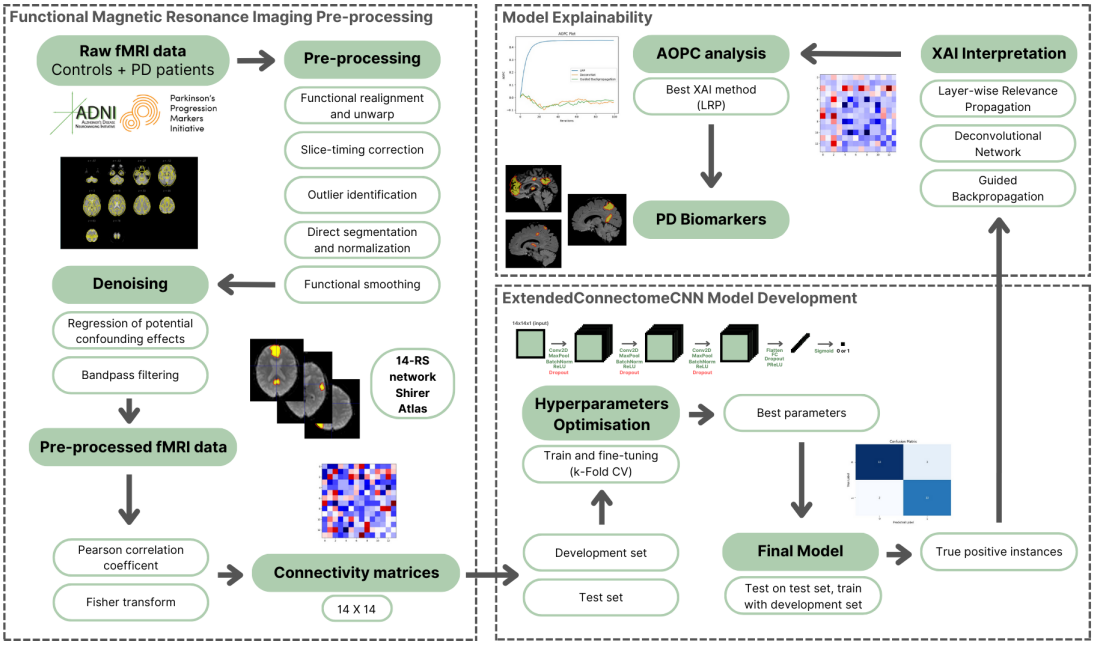

Figure 4.10: Graphical abstract of the dissertation work.

4.6 Additional Methods

In this section, the additional methods apart from the main work pipeline followed in the scope of this dissertation will be presented. The additional methods correspond to an attempt to create a model with the ExtendedConnectomeCNN on the ABIDE data set in order to perform transfer learning of that model on the PPMI data.

4.6.1 Training the Model on ABIDE data set

The size of the data set used for this work (120 PD patients and 153 healthy controls) is considered small in the context of deep learning models. In fact, the lack of medical data, in terms of size and quality, to be applied in machine learning models, is a common issue when developing AI-aided techniques of diagnosis, as mentioned earlier in this dissertation. A solution often proposed in the context of medical imaging is to apply transfer learning with a large-size data set such as the ImageNet [217]. In subsection 2.3.2, there is further information regarding the current challenges in deep learning and neuroimaging.

Throughout this work, the possibility of recurring transfer learning to tackle the lack of data in our data emerged. In this case, it was thought to use a larger data set of the same nature: a set of functional connectivity matrices, with a class of the disease and other controls. One example is the Autism Brain Imaging Data Exchange (ABIDE). The ABIDE is a public source of preprocessed neuroimaging data that includes data from 539 autism spectrum disease patients (ASD) and 573 controls, with an age range of 5-64 years, comprising structural and resting-state functional MRI and phenotypic information. In the case of this data set, the functional connectivity matrices can be retrieved from the already pre-processed

fMRI scans. The used atlas was the AAL, which resulted in FC matrices with 116 × 116 shapes. We have available 468 FC matrices from healthy controls and 403 from ASD patients.

There are multiple papers that refer to the use of deep learning and the study of functional connectivity patterns to classify ASD patients against healthy controls, with a state-of-the-art accuracy of approximately 70 % [218, 219]. However, one of the main challenges of the use of ABIDE is the heterogeneity of the data set. The disease itself is manifested widely differently in between patients since its pathophysiology is still not fully comprehended and involves several brain regions [220]. Furthermore, the ABIDE includes data acquired with various combinations of parameters of acquisition, which may also contribute to the diverseness of the data set.

Nonetheless, an attempt to build a model with the ExtendedConnectomeCNN with ABIDE data set to apply transfer learning to the PPMI + ADNI data set was performed. Similar to the process described in Section 4.3, the best hyperparameters were adjusted by performing cross-validation on the development set. Only the dropout rate was optimised. The remaining parameters were the same as the best model retrieved in the previous methods. Then, the purpose would be to train a final model to be evaluated on the test set, and then the model with the trained weights would be used to transfer learning: the convolutional layers' weights would be frozen, and the fully connected layers would be re-trained on the PPMI + ADNI data set (see Figure 4.11). The dropout rate values that were used are presented in Table 4.3.

Table 4.3: Tested models for the ABIDE classification model. Only the dropout rate was optimised. The hyperparameters were iteratively adjusted accordingly to the model performance on 10-fold cross-validation.

ID	Dropout Rate
1	0.5 in all layers
2	no dropout layers
3	0.1 in conv layers, 0.4 in FC layer

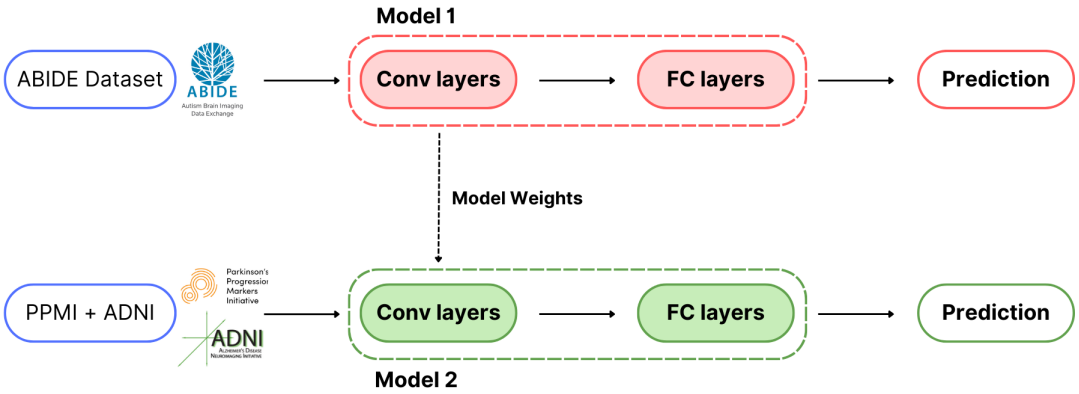

Figure 4.11: Pipeline for transfer learning of the classification CNN model trained on the ABIDE data set, to the PPMI + ADNI set.

Chapter 5

Results and Discussion

In this chapter, the results and findings from the work of this dissertation and the respective discussion are presented. Firstly, we present the results corresponding to the pre-processing of the fMRI data and the computation of the functional connectivity matrices. Afterwards, the results from the design and evaluation of the CNN classification model are introduced and discussed. Additionally, we will examine the results from the application of the XAI methods to our model, and analyse the respective interpretation. Finally, we will inspect if the transfer learning approach with the ABIDE data set was feasible.

5.1 Obtained Pre-processed Functional Magnetic Resonance Imaging Data

Here we introduce the results subsequent to the pre-processing pipeline of the fMRI scans from the PPMI and ADNI data sets. This section of the dissertation's project is important, despite taking a considerably long time to take on. For that reason, it is necessary to guarantee that the steps of this process were correctly followed, as the pre-processing steps highly affect the results for the connectivity matrices [221]. This way, we obtain acceptable FC matrices to be fed to our classification model.

In Figure 5.1, an example from a subject with Parkinson's disease of the resulting segmentation of the ROIs defined by Shirer et al. [83] is presented. Comparing the results from our segmentation with Figure A.1, which represents the segmentation performed by the authors, it is possible to affirm that the segmentation performed in this work is fairly correspondent to the segmentation by Shirer et al. [83]. Thus, we may infer that the segmentation of the ROIs was successful.

Furthermore, analysing the quality assurance plots corresponding to the examination of the registration of the fMRI scans and the anatomical scans, and the normalisation to the MNI coordinates (some examples are represented in Figures 4.3, 4.4, 4.5, 4.6), we can conclude that those processes were performed correctly.

Following the pre-processing and denoising pipeline, we proceeded to the computation of the FC matrices. In Figure 5.2 an example of a functional connectivity matrix from a Parkinson's disease patient is displayed. As expected, after the normalisation of the connectivity values using the Fisher transform, the values comprehended a range of -1 to 1. Moreover, the connectivity matrices are symmetric, as expected. From the CONN computation of the matrices, the values in the diagonal (corresponding to same ROI-ROI connectivity) were outputted with NaN values. During the data preparation, the diagonal values were set to 0, taking into account that those values provide no relevant information. After this

alteration, it was verified if any matrix had a NaN value, to ensure that the computation was successful. It was confirmed that none of the FC matrices had any NaN values. Therefore, the connectivity matrices were considered acceptable to be used in the following steps of the dissertation's work.

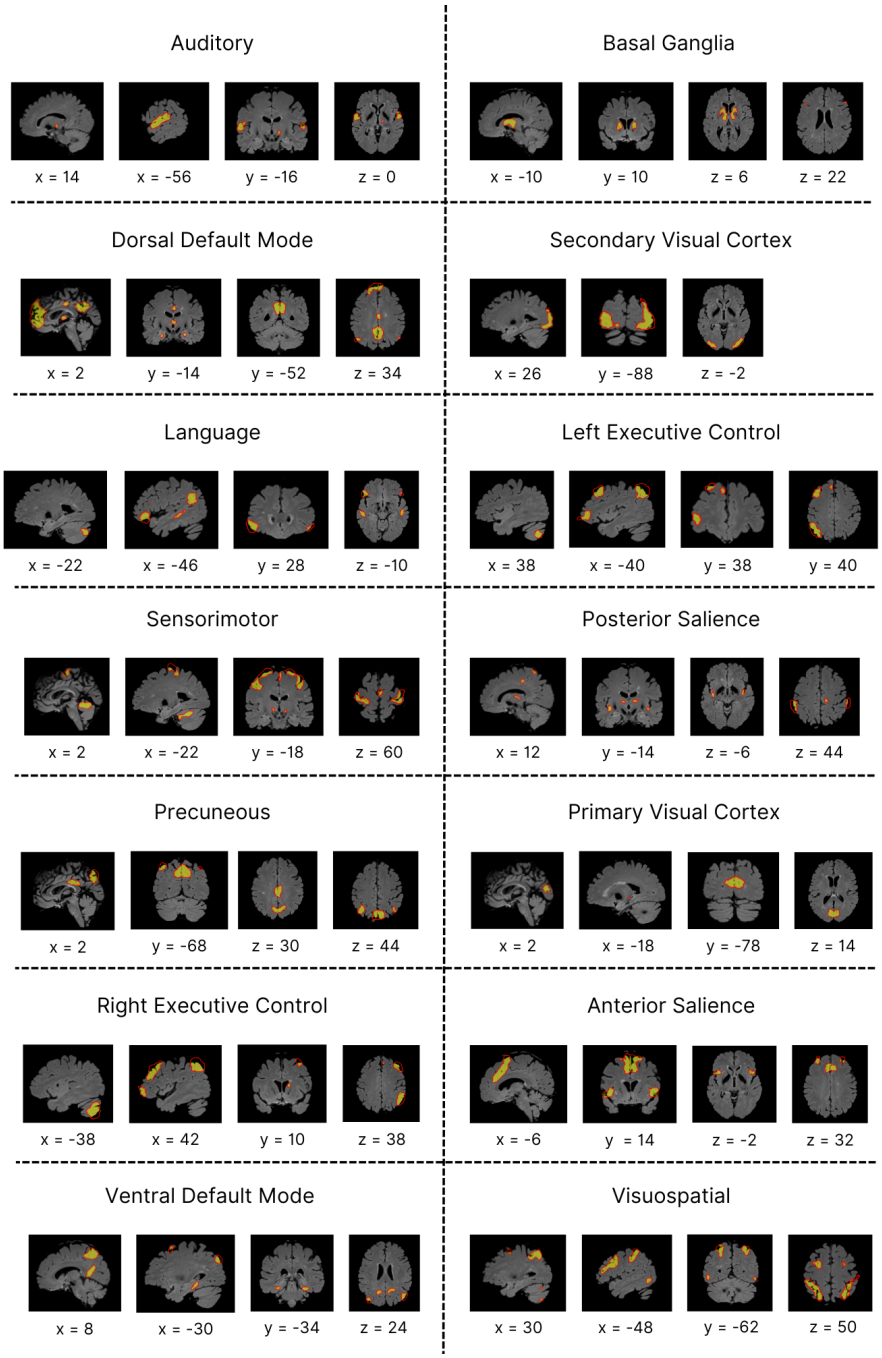

Figure 5.1: Results from the segmentation of the ROIs defined by Shirer et al. [83]. The 14 resting-state networks found by the authors are represented in sections of the anatomical data. The images were obtained from the CONN toolbox for fMRI pre-processing.

Considering that:

1. the segmentation of the resting-state networks defined by Shirer et al. [83] was correctly performed;

2. the quality assurance reports of the registration and normalisation steps showed that these steps of

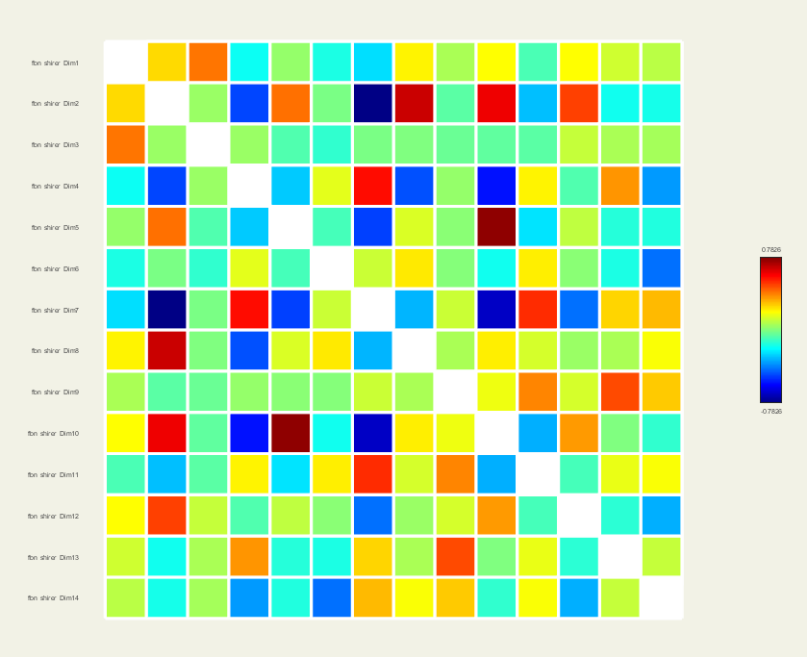

Figure 5.2: Example of ROI-to-ROI connectivity matrix of a Parkinson's disease patient from the PPMI data set. The FC matrix was computed using the CONN toolbox for fMRI pre-processing.

the pre-processing were successful;

3. the ROI-to-ROI functional connectivity matrices were considered as correctly computed;

we can affirm that the pre-processing of the fMRI scans from the PPMI and ADNI data sets was fairly acceptable, therefore, the connectivity matrices are sufficiently adequate to be applied to a deep learning classification model. Nonetheless, the pipeline of pre-processing fMRI data, which is very complex and time-consuming, due to the high dimensionality of the data, poor temporal resolution and low signal-to-noise ratio, is not immune to possible errors and inaccuracies that must not be disregarded.

5.2 Deep Learning Model Performance

This section presents the results related to the development of the deep learning classification model. We start by presenting the results from the model optimisation part, and then the performance of the final model.

5.2.1 Cross-validation Results

The first step in designing the classification model was to define the proper architecture for our problem. As described in the Methods chapter, we opted to inspire in the architecture of the ConnectomeCNN, designed by Meszlényi et al. [213], since it was developed precisely to classify functional connectivity matrices of patients *vs* healthy controls. To create a deeper architecture, in order to capture more features from the matrices, we decided to add one convolutional layer to our architecture and denominate it as ExtendedConnectomeCNN.

Then, we proceeded to the optimisation of the hyperparameters of our architecture. The hyperparameters chosen to optimise were the learning rate, the dropout rate and the presence or not of a batch normalisation layer on each convolutional layer. These parameters were adjusted in order to

achieve the most satisfactory performance on a 10-fold cross-validation. Eight different combinations of hyperparameters were tested, and they are defined in Table 4.2. The models were evaluated through the training and validation accuracy, the AUC ROC, sensitivity, specificity and F1-score.

The main characteristic that we want to be present in our model is generalisation. When applying explainable AI methods, it is crucial to have a model that generalises well. This ensures that the explanations derived from the XAI techniques are meaningful and can be applied to new and unseen data. It is known that overfitting models usually capture noise from the data [222]. Hence, despite having a small data set, we aimed to avoid a model that overfits. Furthermore, sensitivity (or true positive rate) and specificity (or true negative rate) should not be considerably distinct from one another, seeing that the model should capture both disease (or positive) and control (or negative) classes. Finally, we should not underestimate the AUC and F1-score metrics, which also should have a considerable value.

Hereupon, we present the results of the 10-fold cross-validation on the eight designed architectures, represented by the mean values of the considered performance metrics, in Table 5.1.

Table 5.1: Performance metrics on the classification task healthy controls (0) vs Parkinson's disease (1): 10-fold cross-validation. The best model is highlighted in bold.

Cross-Validation: mean values for 10 folds						
ID	Training Accuracy	Validation Accuracy	AUC	Sensitivity	Specificity	F1-Score
1	1.0000	0.7560	0.7181	0.7181	0.7857	0.7519
2	1.0000	0.7760	0.7372	0.7364	0.8071	0.7718
3	0.8991	0.7360	0.6860	0.6818	0.7786	0.7302
4	0.9814	0.7640	0.7183	0.7000	0.8143	0.7571
5	0.8977	0.7440	0.7136	0.7182	0.7643	0.7412
6	0.8845	0.7120	0.6819	0.7182	0.7071	0.7127
7	**0.8814**	**0.7760**	**0.7496**	**0.7636**	**0.7857**	**0.7747**
8	0.5968	0.5680	0.0910	0.0545	0.9714	0.5130

If we look at the training accuracy of combinations 1 and 2, with learning rates of 0.001 and 0.0001 respectively, and with batch normalisation layers and a dropout rate of 0.3 in all layers, both values are equal to 1.000, meaning that all training instances were correctly classified in all 10 folds. Hence, these models are clearly overfitting. Additionally, we notice a slight imbalance between sensitivity and specificity, which is not desirable.

