Anomaly characterization in the MRI data of 'de novo' Parkinson patients

Verónica Muñoz Ramírez

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : **BIS - Biotechnologie, instrumentation, signal et imagerie pour la biologie, la médecine et l'environnement**

Arrêté ministériel : 25 mai 2016

Caractérisation d'anomalies sur les données IRM des patients Parkinsoniens 'de novo'

Anomaly characterization in the MRI data of 'de novo' Parkinson's patients

Présentée par

Verónica MUÑOZ RAMÍREZ

Thèse dirigée par **Michel DOJAT**, Université Grenoble Alpes et codirigée par **Florence FORBES,** Université Grenoble Alpes **;**

préparée au sein du **Grenoble Institut de Neurosciences** dans **l'École Doctorale d'Ingénierie pour la Santé, la Cognition et l'Environnement.**

Thèse soutenue publiquement le **9 décembre 2020**, devant le jury composé de :

Monsieur Michel DOJAT
Directeur de Recherche, INSERM Délégation Alpes, Directeur de thèse

Madame Florence FORBES
Directrice de recherche, INRIA Centre de Grenoble, Co-directrice de thèse

Madame Meritxell BACH CUADRA
Maître de conférences, Université de Lausanne – Suisse, Rapporteuse

Monsieur Olivier COULON
Directeur de Recherche, CNRS – Aix-Marseille, Rapporteur

Monsieur Vincent VIGNERON
Maître de conférences, Université d'Evry / Paris-Saclay, Examinateur

Monsieur Philippe CINQUIN
Professeur des Universités, Université Grenoble Alpes, Examinateur, *Président*

Acknowledgements

I would like to extend my gratitude to anyone reading this manuscript. It is the product of three years of hard work where I got the chance to collaborate with brilliant doctors and scientists. It is an absolute honor to publish my thesis and hopefully be a source of information for somebody else.

...

Tout d'abord, j'aimerais remercier mes directeurs de thèse: Michel DOJAT et Florence FORBES. Merci de m'avoir permis d'effectuer ce doctorat. Merci de vos conséils et votre enthusiasme au cours ce ces trois denières années. Travailler à vos côtés fut un grand plaisir.

J'adresse également mes remerciements aux membres du jury, pour avoir bien voulu juger mon travail et consacrer du temps à la lecture de mon manuscrit