Combination 3, with the same hyperparameters as combinations 1 and 2, except for the learning rate, which is 0.00001, does not overfit, with a training accuracy of 0.8991. However, we notice a greater imbalance between the sensitivity and specificity, which values are respectively 0.6818 and 0.7786, approximately. In particular, the models with this combination of hyperparameters are capturing the negative class (or the control class), significantly better than the positive class (or the PD class). This work aims to build a classification model in order to then interpret it and explain particularly how it is classifying Parkinson's disease. Hence, it is not proper to use a model that misses the classification of the positive class.

Considering that model 3 had a more significant imbalance between specificity and sensitivity, and that combination 2 had a better performance on the validation data than combination 1, because of higher accuracy, AUC and F1-score, we opted to increase the rate of dropout to 0.5, to avoid overfitting, and used the value of learning rate of the model 2, 0.0001. The batch normalisation layers were kept. Combination 4, corresponding to the models with the described hyperparameters, reached a 0.9814 mean training accuracy. They still overfit, but not as combinations 1 and 2. Moreover, the gap between sensitivity and

specificity is still considerable, hence, it still captures the control class better than the PD class.

For combination 5, compared to combination 4, the rate of dropout was increased to 0.6 in all layers, to check if we could avoid overfitting. The result of the training accuracy was approximately 0.8977, so the models with this combination were not overfitting. However, the models still present a denoting amplitude between sensitivity and specificity (0.7182 and 0.7643, respectively).

Afterwards, it was decided to improve combination 3, with a lower learning rate, by increasing the dropout rate only in the fully connected layer to 0.5. The dropout rate in the convolutional layers remained at 0.3. The training accuracy was 0.8845, so combination 6 was not overfitting, and the sensitivity and specificity did not present a high gap between them. Nevertheless, the value of the AUC ROC was low (approximately 0.6819). Therefore, these models are not generalising as expected.

To change the approach to the adjustment of the hyperparameters, we opted to eliminate the batch normalisation layers from the convolutional layers. The learning rate was set to 0.00001, the smallest used value, for combination 7, and the dropout rate was equal to 0.1 in convolutional layers, and 0.4 in the fully connected layer. The models with this combination do not overfit, with a training accuracy of 0.8814. Moreover, the values of sensitivity and specificity were relatively balanced, being approximately 0.7636 and 0.7857, respectively. Furthermore, the models presented satisfactory mean F1-score (0.7747), AUC ROC (0.7496) and validation accuracy (0.7760). When the dropout rates were set to 0.5 in all layers (combination 8), the models' mean performance drastically decreased, with training accuracy of 0.5968, validation accuracy of 0.5869, and AUC ROC of 0.0910.

The models with better performance on validation data are combinations 2 and 7. Nonetheless, since we want to apply XAI methods, we want a model that generalises, hence, combination 7 is preferable because it does not overfit. Lower training accuracy but similar test accuracy suggests that that model might be better at generalising to new data. It is less likely that combination 7 is overfitting to noise in the training data, as its performance on the validation set is comparable to that of combination 2.

Hereupon, the hyperparameters that were considered optimal were respective to the combination 7: learning rate of 0.00001, the dropout rate in convolutional layers equal to 0.1, and 0.4 in the fully connected layer, and no presence of batch normalisation layers.

To visually compare the models' combinations that were evaluated with 10-fold cross-validation, two scatter plots were constructed. Figure 5.3 displays the values of training accuracy, validation accuracy and AUC ROC in terms of the learning rate. Points in the grey area will correspond to values of training accuracy that did not overfit, and values of validation accuracy and AUC ROC that characterise generalisation. Combination 5 (cyan points) and combination 7 (blue points) have all three shapes (corresponding to the three performance metrics in analysis) in the grey region (between 0.7 and 0.9).

Figure 5.4 displays the values of the difference between sensitivity and specificity. Models' combinations with smaller differences will correspond to a balance between sensitivity and specificity, meaning that both classes (positive and negative) are being captured. Combination 6 (green point) and combination 7 (blue point) have the smallest values. It is important to note, however, that the values of AUC ROC, which measure how the model distinguishes positive and negative classes, do not necessarily vary according to the difference between sensitivity and specificity. Although combination 7 presents a considerably satisfactory AUC ROC value, 0.7496, the highest value of all combinations, combination 6 has one of the lowest AUC ROC values, with 0.6819.

The analysis of Figures 5.3 and 5.4 confirm the preference of the values of hyperparameters of combination 7. This combination was then applied to the design of the final model, which will be analysed through XAI methods. We will discuss the results referring to the final model performance in

the next subsection.

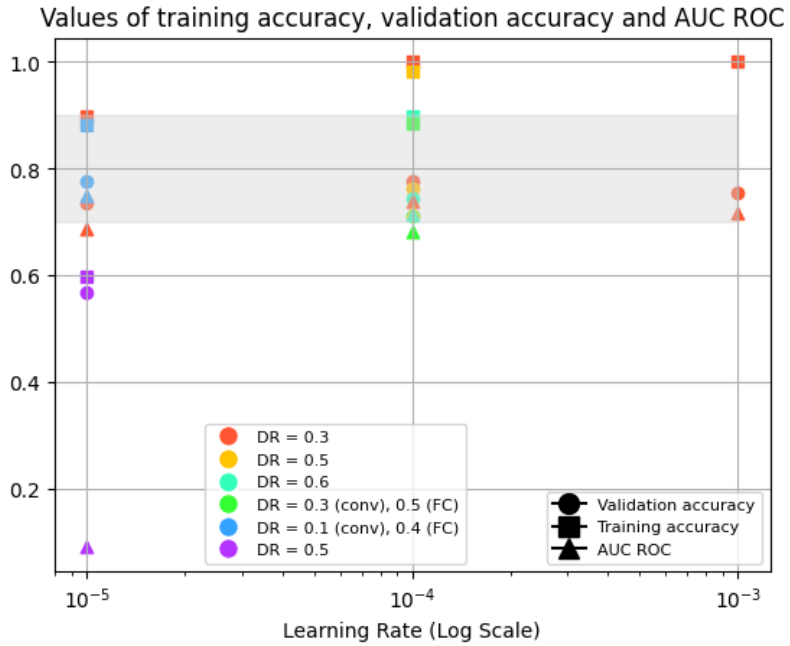

Figure 5.3: Scatter plot of the values of the training accuracy (square points), validation accuracy (circle points) and AUC ROC (triangular points) for each one of the model combinations. DR: dropout rate; conv: convolutional layers; FC: fully connected layer.

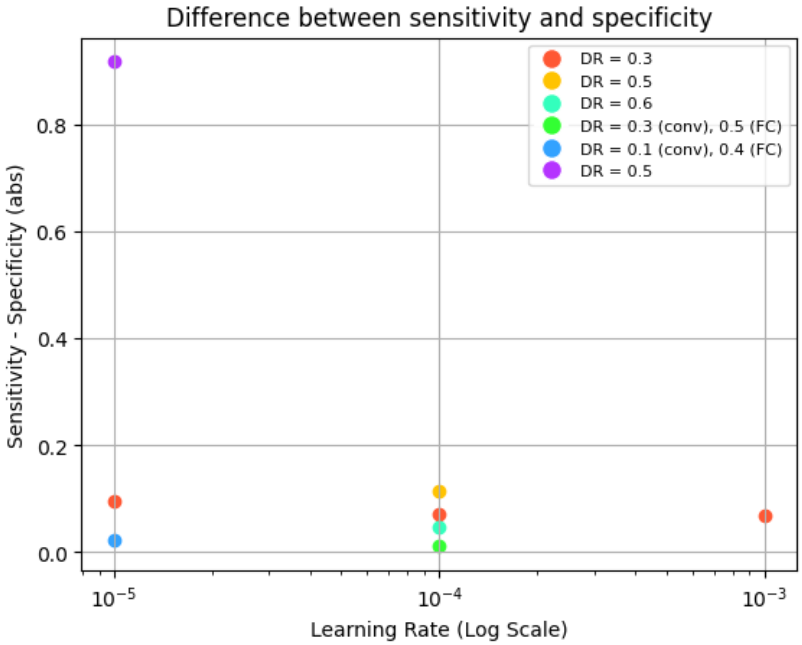

Figure 5.4: Scatter plot of the values of the difference between sensitivity and specificity for each one of the model combinations. DR: dropout rate; conv: convolutional layers; FC: fully connected layer.

5.2.2 Final Model Results

In this part of the work, a model trained with the whole development set was created, with the hyperparameters of combination 7, to be tested on the separate test set. It matters much to point out

that the ideal practice is to use an independent/external test set, for instance, from a different acquisition cohort, so that we assess if the deep learning model is generalising to unseen data, ensuring the model quality [223]. Unfortunately, the lack of availability of open and facilitated access to Parkinson's disease fMRI scans does not allow the use of an independent test set. Instead, as described in the Materials and Methods chapter, we opted to separate our PPMI + ADNI data set into development (245 subjects) and test sets (28).

Henceforth, the results of the performance of the final model (with combination 7 hyperparameters) are presented in Table 5.2. Similarly to the process of cross-validation, we chose as performance metrics training accuracy, test accuracy, AUC ROC, sensitivity, specificity, and F1-score. Additionally, the confusion matrix for the test set is presented in Figure 5.5.

Table 5.2: Performance metrics on the classification task healthy controls (0) vs Parkinson's disease (1): final model on the test set.

Final Model Performance on Test Set					
Training Accuracy	Test Accuracy	AUC	Sensitivity	Specificity	F1-Score
0.8776	0.8214	0.8230	0.8125	0.8333	0.8000

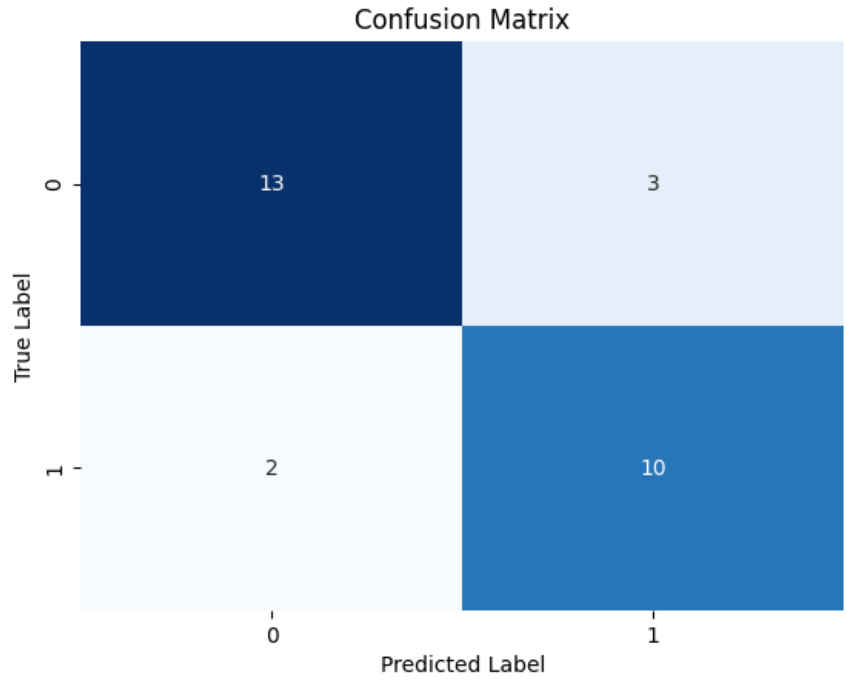

Figure 5.5: Confusion matrix for the classification on the test set with the final model; 0: healthy control, 1: Parkinson's disease.

The values of the training and test accuracies are considered to be close to one another, relatively robust and satisfactory, being 0.8776 and 0.8214 respectively. This result suggests that the model, despite apparently underfitting, is probably suitable to generalise to unseen data. The accuracy values indicate the ratio of correctly predicted instances. It is one of the most common metrics evaluated and analysed in machine learning problems, however, it is not enough to assess a model's quality by itself. Therefore, we considered other metrics in our results analysis.

The AUC ROC measures how the model distinguishes the positive and the negative class. In the case of our model, the AUC ROC is 0.8230, which is a reasonably acceptable result. This can be confirmed when analysing the values of sensitivity and specificity, 0.8125 and 0.8333. These values are close to

each other, which means that both positive and negative classes are being considerably captured.

On the other hand, the F1-score is a harmonic mean of the precision and sensitivity (or recall), as previously explained. An F1-score of 0.8000 suggests that the model is fairly balanced in terms of precision and recall.

Looking into the respective confusion matrix, it is possible to obtain a visual perspective of our results, in particular regarding the distribution of true positives, false positives, false negatives and true negatives. Although the distribution of true instances (true positives and true negatives) is fairly positive, we cannot undervalue the existent false instances. It is necessary to take into account that the model still misses some PD patients classification, which is not ideal since it is our purpose for the model to capture and diagnose the disease, not only in terms of the focus of the classification task but also due to the objective of explaining the respective results and understanding how the model classifies true positive instances. Nonetheless, we keep in mind that, at the moment, considering the limitations and challenges of this work, which will be explored next, it is not realistic to build a perfectly accurate model.

Overall, our model presents balanced, fairly acceptable and satisfactory results. Notwithstanding the fact that the value of training accuracy is not as large as expected in the context of AI and its application to medical imaging analysis, so one may consider that our model is underfitting, it is crucial to remember that the focus of our work is to apply, compare and analyse XAI methods in order to assess how the CNN model is classifying Parkinson's disease. Henceforth, having a model with lower training accuracy, but with similar test accuracy, is preferable to a model with a higher training model (approximately equal to 1), since it is less probable that the model with lower training accuracy is overfitting, capturing noise from the data that it is not meant to be captured by the XAI analysis.

This discussion regarding the final model results is according to the cross-validation results, where we optimise the hyperparameters of the developed model architecture. Therefore, we can consider that the final model achieved the expected results, fulfilling the sought requirements.

However, we need to take into account the small size of our data set. Our test set represents 10% of our data set, meaning that it is composed of 28 instances, 16 healthy controls and 12 PD patients. The fact is that general assumptions regarding the final model performance are being made based on a considerably small data set, which is a significant limitation that cannot be disregarded. In the next subsection this and other challenges encountered during the development and evaluation of the CNN model for the classification of Parkinson's disease with functional connectivity matrices.

5.2.3 General Considerations and Limitations

In both cross-validation results and final model results, in the general scope, the models meet our expectations: all the performance metrics were considered balanced and fairly acceptable. However, in this work, there are some limitations and challenges that need to be taken into account.

Firstly, as mentioned above, the small size of the used data set is a limitation to consider in this work. Despite trying to augment and balance it by importing data of healthy controls from the ADNI data set, the size of the set is still limited, considering that solving problems with deep learning involves a great amount of data. This is a problem that was identified in Chapter 3 while reviewing the use of artificial for Parkinson's disease diagnosis in the literature. Notwithstanding the important results that were obtained from the presented studies, presumptions on the diagnosis of PD based on small data sets were presented. On the other hand, in our work, we opted for using healthy controls from a separate data set, in order to use a larger set than the previous works. Nonetheless, the number of samples should be improved. In light of this limitation, we point out that it is mandatory to improve the available data sets in the context

of Parkinson's disease and acquire more data, so that is possible to create studies and research works with more significant and reliable generalisations. Another solution for the problem of lack of data would be the application of transfer learning from a model trained on similar data. This approach was attempted and will be discussed later.

In addition, although we did not detect any evidence of this matter in our data, we cannot disregard the slight possibility of the existence of a bias related to the inclusion of ADNI healthy controls in our data set. We still considered the best solution to use them, because the correspondent fMRI acquisition parameters were similar to the used parameters for PPMI acquisition, as described in Chapter 4, and considering the need for a larger and more balanced set. Nevertheless, it is important to note that an even better approach could be applied if the data sets from PPMI, in particular fMRI scans for healthy controls, were augmented. However, we cannot disregard the fact that the acquisition of fMRI scans extensively is a costly process that takes a long time, compared with other techniques for functional connectivity assessment, like EEG [224], so it may not be an easy task to augment the PPMI data set.

Finally, it is needed to account for the complexity of the functional MRI data. fMRI data has high dimensionality (time and space) and has several limitations. Spatial resolution is limited mainly by the signal-to-noise ratio since it is needed to rapidly acquire the time-series information [57]. Additionally, the time resolution is also low due to the hemodynamic response time (as explained in subsection 2.2.2). A signal dropout or spatial distortion in frontal orbital and lateral parietal regions, caused by the $\tilde{9}$ ppm difference in magnetic susceptibility at interfaces between air and brain tissue, is also verified in fMRI scans [57]. These are some of the limitations present in the acquisition of the fMRI data, which increase their complexity, therefore, the respective pre-processing will not be facilitated. Henceforth, complications regarding the pre-processing pipeline may affect the computation of the connectivity matrices, and then the performance of the model. There is always room to evaluate and improve the used pre-processing steps. In addition, a suggestion to facilitate the handling of the fMRI scans is the availability of pre-processed data in the PPMI, guaranteeing that the process would be the most suitable for the respective data.

Notwithstanding the presented challenges and limitations on the design and evaluation of the model for the classification of Parkinson's disease, we consider that our approach is an innovative method regarding the study of brain functional connectivity in Parkinson's disease and in the application of deep learning models. With the cross-validation method for hyperparameters optimisation, we ensured that we obtained an architecture that was generalisable given our data set. This was confirmed when the model was trained in the complete development set and tested on the external test set, as our described findings show. Hence, our model is considered acceptable for the application of XAI methods, by fulfilling the following requirements:

1. Satisfactory performance metrics: test accuracy, AUC ROC, sensitivity, specificity and F1-score are all at reasonable values. The results suggest that the model is making meaningful and relevant predictions, while not performing poorly.

2. Balanced metrics: both sensitivity and specificity display a balanced performance in identifying positive (PD) and negative (control) classes. This equilibrium in performance often makes the interpretation and explanation of the model's classification more straightforward.

3. Interpretability potential: despite using a deep neural network, our CNN do not include a large number of layers, meaning that our model's complexity is moderate. Since the model is not overly complex, it is more likely that meaningful explanations can be extracted.

5.3 Model Explanability

After obtaining the final model results, we proceeded to the respective explainable AI analysis. In our analysis, we focused on the explanation and interpretation of the classification of the positive class, and we considered only the true positive cases since our aim is to study how the model classifies the Parkinson's disease class. The application of XAI methods allows the understanding of how the model is making decisions, which is particularly important when the model's predictions influence critical decisions, and enhance the transparency of the model's decision-making. In addition, by assessing the model's explainability, it is possible to verify which resting-state networks correspond to the most relevant feature for PD, hence, which networks can be considered as diagnostic biomarkers of Parkinson's disease.

Hereupon, in this subsection we present the results obtained from the XAI analysis of our CNN model for the classification of Parkinson's disease *vs* healthy controls. Individual heatmaps corresponding to each one of the applied methods, and for each considered subject, were obtained. In Figures 5.6, 5.8 and 5.10, there are examples of heatmaps correspondent to the LRP DeconvNet and guided backpropagation methods, respectively, of one PD patient that was correctly classified as a true instance (true positive). Furthermore, the mean heatmaps of all true positive instances for each applied XAI method were computed. These heatmaps are represented in Figures 5.7 (LRP), 5.9 (DeconvNet) and 5.11 (guided backpropagation). From each mean heatmap, the top 10 most explainable correlation pairs of features were extracted. The values for each one of the methods LRP, DeconvNet and guided backpropagation, are represented in Tables 5.3, 5.4, and 5.5, respectively. Analysing the values from the tables, it is noticed that network number 2, corresponding to the dorsal default mode network, is the most recurrent feature to appear as the most relevant. This finding is common to all the applied XAI methods.

For LRP analysis, the pair dorsal default mode network and ventral default mode network has the highest value of explainability, with approximately 0.3014. For DeconvNet analysis, the most relevant pair was the precuneus network and dorsal default mode network, with a value of explainability of approximately 0.3461. The pair ventral mode network and precuneus was the most relevant in guided backpropagation analysis, with an explainability value of about 0.4924. In terms of evaluation of the explainability values between the different XAI methods, it is not possible to compare them, since for each XAI method there is not a range of values. The obtained values are purely relative to each other, in between the XAI analysis, and are useful for relative visual comparison purposes, and not for absolute quantification.

Note that Figures 5.6, 5.8 and 5.10 correspond to the same subject, hence, they are comparable. Visually comparing the local heatmaps for each XAI method, it is possible to infer that the LRP method seems to present a less scattered, and more localised distribution of the values, meaning that it is clear which are the most relevant features. On the other hand, DeconvNet and guided backpropagation methods present more sparse heatmaps: the values seem to concentrate on relatively relevant values, and the most relevant features are not distinct.

Regarding the mean heatmaps for each method, the same tendency is verified. LRP presents more localised values of explainability than DeconvNet and guided backpropagation methods. Considering this tendency found in the local matrices and in the mean matrices, there seems to be evidence that the explanations provided by the LRP method are more specific and less random. However, we cannot ensure this hypothesis by solely relying on the visual distribution of the explainability values in the heatmaps.

Another curious finding in the heatmaps is the fact that the explanation values are not completely symmetrically distributed in the matrices, for both local and mean heatmaps. It actually would be

expected to obtain symmetric explanations, considering the symmetry of the FC matrices and the context of this study. The resulting non-symmetric XAI heatmaps could be evidence that the symmetric information of the FC matrices is redundant. Nevertheless, it is still believed that eliminating the redundant information from the matrices could affect the convolution operation during feature extracting through the CNN, which would be disadvantageous.

To complete the visual analysis of the XAI heatmaps, we performed the computation of the AOPC values for each one XAI methods. This analysis will allow us to verify which one of the methods delivered better explanations of the classification of the positive (PD) class. Figure 5.12 represents the plot of the AOPC for each XAI method.

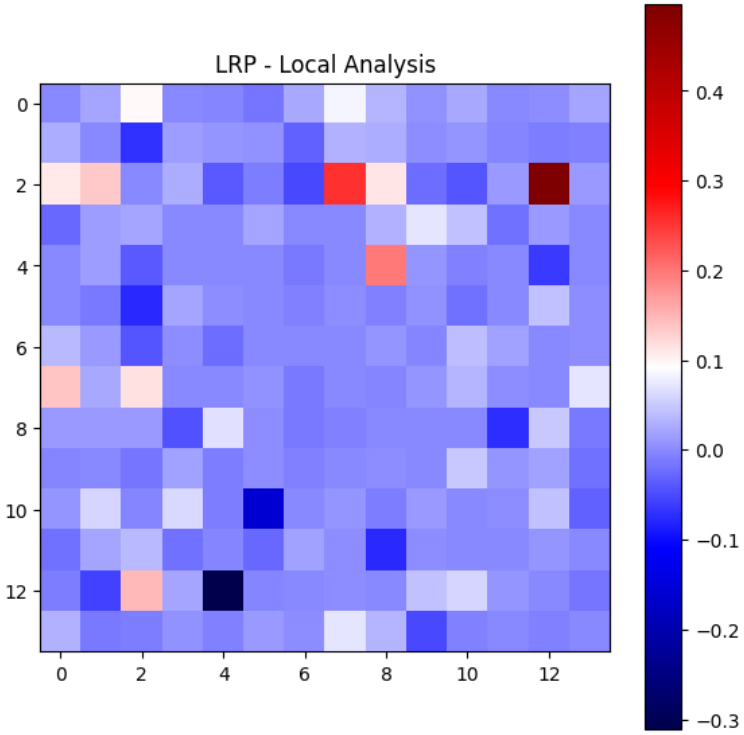

Figure 5.6: Example of a heatmap from XAI analysis with layer-wise relevance propagation method, from one subject. 0: auditory; 1: basal ganglia; 2: dorsal default mode; 3: higher visual cortex; 4: language; 5: left executive control; 6: sensorimotor; 7: posterior salience; 8: precuneus; 9: medial visual cortex; 10: right executive control; 11: anterior salience; 12: ventral default mode; 13: visualspatial.

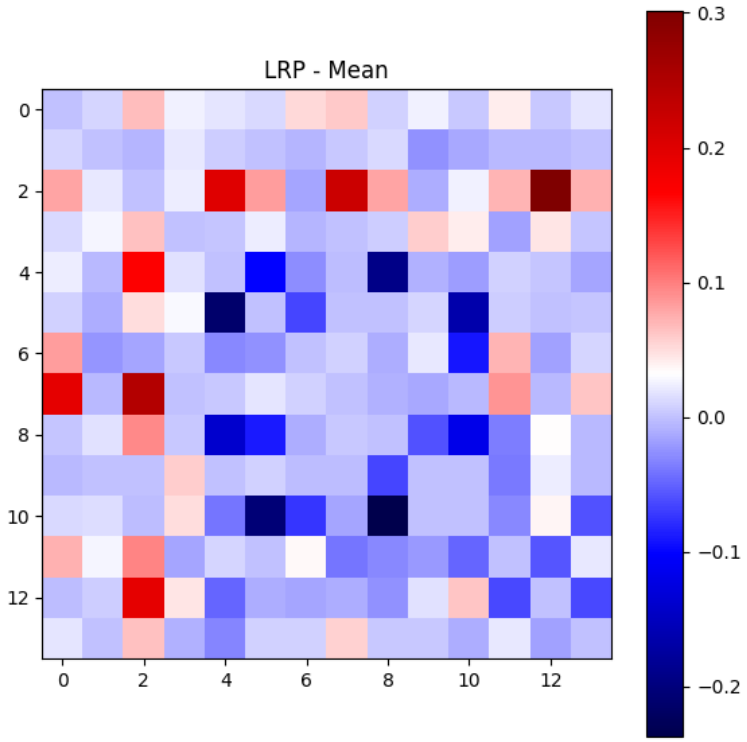

Figure 5.7: Mean heatmap from XAI analysis with layer-wise relevance propagation method. 0: auditory; 1: basal ganglia; 2: dorsal default mode; 3: higher visual cortex; 4: language; 5: left executive control; 6: sensorimotor; 7: posterior salience; 8: precuneus; 9: medial visual cortex; 10: right executive control; 11: anterior salience; 12: ventral default mode; 13: visualspatial.

Table 5.3: Top 10 correlation pairs of resting-state networks, accordingly the explainability values from layer-wise relevance propagation method.

	Layer-wise Relevance Propagation		
Top	Coordinates	Networks	Explainability Value
1	(2, 12)	Dorsal Default Mode, Ventral Default Mode	0.3014
2	(7, 2)	Posterior Salience, Dorsal Default Mode	0.2463
3	(2, 7)	Dorsal Default Mode, Posterior Salience	0.2206
4	(2, 4)	Dorsal Default Mode, Language	0.1994
5	(7, 0)	Posterior Salience, Auditory	0.1934
6	(12, 2)	Ventral Default Mode, Dorsal Default Mode	0.1928
7	(4, 2)	Language, Dorsal Default Mode	0.1673
8	(11, 2)	Anterior Salience, Dorsal Default Mode	0.0956
9	(8, 2)	Precuneus, Dorsal Default Mode	0.0946
10	(7, 11)	Posterior Salience, Anterior Salience	0.0886

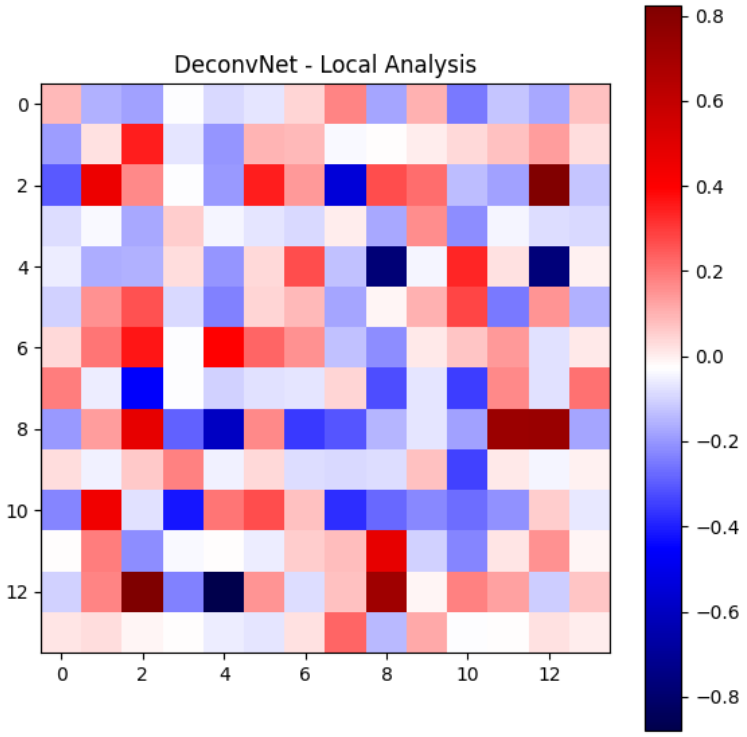

Figure 5.8: Example of a heatmap from XAI analysis with deconvolutional network method, from one subject. 0: auditory; 1: basal ganglia; 2: dorsal default mode; 3: higher visual cortex; 4: language; 5: left executive control; 6: sensorimotor; 7: posterior salience; 8: precuneus; 9: medial visual cortex; 10: right executive control; 11: anterior salience; 12: ventral default mode; 13: visualspatial.

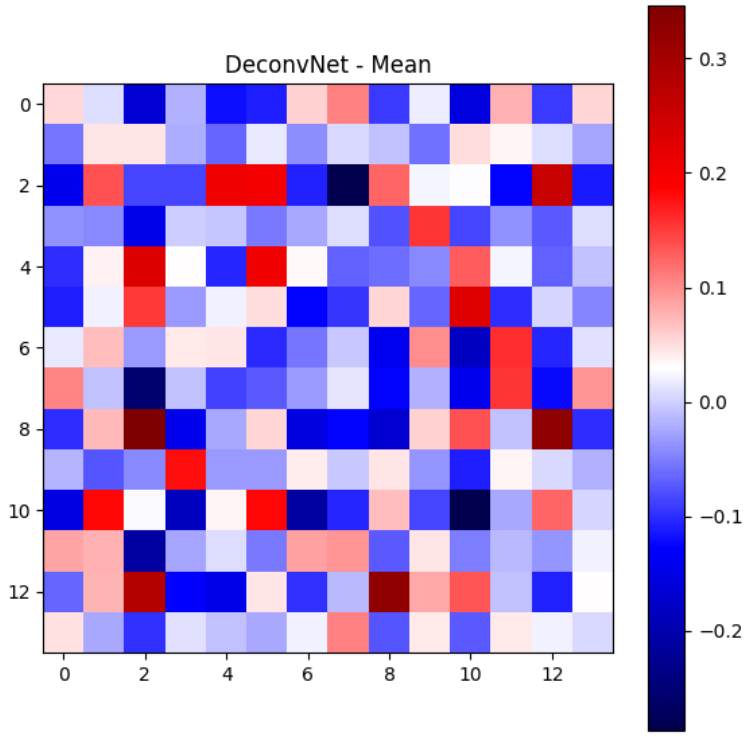

Figure 5.9: Mean heatmap from XAI analysis with deconvolutional network method. 0: auditory; 1: basal ganglia; 2: dorsal default mode; 3: higher visual cortex; 4: language; 5: left executive control; 6: sensorimotor; 7: posterior salience; 8: precuneus; 9: medial visual cortex; 10: right executive control; 11: anterior salience; 12: ventral default mode; 13: visualspatial.

Table 5.4: Top 10 correlation pairs of resting-state networks, accordingly the explainability values from deconvolutional network method.

	Deconvolution Network		
Top	Coordinates	Networks	Explainability Value
1	(8, 2)	Precuneus, Dorsal Default Mode	0.3461
2	(8, 12)	Precuneus, Ventral Default Mode	0.3258
3	(12, 8)	Ventral Default Mode, Precuneus	0.3236
4	(12, 2)	Ventral Default Mode, Dorsal Default Mode	0.2781
5	(2, 12)	Dorsal Default Mode, Ventral Default Mode	0.2547
6	(4, 2)	Language, Dorsal Default Mode	0.2297
7	(5, 10)	Left Executive Control, Right Executive Control	0.22531745
8	(2, 4)	Dorsal Default Mode, Language	0.2042
9	(4, 5)	Language, Left Executive Control	0.2039
10	(2, 5)	Dorsal Default Mode, Left Executive Control	0.1989

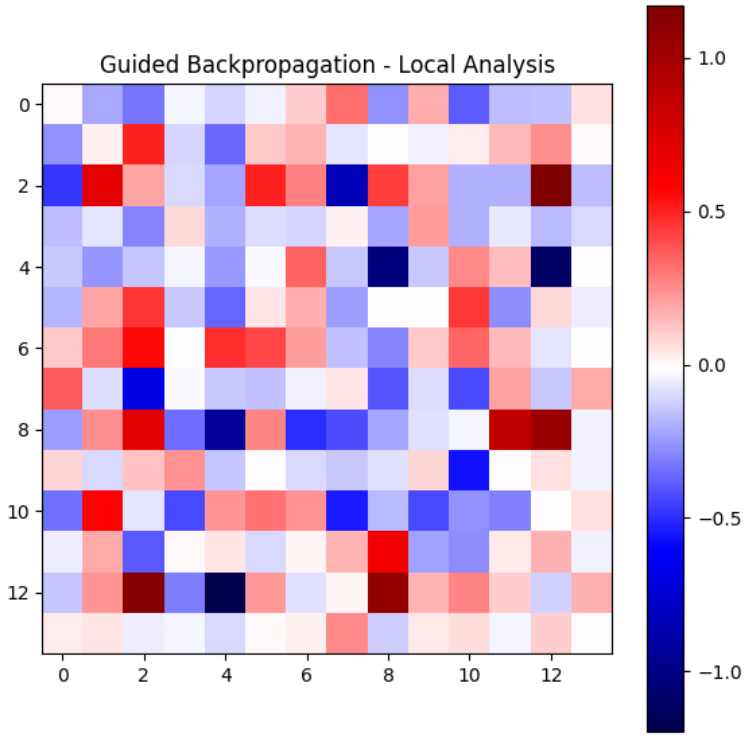

Figure 5.10: Example of a heatmap from XAI analysis with guided backpropagation method, from one subject. 0: auditory; 1: basal ganglia; 2: dorsal default mode; 3: higher visual cortex; 4: language; 5: left executive control; 6: sensorimotor; 7: posterior salience; 8: precuneus; 9: medial visual cortex; 10: right executive control; 11: anterior salience; 12: ventral default mode; 13: visualspatial.

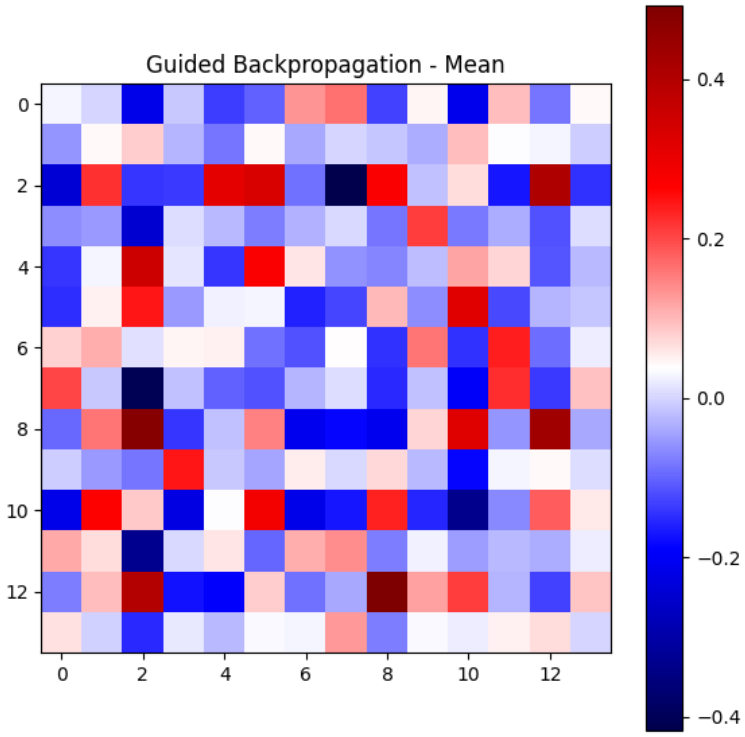

Figure 5.11: Mean heatmap from XAI analysis with guided backpropagation method. 0: auditory; 1: basal ganglia; 2: dorsal default mode; 3: higher visual cortex; 4: language; 5: left executive control; 6: sensorimotor; 7: posterior salience; 8: precuneus; 9: medial visual cortex; 10: right executive control; 11: anterior salience; 12: ventral default mode; 13: visualspatial.

Table 5.5: Top 10 correlation pairs of resting-state networks, accordingly the explainability values from guided backpropagation method.

	Guided Backpropagation		
Top	Coordinates	Networks	Explainability Value
1	(12, 8)	Ventral Default Mode, Precuneus	0.4924
2	(8, 2)	Precuneus, Dorsal Default Mode	0.4761
3	(8, 12)	Precuneus, Ventral Default Mode	0.4313
4	(2, 12)	Dorsal Default Mode, Ventral Default Mode	0.4061
5	(12, 2)	Ventral Default Mode, Dorsal Default Mode	0.3997
6	(4, 2)	Language, Dorsal Default Mode	0.3521
7	(2, 5)	Dorsal Default Mode, Left Executive Control	0.3304
8	(8, 10)	Precuneus, Right Executive Control	0.3198
9	(5, 10)	Left Executive Control, Right Executive Control	0.3153
10	(2, 4)	Dorsal Default Mode, Language	0.3083

Analysing the plot from Figure 5.12, it is noticed that the curve correspondent to the LRP method abruptly increases after around 20 iterations, and then stabilises to a maximum value of approximately 0.4549. This reflects that, in the first 20 iterations, relevant features are being removed from the matrices and then affecting the classification of the model: the increase of the AOPC value is increasing at each iteration, meaning that the 20 MoRF features significantly better explain the classification of Parkinson's disease. After the first 20 iterations, the value of the AOPC is constant to a maximum, which suggests that the remaining MoRF features are not so significant for the classification.

On the other hand, the curves corresponding to the DeconvNet and guided backpropagation methods do not present a crescent tendency, by starting to decrease slightly to negative values until after the first

20 iterations and then stabilising. In addition, these curves are much noisier than the LRP curve. These results indicate that the pairs of resting-state networks that are considered as most relevant do not seem to significantly explain the classification of Parkinson's disease.

Overall, it is possible to infer that the heatmaps computed by LRP have the largest AOPC values than the heatmaps computed by DeconvNet and guided backpropagation. Hence, the LRP method better identifies, by far, the relevant resting-states networks in the FC matrices. The performance of the DeconvNet and guided backpropagation is considered much poorer. It is acknowledged that the explanations produced by these methods are not significant, since they appear to not affect the classification of PD as expected when they are perturbated. As the explanations from the DeconvNet and guided backpropagation do not seem relevant, it makes sense to disregard the respective results in the assessment of the changes in FC related to Parkinson's disease, and solely consider the LRP analysis.

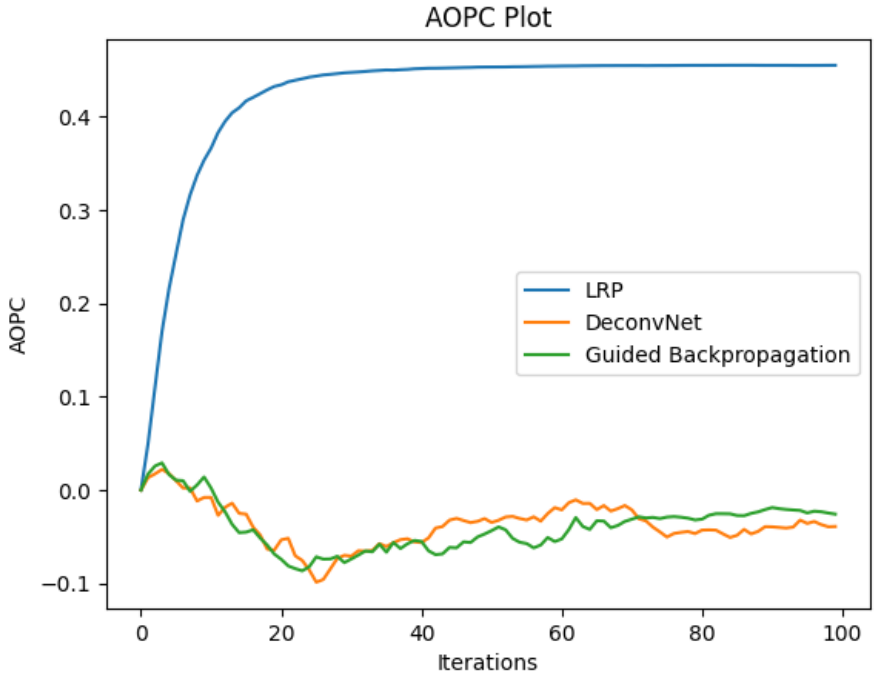

Figure 5.12: Area over the MoRF curve (AOPC) plot for each one of the applied XAI methods.

A significant limitation of the XAI analysis of the CNN model for PD classification is, once more, the size of the data set. In particular, in this section, we strict our analysis to the true positive instances, which are 10. Again, generalisations are based on a small data set. Nevertheless, these results are still significant, since new ground is being broken in the study of the functional connectivity of Parkinson's disease. Additionally, it is crucial to focus our analysis on the true positive classifications, since our aim is to unravel how the model is discriminating the PD class.

Furthermore, it is important to note that other ways to perturb the connectivity matrices, in order to compute the AOPC metric, should be taken into account. For example, noise could be introduced in the MoRF features by applying Gaussian-distributed or uniform-distributed noise to the matrix elements. Perturbations could also be added by simply adding or multiplying a constant in the respective values, to disrupt the connectivity in that element.

Despite the pointed limitations, the XAI analysis proved to surpass the black-box nature of our classification model. Relevant and significant explanations were produced by the LRP model, which results allow us to compare the results with the pathophysiologic findings regarding Parkinson's disease.

We will then proceed to the identification of potential Parkinson's disease biomarkers related to functional connectivity alterations.

5.3.1 Parkinson's Disease Biomarkers

In this subsection, the results from the XAI analysis will be interpreted, in the scope of findings regarding brain functional connectivity alterations related to Parkinson's disease. As previously mentioned, in this study we will only consider the results from the LRP analysis since both DeconvNet and guided backpropagation methods did not produce relevant explanations, according to the AOPC analysis.

Firstly, it is crucial to point out that the basal ganglia network was not identified as a significant feature for Parkinson's disease classification. This network was not found in the top 10 explanations produced by the LRP method, and analysing the mean heatmap corresponding to this method, the line and column corresponding to the network 1 (which is the basal ganglia) present a range of colours that match the explainability values around 0. This was not an expected result, considering the role of the basal ganglia system in the pathophysiology of Parkinson's disease (refer to subsection 2.1.1 for more information), and taking into account the findings from the study performed by Szewczyk-Krolikowski et al., where reduced functional connectivity with the basal ganglia network in PD patients was found (this study is further explained in section 3.2) [171].

However, the LRP method still identified resting-state networks considered relevant in the scope of Parkinson's disease. For this discussion, we will consider the features that are associated with an explainability value greater than 0.2 as significant for Parkinson's disease classification. Following this criterion, the significant pairs of networks are: (2,12), with an explainability value of approximately 0.3014, (7,2), with an explainability value of approximately 0.2463, and (2,7), with an explainability value of approximately 0.2205. Note that the last two pairs correspond to the same connectivity. Hereupon, we have three resting-state networks of interest: dorsal default mode network, ventral default mode network, and posterior salience.

To verify if the values of functional connectivity actually differ between the groups of Parkinson's disease patients and healthy controls, the corresponding boxplots of the connectivity values of the pairs dorsal default mode, ventral default mode networks, and posterior salience, dorsal default mode networks were computed. Then, the objective is to perform a one-sided Mann-Whitney U statistic test to confirm if the median of connectivity values of one group is larger or smaller than the other. Figures 5.13 and 5.14 present the boxplots correspondent to the comparison of the connectivity values of dorsal default mode and ventral default mode networks, and the comparison of the connectivity values of posterior salience and dorsal default mode networks, respectively.

When analysing the boxplot from Figure 5.13, it seems that the values from connectivity between dorsal default mode and ventral default mode networks of the PD group are distributed in larger values than the values from the healthy controls groups: the median of PD is approximately 0.3579, and the median of the controls is 0.2252. On the other hand, the values of connectivity between posterior salience and dorsal default mode networks seem greater in the healthy control group: the median of PD is 0.4225, and the median of the controls is -0.2197. To confirm these assumptions, the one-sided Mann-Whitney U statistic tests were performed. In Table 5.6 the hypothesis H_0 and H_1 for each one of the statistical tests are defined, as well as the significance level $\alpha = 0.05$. The respective results are displayed in Table 5.7.

The test on the connectivity values of dorsal default mode and ventral default mode networks had as

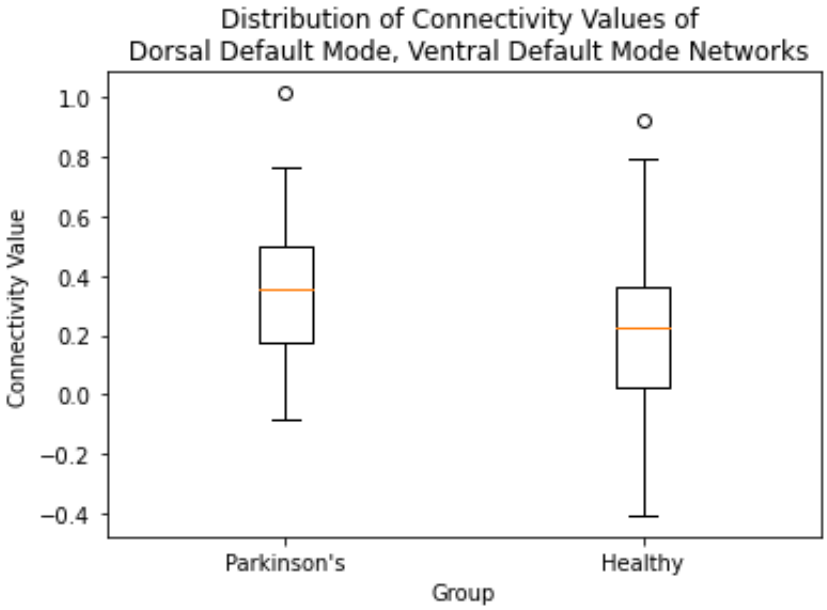

Figure 5.13: Comparison of the connectivity values of dorsal default mode and ventral default mode networks, between the Parkinson's disease patients and healthy controls groups, described in a boxplot.

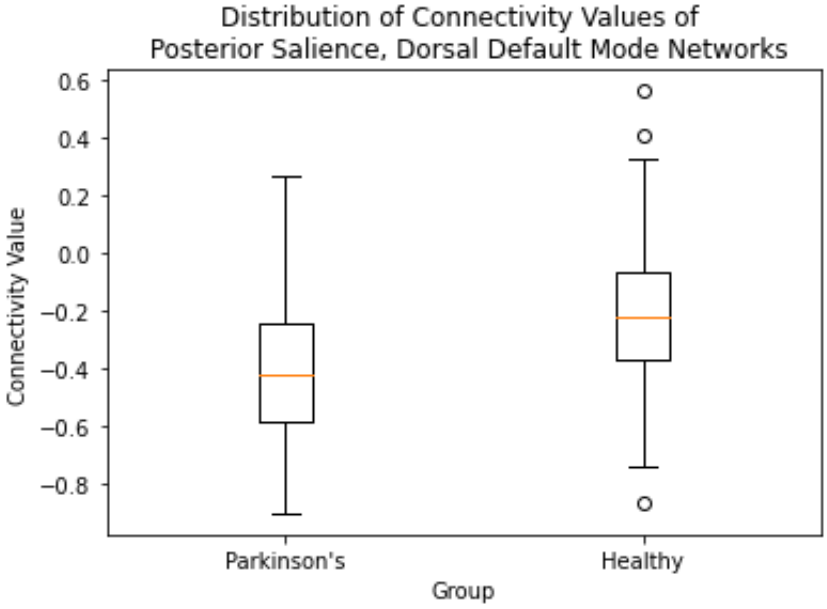

Figure 5.14: Comparison of the connectivity values of posterior salience and dorsal default mode, between the Parkinson's disease patients and healthy controls groups, described in a boxplot.

Table 5.6: Statistical hypothesis for the one-sided Mann-Whitney test for the median. The tests were performed for the connectivity values of the pair dorsal default mode, ventral default mode networks, and the pair posterior salience and dorsal default mode networks. The significance level was set to $\alpha = 0.05$ for both statistical tests.

Connectivity Pair	H_0	H_1	α
Dorsal Default Mode, Ventral Default Mode	"PD $\leq$ controls"	"PD > controls"	0.05
Posterior Salience, Dorsal Default Mode	"PD $\geq$ controls"	"PD < controls"	

the H_0 hypothesis "the values of connectivity in the PD group are smaller or equal to the values in the control group", and the H_1 hypothesis being "the values of connectivity in the PD group are greater than

Table 5.7: Results for the one-sided Mann-Whitney test for the median.

Connectivity Pair	α	Mann-Whitney U statistic	P-value	Result
Dorsal Default Mode, Ventral Default Mode	0.05	12195.0	6.2120e-08	Reject H_0
Posterior Salience, Dorsal Default Mode		5675.0	3.1060e-08	Reject H_0

the values in the control group". On the other hand, the H_0 hypothesis for the test on the connectivity values of posterior salience and dorsal default mode networks is "the values of connectivity in the PD group are greater than or equal to the values in the control group", and the H_1 "the values of connectivity in the PD group are smaller than the values in the control group". In the first test, the resultant p-value was approximately 6.2120e-08, and for the second 3.1060e-08. Considering the significance value of $\alpha = 0.05$, for both tests H_0 was rejected, meaning that:

- the connectivity values of dorsal default mode and ventral default mode networks of the PD group are significantly greater than the control group, which may indicate that this connectivity path is compensating for some defect caused by Parkinson's disease;
- the connectivity values of posterior salience and dorsal default mode networks of the PD group are significantly smaller than the control group, which may indicate that this connectivity path is negatively affected by Parkinson's disease.

To further analyse these findings, we proceed to study how dorsal default mode, ventral default mode, and posterior salience networks are related to Parkinson's disease, by assessing in which processes are those networks involved.

5.3.1.1 Involvement of Dorsal Default Mode Network in Parkinson's Disease

The dorsal default mode network, as previously mentioned, is thought to be involved in the evaluation of imagined future events [86]. Important brain regions that are part of this network are the posterior cingulate cortex and the medial prefrontal cortex.

The posterior cingulate cortex is linked to other posteromedial cortices, as well as paralimbic and limbic systems. These connectivity patterns represent a dorsal-ventral differentiation within the posterior cingulate cortex, with stronger connections of the ventral posterior cingulate cortex regions with structures in the medial temporal lobe, such as the hippocampus, and with structures in the ventromedial prefrontal cortex, such as the ventral anterior cingulate [225]. A study performed by Vogt showed evidence that both anterior and posterior cingulate cortex regions showed reduced thickness in PD patients [226]. This study also points out the potential of the whole cingulate cortex as a biomarker for the diagnosis of the disease.

On the other hand, the medial prefrontal cortex is involved in memory and decision-making mechanisms, affecting tasks like learning associations between context, locations, events, and corresponding adaptive responses, particularly emotional responses [227]. Dysfunction of the medial prefrontal cortex has been found in several disorders, such as depression, anxiety disorders, schizophrenia, autism spectrum disorders, Alzheimer's disease, Parkinson's disease, and addiction. In the case of PD, medial prefrontal cortex function is thought to influence the firing rate of respective neurons, leading to disease-related depression [228].

As noticed, the previous analysis shows evidence that Parkinson's disease affects regions of the dorsal default mode network, which may influence the functional connectivity of this network with others.

These regions, the posterior cingulate cortex and the medial prefrontal cortex affect mostly non-motor manifestations of Parkinson's disease, such as depression. Additionally, the progression of PD seems to lead to nerve injuries in the limbic system [229], which is associated with the posterior cingulate cortex.

5.3.1.2 Involvement of Ventral Default Mode Network in Parkinson's Disease

The ventral default mode network, on the other hand, is believed to be involved in the construction of imagined future events [86]. In this network, the retrosplenial cortex and the medial temporal lobe are relevant brain regions to analyse.

The retrosplenial cortex is a key region of a core network of brain regions that are involved in a range of cognitive functions, including episodic memory, navigation, imagination and planning for the future [230]. A study showed the administration of dopaminergic medication in patients with PD resulted in "decreased retrosplenial cortex BOLD response to emotion relative to off-medication state" [231]. In this experiment, it is shown that upon the alterations in the retrosplenial cortex, dopaminergic medication resulted in restored behavioural accuracy for emotional faces [231]. This could be evidence that a higher BOLD activity in this region of the ventral default mode network in non-medicated PD patients may have induced the higher connectivity between ventral default mode and dorsal default mode networks.

The medial temporal lobe includes the following brain structures: the hippocampus, amygdala and parahippocampal regions. It is involved in processes such as episodic and spatial memory [232]. In the scope of Parkinson's disease, a study shows that medial temporal lobe atrophy was verified in cognitively intact patients, suggesting an early hippocampal involvement in PD [233].

Once more, the literature suggests there are alterations in brain regions that are present in the ventral default mode network related to Parkinson's disease non-motor manifestations.

5.3.1.3 Involvement of Posterior Salience Network in Parkinson's Disease

The posterior salience network includes the left and right thalamus and putamen, which are part of the basal ganglia system. It is then assumed that the damage associated with that brain region may affect the functional connectivity of the posterior salience network. Furthermore, the study performed by Berman et al., presented in the state-of-the-art chapter of this dissertation, showed alterations in the salience network related to the non-administration of medication on PD patients [176].

A brain region that is part of the posterior salience network and is relevant to analysis is the posterior insula. The posterior insula seems to be associated with the recognition, intensity encoding, localisation, learning, and memory of somatically painful events [234]. In relation to Parkinson's disease, Criaud et al. studied the contribution of insula in Parkinson's disease. This work showed evidence that "important functional distribution of different domains within the insular cortex in PD, particularly in relation to non-motor aspects, with an influence of medication effect" [235].

Regarding the posterior salience network, the literature seems to show evidence of brain alterations that are related to non-motor manifestations of Parkinson's disease, and also related to motor manifestations, since structures of the basal ganglia are involved in the posterior salience network.

5.3.1.4 Overview

Considering the explanations extracted from the XAI analysis of the CNN model for Parkinson's disease classification, and their comparison with the literature referent to brain alterations related to this neurodegenerative disease, there is evidence that the resting-state networks identified as relevant by the

LRP method in terms of functional connectivity (dorsal default mode, ventral default mode and posterior salience), are related to the changes of the brain regions that are present in the respective networks due to PD pathophysiology.

Nevertheless, there is a possible limitation that may influence our results. As previously mentioned in the discussion of the results regarding the development and evaluation of the deep learning model, we cannot disregard completely the existent chance of the discrimination between PD patients and healthy controls performed by the algorithm as a result of possible differences between the FC matrices extracted from the PPMI data set and from the ADNI data set. Although this is a probable hypothesis, we have at least two points that verify that this possibility is remote. First, as mentioned in the Methods chapter, studies give evidence that conventional resting-state metrics, such as functional connectivity, are not significantly affected by different TRs [193]. Between fMRI acquisition parameters from PPMI and ADNI, the most significant difference are the values of TR, which are 2500 ms for PPMI and 3000 ms for ADNI. Hence, it is not expected, according to this study, a pronounced difference between controls from ADNI and controls from PPMI. In addition, if indeed there were significant differences between both groups of data sets, the explanations provided by the LRP methods would display significant explanations in a more general distribution of the matrix, i.e., more features would be identified as relevant for the classification, since the matrices between both groups would be completely distinct. However, the LRP method produced very specific and not sparse explanations.

Hence, it is possible to consider the process of XAI analysis and identification of Parkinson's disease biomarkers related to functional connectivity changes was fairly successful and trustful.

5.4 Additional Task: Attempt to Apply Transfer Learning

In this section, we will describe the results regarding the additional task taken. This task involved an attempt to apply transfer learning from an ExtendedConnectomeCNN trained on the ABIDE data set, in order to surpass the limitation of the reduced size of the PPMI + ADNI data set. The ABIDE data set is larger and is composed of a similar type of data of the PPMI + ADNI set, which comprises functional connectivity matrices. However, the sizes of the matrices are different, hence, the PPMI + ADNI set would have to be pre-processed all over again, accordingly the AAL atlas. This way, all the FC matrices would have the 116×116 size.

The first step of this additional task was to retrieve the FC matrices and respective labels from the ABIDE data set. Then, we proceeded to perform 10-fold cross-validation with those data on the ExtendedConnectomeCNN. Three different architectures with different values of dropout rate were experimented with. The respective results are represented in Table 5.8.

Table 5.8: Performance metrics on the classification task healthy controls (0) vs ASD (1): 10-fold cross-validation.

Cross-validation: mean values for 10 folds						
ID	Training Accuracy	Validation Accuracy	AUC ROC	Sensitivity	Specificity	F1-Score
1	0.5379	0.5190	0.1005	0.0730	0.9119	0.4924
2	1.0000	0.6430	0.5739	0.5162	0.7548	0.6355
3	0.8168	0.5937	0.4753	0.3973	0.7667	0.5820

As it is observed in Table 5.8, the performance of all combinations was poor. Combination 1 resulted in a training accuracy of 0.5379 and validation of 0.5190, reflecting that the models were "severely" underfitting and not generalising. Furthermore, the performance on each one of the positive and negative

classes is very imbalanced, which is verified in the AUC ROC, sensitivity, specificity and F1-score metrics. In particular, the model is critically missing the positive class since the sensitivity is only 0.0730 (almost null). Combination 2, on the other hand, is overfitting: the training accuracy is equal to 1. The models still do not generalise: the validation accuracy is 0.6430. The remaining metrics also reflect a poor performance on the data, but better than combination 1. Combination 3, although is not overfitting, the validation accuracy is lower than the validation accuracy from combination 2. The sensitivity value of combination 3 (0.3973) is also lower than the value from combination 2 (0.5162).

It is evident from the cross-validation results that the model resulted in poor performance and imbalanced metrics, not generalising on the data. The performance of the models trained directly in the PPMI + ADNI data set was largely superior. Hence, we can already assume that using the ABIDE data to create a model to apply transfer learning may not be the ideal approach for this problem. Even so, combination 2 was chosen as the final model and it was trained on the entire development set and test set on the test data. The results are displayed in Table 5.9. The respective confusion matrix on the test set is represented in Figure 5.15.

Table 5.9: Performance metrics on the classification task healthy controls (0) vs ASD (1): final model on the test set.

Final Model Performance on Test Set					
Training Accuracy	Validation Accuracy	AUC	Sensitivity	Specificity	F1-Score
1.000	0.6136	0.6087	0.5366	0.6809	0.5641

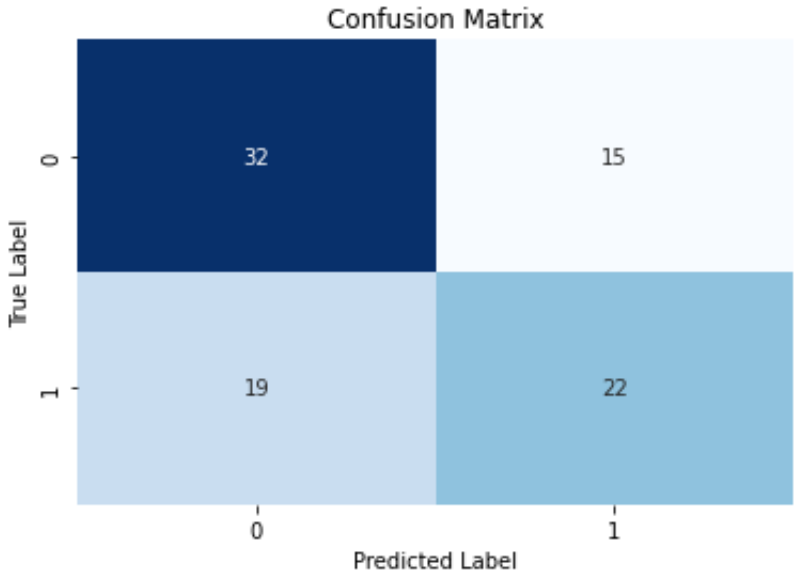

Figure 5.15: Confusion matrix for the classification on the test set with the final model; 0: healthy control, 1: ASD.

Analysing the results from Table 5.9, we can notice that the poor and imbalanced performance remains. The model is overfitting (training accuracy is 1.000) and not generalising on the data (validation accuracy is 0.6136). Considering then that the performance of the model on ABIDE data is worse than the performance on the PPMI + ADNI data set (refer to section 5.2 to assess the deep learning performance results), and the time limitation of this work, it is concluded that using a model trained on the larger ABIDE data set to perform transfer learning on the PPMI + ADNI data set is not a proper and feasible approach in this context.

As aforementioned, in section 4.6, the verified poor performance may be due to the high heterogeneity of the ABIDE data set, caused by the widely variant manifestation and complex

pathophysiology of autism spectrum disease (ASD), and the variety of combinations of fMRI acquisition parameters existent in the set. Therefore, using a CNN model on ABIDE data is not the appropriate approach to construct a classification model. Other DL architectures, such as the graph neural networks, and using other types of data, such as phenotypic data, may be game-changing alternatives to handle ABIDE set.

We end the results and discussion chapter with the key idea that the proposed methods and strategies to study brain functional connectivity in Parkinson's disease, using explainable AI, despite some limitations, show promise of being a novel and innovative approach. In the next chapter, the dissertation will be concluded, where the main points of the work will be presented, as well as the encountered challenges and proposals for future work.

Chapter 6

Conclusion

One of the main challenges related to Parkinson's disease is the diagnosis, which is mostly solely based on the clinical presentation. Entities such as the Movement Disorders Society presented clinical diagnostic criteria for PD for research and clinical use. In some cases, image techniques such as SPECT, in particular, DaTScan, are used to identify the biomarker for nigrostriatal dopaminergic degeneration, the dopamine transporter. However, this imaging technique does not discriminate against other pathologies manifested with parkinsonism and nigrostriatal dopaminergic degeneration. In this dissertation, it is presented an approach that not only allows the identification of potential neuroimaging biomarkers, using fMRI scans and respective FC matrices, and XAI, but also supports the unravelling of the pathophysiologic mechanism of PD.

Initially, this work involved the literature review of the current approaches for the use of artificial intelligence for PD diagnosis, of the study of functional connectivity changes related to PD, and the use of XAI in the field of neuroimaging. The review of all these studies allowed us to define the work pipeline and methods to take on.

The first step of the methods was the pre-processing of the fMRI scans and computation of the FC matrices, from the PPMI data set. We included in our analysis 131 healthy controls from the ADNI data set as well, to balance our set. The results from the pre-processing and the respective quality assurance plots led us to conclude that this step was correctly performed. Nevertheless, there is always room to improve the pre-processing pipeline with novel techniques.

We then proceeded to the development of the ExtendedConnectomeCNN model for PD classification. This process included hyperparameters optimisation through 10-fold cross-validation, and then a final model was tested on a separate test set. A generalisable model with fairly acceptable performance was obtained, despite some challenges that will be later discussed.

Lastly, the XAI methods layer-wise relevance propagation, deconvolutional network and guided backpropagation were applied. According to the AOPC analysis, only the LRP analysis produced sufficiently explainable interpretations, then solely the results from this method were further analysed. We did not identify any alterations related to the basal ganglia network, as it was supposed to be observed. Nonetheless, changes in the dorsal default mode, ventral default mode and posterior salience networks were found, which are also involved in the pathophysiology of Parkinson's disease.

Additionally, an attempt to perform transfer learning to our data set by training a model on the larger ABIDE set was executed. The resulting model presented a poor performance and was not generalising. To repeat the pre-processing of the fMRI data into the AAL atlas, and to fine-tune a model resulting from TL, we would need to take a considerable amount of time, which was not possible in the context of this dissertation. Therefore, we disregarded this possibility, as the model trained on the PPMI + ADNI data

set still presented satisfactory results.

Notwithstanding, some challenges and limitations were encountered during our work. The first one to note is the small size of our data set, which is limited considering that we are using AI methods that demand the use of great amounts of data. However, acceptable results were obtained, considering this limitation. Furthermore, despite not having evidence on this, we cannot disregard the possibility of the use of ADNI controls affecting our results, even if it is remote. Regarding the XAI analysis, more methods could be explored, with different characteristics regarding scope, methodology and usage. Finally, the symptoms and usage of medication from the PD patients were not considered in this work, for simplification purposes. However, it would be greatly advantageous to take these into account since it would give more robustness to our study.

In conclusion, owning to the fact that:

1. the pre-processing of the fMRI scans followed an adequate pipeline and produced acceptable results;

2. the developed CNN model for PD classification proved to be sufficiently generalisable, with satisfactory performance metrics and balanced metrics;

3. the XAI analysis seems to be reliable and accordingly the literature on alterations in resting-state networks related to PD;

we can infer that the approach taken to assess brain functional connectivity changes in Parkinson's disease using XAI methods was fairly successful. In this dissertation, we have been breaking new ground in the computer-aided diagnosis of Parkinson's disease using functional MRI scans and the resultant functional connectivity matrices, and in the application of explainable AI methods interpret the results of the classification model and to seek for potential diagnostic biomarkers by analysis neuroimaging scans. The objectives of the dissertation were achieved, hoping that it will contribute to the progression of the development of alternative and novel techniques for Parkinson's disease diagnosis.

6.1 Future Work

From this work, some additional research topics and study challenges are proposed for future work. As aforementioned, further novel methods of pre-processing of the fMRI data could be explored. In addition, other metrics of functional connectivity can be applied. A multi-metric approach could also be considered for further investigation on this topic. Furthermore, as previously mentioned, more XAI methods, with different characteristics from the already experimented methods, should be explored, in order to assess if other important networks in the diagnosis of PD are captured.

In this work, it was opted to use a convolutional neural network to build a classification. However, given the complexity of fMRI data, more sophisticated and elaborate DNN architectures could be explored. One example that would capture the complexity of the FC matrices would be the graph neural networks, a type of model that is designed to process and analyse data structured as graphs by capturing and propagating information between interconnected nodes. There are already some studies that opt for this approach on FC data, showing promising results.

Moreover, as an alternative to the binary classification approach, a multiclass classification task, between Parkinson's, prodromal, and controls, could be done. The objective would be to additionally assess how brain functional connectivity changes upon the onset of prodromal symptoms related to

parkinsonism, and how would it differ from a healthy state and a Parkinson's disease state. This approach would allow a deeper understanding of the pathophysiology of Parkinson's disease.

At last, the used data to feed the DL model could be more complex, for example, by incorporating a neuroimaging multimodality approach, including, for instance, structural connectivity from anatomical MRI scans, or diffusion tensor imaging. This would provide more information, and, hence, create a more specific analysis of potential biomarkers for Parkinson's disease. Furthermore, adding data such as phenotypical manifestations and symptoms of the PD patients, demographic information such as age, the corresponding subtype of PD, and information regarding the medication, would deepen the study on brain connectivity of Parkinson's disease since the related factors definitely affect connectivity and neuroplasticity of the brain. Henceforth, it would be advantageous to include a broader range of this essential information in future research concerning the diagnosis of Parkinson's disease and the identification of potential biomarkers.

References

[1] G. D. Hammer and S. J. McPhee, *Parkinson Disease*. McGraw-Hill Education, 2019, p. 418–421.

[2] J. J. Ferreira, N. Gonçalves, A. Valadas, C. Januário, M. R. Silva, L. Nogueira, J. L. Vieira, and A. B. Lima, "Prevalence of Parkinson's disease: A population-based study in Portugal," *European Journal of Neurology*, vol. 24, no. 5, p. 748–750, 2017.

[3] R. L. Nussbaum and C. E. Ellis, "Alzheimer's disease and Parkinson's disease," *New England Journal of Medicine*, vol. 348, no. 14, p. 1356–1364, 2003.

[4] O.-B. Tysnes and A. Storstein, "Epidemiology of Parkinson's disease," *Journal of Neural Transmission*, vol. 124, no. 8, p. 901–905, 2017.

[5] "Parkinson disease," Jun 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/parkinson-disease

[6] E. R. Dorsey, A. Elbaz, E. Nichols, N. Abbasi, F. Abd-Allah, A. Abdelalim, J. C. Adsuar, M. G. Ansha, C. Brayne, J.-Y. J. Choi, and et al., "Global, regional, and national burden of Parkinson's disease, 1990–2016: A systematic analysis for the global burden of disease study 2016," *The Lancet Neurology*, vol. 17, no. 11, p. 939–953, 2018.

[7] S. Siuly and Y. Zhang, "Medical big data: Neurological diseases diagnosis through medical data analysis," *Data Science and Engineering*, vol. 1, no. 2, p. 54–64, Jun 2016.

[8] L. Zhang, M. Wang, M. Liu, and D. Zhang, "A survey on deep learning for neuroimaging-based brain disorder analysis," *Frontiers in Neuroscience*, vol. 14, Oct 2020.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436–444, May 2015.

[10] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608014002135

[11] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841517301135

[12] S. Vieira, W. H. Pinaya, and A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications," *Neuroscience amp; Biobehavioral Reviews*, vol. 74, p. 58–75, Mar 2017.

[13] A. Das and P. Rad, "Opportunities and challenges in explainable artificial intelligence (XAI): A survey," *CoRR*, vol. abs/2006.11371, 2020. [Online]. Available: https://arxiv.org/abs/2006.11371

[14] B. H. van der Velden, H. J. Kuijf, K. G. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," *Medical Image Analysis*, vol. 79, p. 102470, Jul 2022.

[15] L. M. de Lau and M. M. Breteler, "Epidemiology of Parkinson's disease," *The Lancet Neurology*, vol. 5, no. 6, pp. 525–535, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474442206704719

[16] D. Purves, *Neuroscience*. Sinauer, 2008.

[17] J. L. Lanciego, N. Luquin, and J. A. Obeso, "Functional neuroanatomy of the basal ganglia," *Cold Spring Harbor Perspectives in Medicine*, vol. 2, no. 12, 2012.

[18] "Basal ganglia: Group of structures in the brain," May 2022. [Online]. Available: https://brainmadesimple.com/basal-ganglia-structure/

[19] N. Bidesi, I. Vang Andersen, A. Windhorst, V. Shalgunov, and M. Herth, "The role of neuroimaging in Parkinson's disease," *Journal of Neurochemistry*, vol. 159, no. 4, p. 660–689, Sep 2021.

[20] G. E. Vázquez-Vélez and H. Y. Zoghbi, "Parkinson's disease genetics and pathophysiology," *Annual Review of Neuroscience*, vol. 44, no. 1, p. 87–108, Jul 2021.

[21] H.-C. Cheng, C. M. Ulane, and R. E. Burke, "Clinical progression in Parkinson disease and the neurobiology of axons," *Annals of Neurology*, vol. 67, no. 6, p. 715–725, Jun 2010.

[22] T. Iwatsubo, "Aggregation of α-synuclein in the pathogenesis of Parkinson's disease," *Journal of Neurology*, vol. 250, no. 0, p. 1–1, Oct 2003.

[23] V. N. Uversky, J. Li, K. Bower, and A. L. Fink, "Synergistic effects of pesticides and metals on the fibrillation of α-synuclein: Implications for Parkinson's disease," *NeuroToxicology*, vol. 23, no. 4–5, p. 527–536, Oct 2002.

[24] H. Braak, K. D. Tredici, U. Rüb, R. A. de Vos, E. N. Jansen Steur, and E. Braak, "Staging of brain pathology related to sporadic Parkinson's disease," *Neurobiology of Aging*, vol. 24, no. 2, pp. 197–211, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0197458002000659

[25] A. Kouli, K. M. Torsney, and W.-L. Kuan, "Parkinson's disease: Etiology, neuropathology, and pathogenesis," *Parkinson's Disease: Pathogenesis and Clinical Aspects*, p. 3–26, Dec 2018.

[26] D. J. Gelb, E. Oliver, and S. Gilman, "Diagnostic Criteria for Parkinson Disease," *Archives of Neurology*, vol. 56, no. 1, pp. 33–39, 01 1999. [Online]. Available: https://doi.org/10.1001/archneur.56.1.33

[27] J. Jankovic, J. C. Mazziotta, S. L. Pomeroy, and N. J. Newman, *Diagnosis and Assessment of Parkinson Disease and Other Movement Disorders*. Elsevier, 2022.

[28] M. J. Armstrong and M. S. Okun, "Diagnosis and Treatment of Parkinson Disease: A Review," *JAMA*, vol. 323, no. 6, pp. 548–560, 02 2020. [Online]. Available: https://doi.org/10.1001/jama.2019.22360

[29] A. A. Moustafa, S. Chakravarthy, J. R. Phillips, A. Gupta, S. Keri, B. Polner, M. J. Frank, and M. Jahanshahi, "Motor symptoms in Parkinson's disease: A unified framework," *Neuroscience Biobehavioral Reviews*, vol. 68, pp. 727–740, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0149763415300919

[30] A. H. Schapira, K. R. Chaudhuri, and P. Jenner, "Non-motor features of Parkinson disease," *Nature Reviews Neuroscience*, vol. 18, no. 7, p. 435–450, Jun 2017.

[31] H. Tibar, K. El Bayad, A. Bouhouche, E. H. Ait Ben Haddou, A. Benomar, M. Yahyaoui, A. Benazzouz, and W. Regragui, "Non-motor symptoms of Parkinson's disease and their impact on quality of life in a cohort of moroccan patients," *Frontiers in Neurology*, vol. 9, Apr 2018.

[32] C. Pellicano, D. Benincasa, V. Pisani, F. R. Buttarelli, M. Giovannelli, and F. E. Pontieri, "Prodromal non-motor symptoms of Parkinson's disease," *Neuropsychiatric Disease and Treatment*, vol. 3, no. 1, p. 145–151, Feb 2007.

[33] A. H. Rajput, "Levodopa prolongs life expectancy and is non-toxic to substantia nigra," *Parkinsonism amp; Related Disorders*, vol. 8, no. 2, p. 95–100, 2001.

[34] J. D. Obray, C. A. Small, E. K. Baldwin, E. Y. Jang, J. G. Lee, C. H. Yang, J. T. Yorgason, and S. C. Steffensen, "Dopamine D2-subtype receptors outside the blood-brain barrier mediate enhancement of mesolimbic dopamine release and conditioned place preference by intravenous dopamine," *Frontiers in Cellular Neuroscience*, vol. 16, Jul 2022.

[35] J. J. Ferreira, R. Katzenschlager, B. R. Bloem, U. Bonuccelli, D. Burn, G. Deuschl, E. Dietrichs, G. Fabbrini, A. Friedman, P. Kanovsky, and et al., "Summary of the recommendations of the EFNS/MDS-es review on therapeutic management of Parkinson's disease," *European Journal of Neurology*, vol. 20, no. 1, p. 5–15, Dec 2012.

[36] J. Zhang and L. Chew-Seng Tan, "Revisiting the medical management of Parkinson's disease: Levodopa versus dopamine agonist," *Current Neuropharmacology*, vol. 14, no. 4, p. 356–363, Apr 2016.

[37] R. P. Munhoz, M. Picillo, S. H. Fox, V. Bruno, M. Panisset, C. R. Honey, and A. Fasano, "Eligibility criteria for deep brain stimulation in Parkinson's disease, tremor, and dystonia," *Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques*, vol. 43, no. 4, p. 462–471, May 2016.

[38] M. Bucur and C. Papagno, "Deep brain stimulation in Parkinson disease: A meta-analysis of the long-term neuropsychological outcomes," *Neuropsychology Review*, vol. 33, no. 2, p. 307–346, Mar 2022.

[39] U. V. Mahajan, V. K. Ravikumar, K. K. Kumar, S. Ku, D. I. Ojukwu, C. Kilbane, P. Ghanouni, J. M. Rosenow, S. C. Stein, and C. H. Halpern, "Bilateral deep brain stimulation is the procedure to beat for advanced Parkinson disease: A meta-analytic, cost-effective threshold analysis for focused ultrasound," *Neurosurgery*, vol. 88, no. 3, p. 487–496, Dec 2020.

[40] R. B. Postuma, D. Berg, M. Stern, W. Poewe, C. W. Olanow, W. Oertel, J. Obeso, K. Marek, I. Litvan, A. E. Lang, and et al., "MDS clinical diagnostic criteria for Parkinson's disease," *Movement Disorders*, vol. 30, no. 12, p. 1591–1601, Oct 2015.

[41] S. H. Isaacson, S. Fisher, F. Gupta, N. Hermanowicz, D. E. Kremens, M. F. Lew, K. Marek, R. Pahwa, D. S. Russell, J. Seibyl, and et al., "Clinical utility of DaTSCAN™ imaging in the evaluation of patients with parkinsonism: A US perspective," *Expert Review of Neurotherapeutics*, vol. 17, no. 3, p. 219–225, Nov 2016.

[42] R. de la Fuente-Fernandez, "Role of DaTSCAN and clinical diagnosis in Parkinson disease," *Neurology*, vol. 78, no. 10, p. 696–701, Feb 2012.

[43] L. Marsili, G. Rizzo, and C. Colosimo, "Diagnostic criteria for Parkinson's disease: From James Parkinson to the concept of prodromal disease," *Frontiers in Neurology*, vol. 9, Mar 2018.

[44] S. N. Gomperts, "Lewy body dementias: Dementia with Lewy Bodies and Parkinson Disease Dementia," *CONTINUUM: Lifelong Learning in Neurology*, vol. 22, no. 2, Dementia, p. 435–463, Apr 2016.

[45] A. Fanciulli and G. K. Wenning, "Multiple-system atrophy," *New England Journal of Medicine*, vol. 372, no. 3, p. 249–263, Jan 2015.

[46] D. R. Williams and A. J. Lees, "Progressive supranuclear palsy: clinicopathological concepts and diagnostic challenges," *The Lancet Neurology*, vol. 8, no. 3, pp. 270–279, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474442209700420

[47] R. Mathew, T. H. Bak, and J. R. Hodges, "Diagnostic criteria for corticobasal syndrome: A comparative study," *Journal of Neurology, Neurosurgery amp; Psychiatry*, vol. 83, no. 4, p. 405–410, Oct 2011.

[48] G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, and G. Logroscino, "Accuracy of clinical diagnosis of Parkinson disease," *Neurology*, vol. 86, no. 6, pp. 566–576, 2016. [Online]. Available: https://n.neurology.org/content/86/6/566

[49] E. Tolosa, T. V. Borght, and E. Moreno, "Accuracy of DaTSCAN (123i-ioflupane) SPECT in diagnosis of patients with clinically uncertain parkinsonism: 2-year follow-up of an open-label study," *Movement Disorders*, vol. 22, no. 16, p. 2346–2351, Oct 2007.

[50] R. W. Brown, *Magnetic Resonance Imaging: Physical principles and sequence design*. Wiley-Blackwell, 2014.

[51] 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. 246–255, Sep 2015.

[52] N. H. Kalin, "Understanding the value and limitations of mri neuroimaging in psychiatry," *American Journal of Psychiatry*, vol. 178, no. 8, p. 673–676, Aug 2021.

[53] B. Ibrahim, S. Suppiah, N. Ibrahim, M. Mohamad, H. A. Hassan, N. S. Nasser, and M. I. Saripan, "Diagnostic power of resting-state fMRI for detection of network connectivity in Alzheimer's disease and mild cognitive impairment: A systematic review," *Human Brain Mapping*, vol. 42, no. 9, p. 2941–2968, May 2021.

[54] B. H. Brown, R. H. Smallwood, and D. Hose, *Biomedical physics and engineering*. Institute of Physics, 1998.

[55] R. W. Chan, J. Y. Lau, W. W. Lam, and A. Z. Lau, "Magnetic resonance imaging," *Encyclopedia of Biomedical Engineering*, p. 574–587, 2019.

[56] A. Elnakib, "Developing advanced mathematical models for detecting abnormalities in 2D/3D medical structures," Ph.D. dissertation, 12 2013.

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

[58] R. B. Buxton, E. C. Wong, and L. R. Frank, "Dynamics of blood flow and oxygenation changes during brain activation: The balloon model," *Magnetic Resonance in Medicine*, vol. 39, no. 6, p. 855–864, Jun 1998.

[59] T. L. Davis, K. K. Kwong, R. M. Weisskoff, and B. R. Rosen, "Calibrated functional MRI: Mapping the dynamics of oxidative metabolism," *Proceedings of the National Academy of Sciences*, vol. 95, no. 4, p. 1834–1839, Feb 1998.

[60] J. E. Chen and G. H. Glover, "Functional magnetic resonance imaging methods," *Neuropsychology Review*, vol. 25, no. 3, p. 289–313, Aug 2015.

[61] R. B. Buxton, "The physics of functional magnetic resonance imaging (fMRI)," *Reports on Progress in Physics*, vol. 76, no. 9, p. 096601, Sep 2013.

[62] K. R. Thulborn, J. C. Waterton, P. M. Matthews, and G. K. Radda, "Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field," *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 714, no. 2, p. 265–270, Feb 1982.

[63] J. L. Boxerman, P. A. Bandettini, K. K. Kwong, J. R. Baker, T. L. Davis, B. R. Rosen, and R. M. Weisskoff, "The intravascular contribution to fMRI signal change: Monte carlo modeling and diffusion-weighted studies in vivo," *Magnetic Resonance in Medicine*, vol. 34, no. 1, p. 4–10, Jul 1995.

[64] J. Siero, A. Bhogal, and J. Jansma, "Blood oxygenation level–dependent/functional magnetic resonance imaging : Underpinnings, practice, and perspectives," *PET Clinics*, vol. 8, p. 329–344, 07 2013.

[65] S. Lang, N. Duncan, and G. Northoff, "Resting-state functional magnetic resonance imaging," *Neurosurgery*, vol. 74, no. 5, p. 453–465, May 2014.

[66] M. Lee, C. Smyser, and J. Shimony, "Resting-state fmri: A review of methods and clinical applications," *American Journal of Neuroradiology*, vol. 34, no. 10, p. 1866–1872, Aug 2012.

[67] E. W. Lang, A. M. Tomé, I. R. Keck, J. M. Górriz-Sáez, and C. G. Puntonet, "Brain connectivity analysis: A short survey," *Computational Intelligence and Neuroscience*, vol. 2012, p. 1–21, 2012.

[68] P. Babaeeghazvini, L. M. Rueda-Delgado, J. Gooijers, S. P. Swinnen, and A. Daffertshofer, "Brain structural and functional connectivity: A review of combined works of diffusion magnetic resonance imaging and electro-encephalography," *Frontiers in Human Neuroscience*, vol. 15, Oct 2021.

[69] H. E. Wang, C. G. BÃ©nar, P. P. Quilichini, K. J. Friston, V. K. Jirsa, and C. Bernard, "A systematic framework for functional connectivity measures," *Frontiers in Neuroscience*, vol. 8, Dec 2014.

[70] B. Ellenbroek and J. Youn, "Schizophrenia," *Gene-Environment Interactions in Psychiatry*, p. 289–322, 2016.

[71] S. Eickhoff and V. Müller, "Functional connectivity," *Brain Mapping*, p. 187–201, 2015.

[72] J. Zhang, A. Kucyi, J. Raya, A. N. Nielsen, J. S. Nomi, J. S. Damoiseaux, D. J. Greene, S. G. Horovitz, L. Q. Uddin, S. Whitfield-Gabrieli, and et al., "What have we really learned from functional connectivity in clinical populations?" *NeuroImage*, vol. 242, p. 118466, Nov 2021.

[73] L. Sang, J. Zhang, L. Wang, J. Zhang, Y. Zhang, P. Li, J. Wang, and M. Qiu, "Alteration of brain functional networks in early-stage Parkinson's disease: A resting-state fMRI study," *PLOS ONE*, vol. 10, no. 10, Oct 2015.

[74] A. Nieto-Castanon, "fMRI minimal preprocessing pipeline," *Handbook of functional connectivity Magnetic Resonance Imaging methods in CONN*, p. 3–16, Feb 2020.

[75] H. Lv, Z. Wang, E. Tong, L. Williams, G. Zaharchuk, M. Zeineh, A. Goldstein-Piekarski, T. Ball, C. Liao, M. Wintermark, and et al., "Resting-state functional MRI: Everything that nonexperts have always wanted to know," *American Journal of Neuroradiology*, Jan 2018.

[76] 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, p. 537–541, Oct 1995.

[77] R. Mohanty, W. A. Sethares, V. A. Nair, and V. Prabhakaran, "Rethinking measures of functional connectivity via feature extraction," *Scientific Reports*, vol. 10, no. 1, Jan 2020.

[78] R. A. Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, vol. 10, no. 4, p. 507, May 1915.

[79] V. Calhoun, T. Adali, G. Pearlson, and J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Human Brain Mapping*, vol. 14, no. 3, p. 140–151, 2001.

[80] H. Yang, X.-Y. Long, Y. Yang, H. Yan, C.-Z. Zhu, X.-P. Zhou, Y.-F. Zang, and Q.-Y. Gong, "Amplitude of low frequency fluctuation within visual areas revealed by resting-state functional MRI," *NeuroImage*, vol. 36, no. 1, p. 144–152, May 2007.

[81] E. T. Rolls, C.-C. Huang, C.-P. Lin, J. Feng, and M. Joliot, "Automated anatomical labelling atlas 3," *NeuroImage*, vol. 206, p. 116189, Feb 2020.

[82] 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, and et al., "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of Neurophysiology*, vol. 106, no. 3, p. 1125–1165, Sep 2011.

[83] 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, p. 158–165, May 2011.

[84] J. J. Kuiper, Y.-H. Lin, I. M. Young, M. Y. Bai, R. G. Briggs, O. Tanglay, R. D. Fonseka, J. Hormovas, V. Dhanaraj, A. K. Conner, and et al., "A parcellation-based model of the auditory network," *Hearing Research*, vol. 396, p. 108078, Oct 2020.

[85] D. J. Greene, S. Marek, E. M. Gordon, J. S. Siegel, C. Gratton, T. O. Laumann, A. W. Gilmore, J. J. Berg, A. L. Nguyen, D. Dierker, and et al., "Integrative and network-specific connectivity of the basal ganglia and thalamus defined in individuals," *Neuron*, vol. 105, no. 4, Feb 2020.

[86] 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," *The Journal of Neuroscience*, vol. 41, no. 24, p. 5243–5250, May 2021.

[87] O. E. Arslan, "Computational basis of neural elements," *Artificial Neural Network for Drug Design, Delivery and Disposition*, p. 29–82, 2016.

[88] L. Jin, C. Li, Y. Zhang, T. Yuan, J. Ying, Z. Zuo, and S. Gui, "The functional reorganization of language network modules in glioma patients: New insights from resting state fMRI study," *Frontiers in Oncology*, vol. 11, Feb 2021.

[89] W. Xu, S. Chen, C. Xue, G. Hu, W. Ma, W. Qi, X. Lin, and J. Chen, "Functional MRI-specific alterations in executive control network in mild cognitive impairment: An ale meta-analysis," *Frontiers in Aging Neuroscience*, vol. 12, Oct 2020.

[90] 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, Jun 2016.

[91] E. F. Pace-Schott and D. Picchioni, "Neurobiology of dreaming," *Principles and Practice of Sleep Medicine*, 2017.

[92] D. Borsook, N. Maleki, and R. Burstein, "Migraine," *Neurobiology of Brain Disorders*, p. 693–708, 2015.

[93] G. Raffa, M. C. Quattropani, G. Marzano, A. Curcio, V. Rizzo, G. Sebestyén, V. Tamás, A. Büki, and A. Germanò, "Mapping and preserving the visuospatial network by repetitive NTMS and DTI tractography in patients with right parietal lobe tumors," *Frontiers in Oncology*, vol. 11, Jun 2021.

[94] C. Zhang and Y. Lu, "Study on artificial intelligence: The state of the art and future prospects," *Journal of Industrial Information Integration*, vol. 23, p. 100224, Sep 2021.

[95] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Future Healthcare Journal*, vol. 6, no. 2, p. 94–98, Jun 2019.

[96] P. A. Flach, *Machine learning: The art and science of algorithms that make sense of data*. Cambridge University Press, 2017.

[97] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, p. 53040–53065, 2019.

[98] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Computer Science Review*, vol. 40, May 2021.

[99] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A survey of deep learning and its applications: A new paradigm to machine learning," *Archives of Computational Methods in Engineering*, vol. 27, no. 4, p. 1071–1092, Jun 2019.

[100] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, pp. 115–133, 1943.

[101] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational Intelligence and Neuroscience*, vol. 2018, p. 1–13, 2018.

[102] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain." *Psychological Review*, vol. 65, no. 6, p. 386–408, 1958.

[103] Y. Sai, R. Jinxia, and L. Zhongxia, "Learning of neural networks based on weighted mean squares error function," in *2009 Second International Symposium on Computational Intelligence and Design*, vol. 1, 2009, pp. 241–244.

[104] U. Ruby and V. Yendapalli, "Binary cross entropy with deep learning technique for image classification," *Int. J. Adv. Trends Comput. Sci. Eng*, vol. 9, no. 10, 2020.

[105] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[106] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," *CoRR*, vol. abs/1502.01852, 2015. [Online]. Available: http://arxiv.org/abs/1502.01852

[107] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MITP, 2018.

[108] K. Stąpor, "Evaluating and comparing classifiers: Review, some recommendations and limitations," *Advances in Intelligent Systems and Computing*, p. 12–21, May 2017.

[109] [Online]. Available: https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc

[110] N. Aloysius and M. Geetha, "A review on deep convolutional neural networks," in *2017 International Conference on Communication and Signal Processing (ICCSP)*, 2017, pp. 0588–0592.

[111] M. Chen, H. Li, H. Fan, J. R. Dillman, H. Wang, M. Altaye, B. Zhang, N. A. Parikh, and L. He, "ConceptCNN: A novel multi-filter convolutional neural network for the prediction of neurodevelopmental disorders using brain connectome," *Medical Physics*, vol. 49, no. 5, p. 3171–3184, Mar 2022.

[112] Z. Sherkatghanad, M. Akhondzadeh, S. Salari, M. Zomorodi-Moghadam, M. Abdar, U. R. Acharya, R. Khosrowabadi, and V. Salari, "Automated detection of autism spectrum disorder using a convolutional neural network," *Frontiers in Neuroscience*, vol. 13, Jan 2020.

[113] J. Smucny, G. Shi, and I. Davidson, "Deep learning in neuroimaging: Overcoming challenges with emerging approaches," *Frontiers in Psychiatry*, vol. 13, Jun 2022.

[114] M. Yousefnezhad, A. Selvitella, D. Zhang, A. J. Greenshaw, and R. Greiner, "Shared Space Transfer Learning for analyzing multi-site fMRI data," *arXiv e-prints*, p. arXiv:2010.15594, Oct. 2020.

[115] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Computers in biology and medicine*, vol. 111, p. 103345, 2019.

[116] M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, E. Marchiori, M. Pesteie, C. R. G. Guttmann, F.-E. de Leeuw, C. M. Tempany, B. van Ginneken, A. Fedorov, P. Abolmaesumi, B. Platel, and I. Wells, William M., "Transfer Learning for Domain Adaptation in MRI: Application in Brain Lesion Segmentation," *arXiv e-prints*, p. arXiv:1702.07841, Feb. 2017.

[117] M. Wang, D. Zhang, J. Huang, P.-T. Yap, D. Shen, and M. Liu, "Identifying autism spectrum disorder with multi-site fMRI via low-rank domain adaptation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 3, pp. 644–655, 2020.

[118] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond Empirical Risk Minimization," *arXiv e-prints*, p. arXiv:1710.09412, Oct. 2017.

[119] M. Nauta, J. Trienes, S. Pathak, E. Nguyen, M. Peters, Y. Schmitt, J. Schlötterer, M. van Keulen, and C. Seifert, "From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable AI," *CoRR*, vol. abs/2201.08164, 2022. [Online]. Available: https://arxiv.org/abs/2201.08164

[120] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 11, p. 4793–4813, Nov 2021.

[121] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, p. 52138–52160, 2018.

[122] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PLOS ONE*, vol. 10, no. 7, Jul 2015.

[123] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K.-R. Müller, "Layer-wise relevance propagation: An overview," *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, p. 193–209, 2019.

[124] G. Montavon, W. Samek, and K.-R. Müller, "Methods for interpreting and understanding deep neural networks," *Digital Signal Processing*, vol. 73, p. 1–15, Feb 2018.

[125] J. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," 12 2014.

[126] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," *CoRR*, vol. abs/1311.2901, 2013. [Online]. Available: http://arxiv.org/abs/1311.2901

[127] S. Mostafa, D. Mondal, M. A. Beck, C. P. Bidinosti, C. J. Henry, and I. Stavness, "Leveraging guided backpropagation to select convolutional neural networks for plant classification," *Frontiers in Artificial Intelligence*, vol. 5, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2022.871162

[128] S. Saravanan, K. Ramkumar, K. Adalarasu, V. Sivanandam, S. R. Kumar, S. Stalin, and R. Amirtharajan, "A systematic review of artificial intelligence (AI) based approaches for the diagnosis of Parkinson's disease," *Archives of Computational Methods in Engineering*, vol. 29, no. 6, p. 3639–3653, Jan 2022.

[129] S. Paul, M. Maindarkar, S. Saxena, L. Saba, M. Turk, M. Kalra, P. R. Krishnan, and J. S. Suri, "Bias investigation in artificial intelligence systems for early detection of Parkinson's disease: A narrative review," *Diagnostics*, vol. 12, no. 1, p. 166, Jan 2022.

[130] D. H. Kim, H. Wit, and M. Thurston, "Artificial intelligence in the diagnosis of Parkinson's disease from ioflupane-123 single-photon emission computed tomography dopamine transporter scans using transfer learning," *Nuclear Medicine Communications*, vol. 39, no. 10, p. 887–893, Oct 2018.

[131] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016.

[132] F. J. Martinez-Murcia, A. Ortiz, J. M. Górriz, J. Ramírez, F. Segovia, D. Salas-Gonzalez, D. Castillo-Barnes, and I. A. Illán, "A 3D convolutional neural network approach for the diagnosis of Parkinson's disease," *Natural and Artificial Computation for Biomedicine and Neuroscience*, p. 324–333, 2017.

[133] F. J. Martinez-Murcia, A. Ortiz, J. M. Gorriz, J. Ramirez, D. Castillo-Barnes, D. Salas-Gonzalez, and F. Segovia, "Deep convolutional autoencoders vs PCA in a highly-unbalanced Parkinson's disease dataset: A textDaTSCAN study," *Advances in Intelligent Systems and Computing*, p. 47–56, Jun 2018.

[134] S. Sivaranjini and C. M. Sujatha, "Deep learning based diagnosis of Parkinson's disease using convolutional neural network," *Multimedia Tools and Applications*, vol. 79, no. 21-22, p. 15467–15479, Mar 2019.

[135] J. Xu and M. Zhang, "Use of magnetic resonance imaging and artificial intelligence in studies of diagnosis of Parkinson's disease," *ACS Chemical Neuroscience*, vol. 10, no. 6, p. 2658–2667, May 2019.

[136] Z. Zhang, G. Li, Y. Xu, and X. Tang, "Application of artificial intelligence in the MRI classification task of human brain neurological and psychiatric diseases: A scoping review," *Diagnostics*, vol. 11, no. 8, p. 1402, Aug 2021.

[137] A. Nawaz, A. U. Rehman, T. M. Ali, Z. Hayat, A. Rahim, U. K. Uz Zaman, and A. R. Ali, "A comprehensive literature review of application of artificial intelligence in functional magnetic resonance imaging for disease diagnosis," *Applied Artificial Intelligence*, vol. 35, no. 15, p. 1420–1438, Oct 2021.

[138] D. Shi, H. Zhang, G. Wang, S. Wang, X. Yao, Y. Li, Q. Guo, S. Zheng, and K. Ren, "Machine learning for detecting Parkinson's disease by resting-state functional magnetic resonance imaging: A multicenter radiomics analysis," *Frontiers in Aging Neuroscience*, vol. 14, Mar 2022.

[139] Y. Chen, J. Storrs, L. Tan, L. J. Mazlack, J.-H. Lee, and L. J. Lu, "Detecting brain structural changes as biomarker from magnetic resonance images using a local feature based SVM approach," *Journal of Neuroscience Methods*, vol. 221, p. 22–31, Jan 2014.

[140] Y. Tang, L. Meng, C.-m. Wan, Z.-h. Liu, W.-h. Liao, X.-x. Yan, X.-y. Wang, B.-s. Tang, and J.-f. Guo, "Identifying the presence of Parkinson's disease using low-frequency fluctuations in BOLD signals," *Neuroscience Letters*, vol. 645, p. 1–6, Apr 2017.

[141] Y. Chen, W. Yang, J. Long, Y. Zhang, J. Feng, Y. Li, and B. Huang, "Discriminative analysis of Parkinson's disease based on whole-brain functional connectivity," *PLOS ONE*, vol. 10, no. 4, Apr 2015.

[142] R. N. Pläschke, E. C. Cieslik, V. I. Müller, F. Hoffstaedter, A. Plachti, D. P. Varikuti, M. Goosses, A. Latz, S. Caspers, C. Jockwitz, and et al., "On the integrity of functional brain networks in schizophrenia, Parkinson's disease, and advanced age: Evidence from connectivity-based single-subject classification," *Human Brain Mapping*, vol. 38, no. 12, p. 5845–5858, Sep 2017.

[143] B. Peng, S. Wang, Z. Zhou, Y. Liu, B. Tong, T. Zhang, and Y. Dai, "A multilevel-ROI-features-based machine learning method for detection of morphometric biomarkers in Parkinson's disease," *Neuroscience Letters*, vol. 651, p. 88–94, Jun 2017.

[144] I. D. Dinov, B. Heavner, M. Tang, G. Glusman, K. Chard, M. Darcy, R. Madduri, J. Pa, C. Spino, C. Kesselman, and et al., "Predictive big data analytics: A study of Parkinson's disease using large, complex, heterogeneous, incongruent, multi-source and incomplete observations," *PLOS ONE*, vol. 11, no. 8, Aug 2016.

[145] E. Adeli, F. Shi, L. An, C.-Y. Wee, G. Wu, T. Wang, and D. Shen, "Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data," *NeuroImage*, vol. 141, p. 206–219, Nov 2016.

[146] C. Rubbert, C. Mathys, C. Jockwitz, C. J. Hartmann, S. B. Eickhoff, F. Hoffstaedter, S. Caspers, C. R. Eickhoff, B. Sigl, N. A. Teichert, and et al., "Machine-learning identifies Parkinson's disease patients based on resting-state between-network functional connectivity," *The British Journal of Radiology*, vol. 92, no. 1101, p. 20180886, Sep 2019.

[147] S. Esmaeilzadeh, Y. Yang, and E. Adeli, "End-to-end Parkinson disease diagnosis using brain MR-images by 3D-CNN," *arXiv*, Jun 2018.

[148] X. Zhang, L. He, K. Chen, Y. Luo, J. Zhou, and F. Wang, "Multi-view graph convolutional network and its applications on neuroimage analysis for Parkinson's disease," *AMIA Annual Symposium proceedings*, Dec 2018.

[149] C. McDaniel and S. Quinn, "Developing a graph convolution-based analysis pipeline for multi-modal neuroimage data: An application to Parkinson's disease," *Proceedings of the 18th Python in Science Conference*, 2019.

[150] A. Safai, N. Vakharia, S. Prasad, J. Saini, A. Shah, A. Lenka, P. K. Pal, and M. Ingalhalikar, "Multimodal brain connectomics-based prediction of Parkinson's disease using graph attention networks," *Frontiers in Neuroscience*, vol. 15, Feb 2022.

[151] A. Dehsarvi and S. L. Smith, "Classification of resting-state fMRI using evolutionary algorithms: Towards a brain imaging biomarker for Parkinson's disease," *arXiv.org*, Oct 2019.

[152] X. Guo, S. Tinaz, and N. Dvornek, "Early disease stage characterization in parkinson's disease from resting-state fmri data using a long short-term memory network," Mar 2022.

[153] M. M. Hoehn and M. D. Yahr, "Parkinsonism: Onset, progression, and mortality," *Neurology*, vol. 17, no. 5, p. 427–427, May 1967.

[154] A. Tessitore, M. Cirillo, and R. De Micco, "Functional connectivity signatures of Parkinson's disease," *Journal of Parkinson's Disease*, vol. 9, no. 4, p. 637–652, Oct 2019.

[155] J. Fang, H. Chen, Z. Cao, Y. Jiang, L. Ma, H. Ma, and T. Feng, "Impaired brain network architecture in newly diagnosed Parkinson's disease based on graph theoretical analysis," *Neuroscience Letters*, vol. 657, p. 151–158, Sep 2017.

[156] T. Wu, X. Long, L. Wang, M. Hallett, Y. Zang, K. Li, and P. Chan, "Functional connectivity of cortical motor areas in the resting state in Parkinson's disease," *Human Brain Mapping*, vol. 32, no. 9, p. 1443–1457, Aug 2010.

[157] M. Rolinski, L. Griffanti, K. Szewczyk-Krolikowski, R. A. Menke, G. K. Wilcock, N. Filippini, G. Zamboni, M. T. Hu, and C. E. Mackay, "Aberrant functional connectivity within the basal ganglia of patients with Parkinson's disease," *NeuroImage: Clinical*, vol. 8, p. 126–132, 2015.

[158] H.-M. Chen, Z.-J. Wang, J.-P. Fang, L.-Y. Gao, L.-Y. Ma, T. Wu, Y.-N. Hou, J.-R. Zhang, and T. Feng, "Different patterns of spontaneous brain activity between tremor-dominant and postural instability/gait difficulty subtypes of Parkinson's disease: A resting-state fmri study," *CNS Neuroscience amp; Therapeutics*, vol. 21, no. 10, p. 855–866, Sep 2015.

[159] M. Rolinski, L. Griffanti, P. Piccini, A. A. Roussakis, K. Szewczyk-Krolikowski, R. A. Menke, T. Quinnell, Z. Zaiwalla, J. C. Klein, C. E. Mackay, and et al., "Basal ganglia dysfunction in idiopathic REM sleep behaviour disorder parallels that in early Parkinson's disease," *Brain*, vol. 139, no. 8, p. 2224–2234, Jun 2016.

[160] M. Su, S. Wang, W. Fang, Y. Zhu, R. Li, K. Sheng, D. Zou, Y. Han, X. Wang, O. Cheng, and et al., "Alterations in the limbic/paralimbic cortices of Parkinson's disease patients with hyposmia under resting-state functional MRI by regional homogeneity and functional connectivity analysis," *Parkinsonism amp; Related Disorders*, vol. 21, no. 7, p. 698–703, Jul 2015.

[161] Y. Lou, P. Huang, D. Li, Z. Cen, B. Wang, J. Gao, M. Xuan, H. Yu, M. Zhang, W. Luo, and et al., "Altered brain network centrality in depressed Parkinson's disease patients," *Movement Disorders*, vol. 30, no. 13, p. 1777–1784, 2015.

[162] A. Tessitore, F. Esposito, C. Vitale, G. Santangelo, M. Amboni, A. Russo, D. Corbo, G. Cirillo, P. Barone, G. Tedeschi, and et al., "Default-mode network connectivity in cognitively unimpaired patients with Parkinson disease," *Neurology*, vol. 79, no. 23, p. 2226–2232, Oct 2012.

[163] K. T. Olde Dubbelink, M. M. Schoonheim, J. B. Deijen, J. W. Twisk, F. Barkhof, and H. W. Berendse, "Functional connectivity and cognitive decline over 3 years in Parkinson disease," *Neurology*, vol. 83, no. 22, p. 2046–2053, Oct 2014.

[164] A. Cerasa, G. Koch, G. Donzuso, G. Mangone, M. Morelli, L. Brusa, M. Stampanoni Bassi, V. Ponzo, S. Picazio, L. Passamonti, and et al., "A network centred on the inferior frontal cortex is critically involved in levodopa-induced dyskinesias," *Brain*, vol. 138, no. 2, p. 414–427, Nov 2014.

[165] P. Manza, S. Zhang, C.-S. R. Li, and H.-C. Leung, "Resting-state functional connectivity of the striatum in early-stage Parkinson's disease: Cognitive decline and motor symptomatology," *Human Brain Mapping*, vol. 37, no. 2, p. 648–662, Nov 2015.

[166] L. I. Boon, D. H. Hepp, L. Douw, N. van Geenen, T. A. Broeders, J. J. Geurts, H. W. Berendse, and M. M. Schoonheim, "Functional connectivity between resting-state networks reflects decline in executive function in Parkinson's disease: A longitudinal fMRI study," *NeuroImage: Clinical*, vol. 28, p. 102468, 2020.

[167] F. Cao, X. Guan, Y. Ma, Y. Shao, and J. Zhong, "Altered functional network associated with cognitive performance in early Parkinson disease measured by eigenvector centrality mapping," *Frontiers in Aging Neuroscience*, vol. 12, 2020.

[168] N. M. Gage and B. J. Baars, "Humans are social beings," *Fundamentals of Cognitive Neuroscience*, p. 321–356, 2018.

[169] K. Tanaka, "Temporal lobe," *International Encyclopedia of the Social amp; Behavioral Sciences*, p. 15595–15599, 2001.

[170] O. Dipasquale and M. Cercignani, "Network functional connectivity and whole-brain functional connectomics to investigate cognitive decline in neurodegenerative conditions," *Functional Neurology*, 2016.

[171] K. Szewczyk-Krolikowski, R. A. Menke, M. Rolinski, E. Duff, G. Salimi-Khorshidi, N. Filippini, G. Zamboni, M. T. Hu, and C. E. Mackay, "Functional connectivity in the basal ganglia network differentiates PD patients from controls," *Neurology*, vol. 83, no. 3, p. 208–214, Jun 2014.

[172] D. Putcha, R. S. Ross, A. Cronin-Golomb, A. C. Janes, and C. E. Stern, "Altered intrinsic functional coupling between core neurocognitive networks in Parkinson's disease," *NeuroImage: Clinical*, vol. 7, p. 449–455, 2015.

[173] P. R. Karunanayaka, E.-Y. Lee, M. M. Lewis, S. Sen, P. J. Eslinger, Q. X. Yang, and X. Huang, "Default mode network differences between rigidity- and tremor-predominant parkinson's disease," *Cortex*, vol. 81, p. 239–250, Aug 2016.

[174] L. Wei, J. Zhang, Z. Long, G.-R. Wu, X. Hu, Y. Zhang, and J. Wang, "Reduced topological efficiency in cortical-basal ganglia motor network of Parkinson's disease: A resting state fMRI study," *PLoS ONE*, vol. 9, no. 10, Oct 2014.

[175] D. Zhang, X. Liu, J. Chen, B. Liu, and J. Wang, "Widespread increase of functional connectivity in Parkinson's disease with tremor: A resting-state fMRI study," *Frontiers in Aging Neuroscience*, vol. 7, Feb 2015.

[176] B. D. Berman, J. Smucny, K. P. Wylie, E. Shelton, E. Kronberg, M. Leehey, and J. R. Tregellas, "Levodopa modulates small-world architecture of functional brain networks in Parkinson's disease," *Movement Disorders*, vol. 31, no. 11, p. 1676–1684, Jul 2016.

[177] O. Kaut, C. Mielacher, R. Hurlemann, and U. Wüllner, "Resting-state fMRI reveals increased functional connectivity in the cerebellum but decreased functional connectivity of the caudate nucleus in Parkinson's disease," *Neurological Research*, vol. 42, no. 1, p. 62–67, Jan 2020.

[178] F. V. Farahani, K. Fiok, B. Lahijanian, W. Karwowski, and P. K. Douglas, "Explainable AI: A review of applications to neuroimaging data," *Frontiers in Neuroscience*, vol. 16, Dec 2022.

[179] R.-K. Sheu and M. S. Pardeshi, "A survey on medical explainable ai (XAI): Recent progress, explainability approach, human interaction and scoring system," *Sensors*, vol. 22, no. 20, p. 8068, Oct 2022.

[180] M. Böhle, F. Eitel, M. Weygandt, and K. Ritter, "Layer-wise relevance propagation for explaining deep neural network decisions in MRI-based Alzheimer's disease classification," *Frontiers in Aging Neuroscience*, vol. 11, Jul 2019.

[181] S. Chakraborty, S. Aich, and H.-C. Kim, "Detection of Parkinson's disease from 3T T1 weighted MRI scans using
text3D convolutional neural network," *Diagnostics*, vol. 10, no. 6, p. 402, Jun 2020.

[182] F. Dubost, H. Adams, P. Yilmaz, G. Bortsova, G. v. Tulder, M. A. Ikram, W. Niessen, M. W. Vernooij, and M. d. Bruijne, "Weakly supervised object detection with 2D and 3D regression neural networks," *Medical Image Analysis*, vol. 65, p. 101767, Oct 2020.

[183] H. Choi, Y. K. Kim, E. J. Yoon, J.-Y. Lee, and D. S. Lee, "Cognitive signature of brain FDG pet based on deep learning: Domain transfer from Alzheimer's disease to Parkinson's disease," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 47, no. 2, p. 403–412, Nov 2019.

[184] T. Fuchigami, S. Akahori, T. Okatani, and Y. Li, "A hyperacute stroke segmentation method using 3D U-Net integrated with physicians' knowledge for NCCT," *Medical Imaging 2020: Computer-Aided Diagnosis*, Mar 2020.

[185] Z. Lin, S. Li, D. Ni, Y. Liao, H. Wen, J. Du, S. Chen, T. Wang, and B. Lei, "Multi-task learning for quality assessment of fetal head ultrasound images," *Medical Image Analysis*, vol. 58, p. 101548, Dec 2019.

[186] J. Kubach, A. Muhlebner-Fahrngruber, F. Soylemezoglu, H. Miyata, P. Niehusmann, M. Honavar, F. Rogerio, S. Kim, E. Aronica, R. Garbelli, and et al., "Same same but different: A web-based deep learning application revealed classifying features for the histopathologic distinction of cortical malformations," *Epilepsia*, vol. 61, no. 3, p. 421–432, Mar 2020.

[187] J. Mourão-Miranda, A. L. Bokde, C. Born, H. Hampel, and M. Stetter, "Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data," *NeuroImage*, vol. 28, no. 4, p. 980–995, Dec 2005.

[188] Z. Wang, A. R. Childress, J. Wang, and J. A. Detre, "Support vector machine learning-based fMRI data group analysis," *NeuroImage*, vol. 36, no. 4, p. 1139–1151, Jul 2007.

[189] X. Li, N. C. Dvornek, J. Zhuang, P. Ventola, and J. S. Duncan, "Brain biomarker interpretation in ASD using deep learning and fMRI," *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, p. 206–214, 2018.

[190] A. W. Thomas, H. R. Heekeren, K.-R. Müller, and W. Samek, "Analyzing neuroimaging data through recurrent deep learning models," *Frontiers in Neuroscience*, vol. 13, Dec 2019.

[191] K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, C. Coffey, K. Kieburtz, E. Flagg, S. Chowdhury, and et al., "The Parkinson Progression Marker Initiative (PPMI)," *Progress in Neurobiology*, vol. 95, no. 4, p. 629–635, Dec 2011.

[192] [Online]. Available: https://www.ppmi-info.org/study-design

[193] 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, 2019. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2019.00279

[194] [Online]. Available: https://adni.loni.usc.edu/

[195] S. Whitfield-Gabrieli and A. Nieto-Castanon, "CONN: A functional connectivity toolbox for correlated and anticorrelated brain networks," *Brain Connectivity*, vol. 2, no. 3, p. 125–141, 2012.

[196] T. M. Inc., "MATLAB version: 9.13.0 (r2022b)," Natick, Massachusetts, United States, 2022. [Online]. Available: https://www.mathworks.com

[197] A. Nieto-Castanon, *Handbook of Functional Connectivity Magnetic Resonance Imaging Methods in CONN*, 2020.

[198] J. L. Andersson, C. Hutton, J. Ashburner, R. Turner, and K. Friston, "Modeling geometric deformations in EPI time series," *NeuroImage*, vol. 13, no. 5, pp. 903–919, 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811901907463

[199] K. J. Friston, J. Ashburner, C. D. Frith, J.-B. Poline, J. D. Heather, and R. S. Frackowiak, "Spatial registration and normalization of images," *Human Brain Mapping*, vol. 3, no. 3, p. 165–189, 1995.

[200] R. Henson, C. Buechel, O. Josephs, and K. Friston, "The slice-timing problem in event-related fMRI," *Henson, R.N.A. and Buechel, C. and Josephs, O. and Friston, K.J. (1999) The slice-timing problem in event-related fMRI. In: 5th International Conference on Functional Mapping of the Human Brain (HBM'99) and Educational Brain Mapping Course, June 22 - 26, 1999, Düsseldorf, Germany.*, vol. 9, 01 1999.

[201] R. Sladky, K. J. Friston, J. Tröstl, R. Cunnington, E. Moser, and C. Windischberger, "Slice-timing effects and their correction in functional MRI," *NeuroImage*, vol. 58, no. 2, pp. 588–594, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811911007245

[202] S. Whitfield-Gabrieli, A. Nieto-Castanon, and S. Ghosh, "Artifact detection tools (ART)," *Cambridge, MA. Release Version*, vol. 7, no. 19, p. 11, 2011.

[203] J. D. Power, A. Mitra, T. O. Laumann, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Methods to detect, characterize, and remove motion artifact in resting state fMRI," *NeuroImage*, vol. 84, p. 320–341, 2014.

[204] A. Nieto-Castanon, "Preparing fMRI data for statistical analysis," 2022.

[205] V. Calhoun, T. Wager, A. Krishnan, K. Rosch, K. Seymour, M. Nebel, S. Mostofsky, P. Nyalakanai, and K. Kiehl, "The impact of T1 versus EPI spatial normalization templates for fMRI data analyses," *Human Brain Mapping*, vol. 38, no. 11, p. 5331–5342, Jul 2017.

[206] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, no. 3, p. 839–851, Jul 2005.

[207] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage*, vol. 38, no. 1, pp. 95–113, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811907005848

[208] K. J. Friston, S. Williams, R. Howard, R. S. Frackowiak, and R. Turner, "Movement-related effects in fMRI time-series," *Magnetic Resonance in Medicine*, vol. 35, no. 3, p. 346–355, Mar 1996.

[209] M. N. Hallquist, K. Hwang, and B. Luna, "The nuisance of nuisance regression: Spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures functional connectivity," *NeuroImage*, vol. 82, p. 208–225, Nov 2013.

[210] Y. Behzadi, K. Restom, J. Liau, and T. T. Liu, "A component based noise correction method (CompCor) for BOLD and perfusion based fMRI," *NeuroImage*, vol. 37, no. 1, p. 90–101, Aug 2007.

[211] X. J. Chai, A. N. Castañón, D. Öngür, and S. Whitfield-Gabrieli, "Anticorrelations in resting state networks without global signal regression," *NeuroImage*, vol. 59, no. 2, p. 1420–1428, Jan 2012.

[212] F. Chollet *et al.*, "Keras," https://github.com/fchollet/keras, 2015.

[213] R. J. Meszlényi, K. Buza, and Z. Vidnyánszky, "Resting state fMRI functional connectivity-based classification using a convolutional neural network architecture," *Frontiers in Neuroinformatics*, vol. 11, 2017. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fninf.2017.00061

[214] F. Chollet, "The Keras blog." [Online]. Available: https://blog.keras.io/building-autoencoders-in-keras.html

[215] M. Alber, S. Lapuschkin, P. Seegerer, M. Hägele, K. T. Schütt, G. Montavon, W. Samek, K.-R. Müller, S. Dähne, and P.-J. Kindermans, "iNNvestigate neural networks!" *Journal of Machine Learning Research*, vol. 20, no. 93, pp. 1–8, 2019. [Online]. Available: http://jmlr.org/papers/v20/18-540.html

[216] W. Samek, A. Binder, G. Montavon, S. Bach, and K.-R. Müller, "Evaluating the visualization of what a deep neural network has learned," 2015.

[217] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin cancer classification using deep learning and transfer learning," in *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*, 2018, pp. 90–93.

[218] M. Ingalhalikar, S. Shinde, A. Karmarkar, A. Rajan, D. Rangaprakash, and G. Deshpande, "Functional connectivity-based prediction of autism on site harmonized ABIDE dataset," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 12, pp. 3628–3637, 2021.

[219] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the ABIDE dataset," *NeuroImage: Clinical*, vol. 17, pp. 16–23, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213158217302073

[220] S. Georgiades, P. Szatmari, and M. Boyle, "Importance of studying heterogeneity in autism," *Neuropsychiatry*, vol. 3, no. 2, p. 123, 2013.

[221] M. Gavrilescu, G. W. Stuart, S. Rossell, K. Henshall, C. McKay, A. A. Sergejew, D. Copolov, and G. F. Egan, "Functional connectivity estimation in fMRI data: Influence of preprocessing and time course selection," *Human Brain Mapping*, vol. 29, no. 9, p. 1040–1052, Sep 2008.

[222] S. Mutasa, S. Sun, and R. Ha, "Understanding artificial intelligence based radiology studies: What is overfitting?" *Clinical Imaging*, vol. 65, pp. 96–99, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0899707120301376

[223] S. Y. Ho, K. Phua, L. Wong, and W. W. Bin Goh, "Extensions of the external validation for checking learned model interpretability and generalizability," *Patterns*, vol. 1, no. 8, p. 100129, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666389920301707

[224] B. Crosson, A. Ford, K. M. McGregor, M. Meinzer, S. Cheshkov, X. Li, D. Walker-Batson, and R. W. Briggs, "Functional imaging and related techniques: An introduction for rehabilitation researchers," *The Journal of Rehabilitation Research and Development*, vol. 47, no. 2, p. vii, 2010.

[225] R. Leech and J. Smallwood, "Chapter 5 - the posterior cingulate cortex: Insights from structure and function," in *Cingulate Cortex*, ser. Handbook of Clinical Neurology, B. A. Vogt, Ed. Elsevier, 2019, vol. 166, pp. 73–85. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780444641960000054

[226] B. A. Vogt, "Cingulate cortex in Parkinson's disease," *Cingulate Cortex*, p. 253–266, 2019.

[227] D. Euston, A. Gruber, and B. McNaughton, "The role of medial prefrontal cortex in memory and decision making," *Neuron*, vol. 76, no. 6, p. 1057–1070, Dec 2012.

[228] P. Xu, A. Chen, Y. Li, X. Xing, and H. Lu, "Medial prefrontal cortex in neurological diseases," *Physiological Genomics*, vol. 51, no. 9, p. 432–442, Sep 2019.

[229] M. Banwinkler, H. Theis, S. Prange, and T. van Eimeren, "Imaging the limbic system in Parkinson's disease—a review of limbic pathology and clinical symptoms," *Brain Sciences*, vol. 12, no. 9, p. 1248, Sep 2022.

[230] S. D. Vann, J. P. Aggleton, and E. A. Maguire, "What does the retrosplenial cortex do?" *Nature Reviews Neuroscience*, vol. 10, no. 11, p. 792–802, Oct 2009.

[231] R. Dan, F. Růžička, O. Bezdicek, J. Roth, E. Růžička, J. Vymazal, G. Goelman, and R. Jech, "Impact of dopamine and cognitive impairment on neural reactivity to facial emotion in Parkinson's disease," *European Neuropsychopharmacology*, vol. 29, no. 11, p. 1258–1272, Nov 2019.

[232] V. Cutsuridis and M. Yoshida, "Editorial: Memory processes in medial temporal lobe: Experimental, theoretical and computational approaches," *Frontiers in Systems Neuroscience*, vol. 11, 2017. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnsys.2017.00019

[233] C. W. Tam, E. J. Burton, I. G. McKeith, D. J. Burn, and J. T. O'Brien, "Temporal lobe atrophy on MRI in Parkinson disease with dementia," *Neurology*, vol. 64, no. 5, p. 861–865, Mar 2005.

[234] S. J. Coen, A. R. Hobson, and Q. Aziz, "Chapter 23 - processing of gastrointestinal sensory signals in the brain," in *Physiology of the Gastrointestinal Tract (Fifth Edition)*, fifth edition ed., L. R. Johnson, F. K. Ghishan, J. D. Kaunitz, J. L. Merchant, H. M. Said, and J. D. Wood, Eds. Boston: Academic Press, 2012, pp. 689–702. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123820266000233

[235] M. Criaud, L. Christopher, P. Boulinguez, B. Ballanger, A. E. Lang, S. S. Cho, S. Houle, and A. P. Strafella, "Contribution of insula in Parkinson's disease: A quantitative meta-analysis study," *Human Brain Mapping*, vol. 37, no. 4, p. 1375–1392, Jan 2016.

Appendix A

Scientific Background

A.1 Brain Connectivity and Functional Magnetic Resonance Imaging

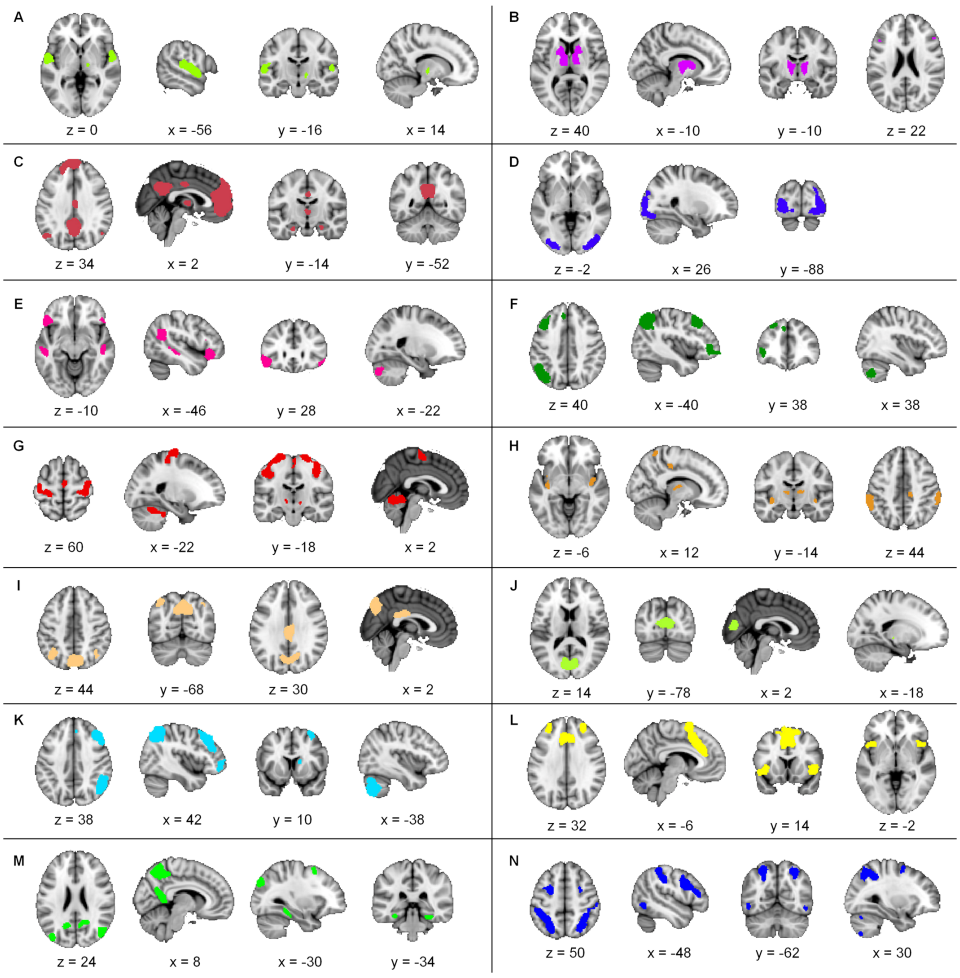

Figure A.1: Identified resting-state networks by Shirer et al.: A. Auditory; B. Basal Ganglia; C. Dorsal Default Mode; D. Secondary Visual Cortex (V2); E. Language; F. Left Executive Control; G. Sensorimotor; H. Posterior Salience; I. Precuneus; J. Primary Visual Cortex (V1); K. Right Executive Control; L. Anterior Salience Network; M. Ventral Default Mode; N. Visuospatial (from [83]).

Table A.1: Anatomic location of the functional networks identified by Shirer et al. (auditory, basal ganglia, dorsal default mode, V2, language networks). Adapted from [83].

Functional Network	Anatomical Location
Auditory	Heschl's Gyrus Left and Right Superior Temporal Gyrus Right Thalamus
Basal Ganglia	Left and Right Thalamus Caudate Putamen Left and Right Inferior Frontal Gyrus Pons
Dorsal Default Mode	Medial Prefrontal Cortex Anterior and Posterior Cingulate Cortex Orbifrontal Cortex Left and Right Angular Gyrus Precuneus Midcingulate Gyrus Left and Right Thalamus Left and Right Hippocampus
V2	Left and Right Middle Occipital Gyrus Superior Occipital Gyrus
Language	Inferior Frontal Gyrus Left Middle Temporal Gyrus Angular Gyrus Superior Temporal Gyrus Supramarginal Gyrus Right Inferior Frontal Gyrus Right Supramarginal Gyrus Middle Temporal Gyrus Left Crus I

Table A.2: Anatomic location of the functional networks identified by Shirer et al. (left executive control, sensorimotor, posterior salience, precuneus, V1). Adapted from [83].

Functional Network	Anatomical Location
Left Executive Control	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 and Right Precentral Gyrus Postcentral Gyrus Right Supplementary Motor Area Left Thalamus Bilateral Lobule IV, Lobule V, Lobule VI Right Thalamus
Posterior Salience	Left Middle Frontal Gyrus Left and Right Supramarginal Gyrus Inferior Parietal Gyrus Left Precuneus Right Midcingulate Cortex Right Superior Parietal Gyrus Precuneus Left and Right Thalamus Lobule IV, VI Left and Right Posterior Insula Putamen
Precuneus	Midcingulate Cortex Posterior Cingulate Cortex Precuneus Left and Right Angular Gyrus
V1	Calcarine Sulcus Left Thalamus

Table A.3: Anatomic location of the functional networks identified by Shirer et al. (right executive control, anterior salience, ventral default mode, visuospatial). Adapted from [83].

Functional Network	Anatomical Location
Right Executive Control	Right Superior Frontal Gyrus Right Middle Frontal Gyrus Right Inferior Frontal Gyrus Supramarginal Gyrus Angular Gyrus Left Crus I Crus II Lobule VI Right Caudate
Anterior Salience	Left and Right Middle Frontal Gyrus Left and Right Insula Anterior Cingulate Cortex Medial Prefrontal Cortex Supplementary Motor Area Left and Right Lobule VI Crus I
Ventral Default Mode	Left and Right Retrosplenial Cortex Posterior Cingulate Cortex Left Middle Frontal Gyrus Left and Right Parahippocampal Gyrus Left Middle Occipital Gyrus Precuneus Right Superior Frontal Gyrus Middle Frontal Gyrus Right Angular Gyrus Middle Occipital Gyrus Right Lobule IX
Visuospatial	Left and Right Middle Frontal Gyrus Superior and Inferior Frontal Gyrus Precentral Gyrus Left and Right Inferior Parietal Sulcus Left and Right Frontal Operculum Left Inferior Temporal Gyrus Inferior Frontal Gyrus Right Middle Temporal Gyrus Left and Right Lobule VIII Lobule VIIb Right Lobule VI Crus I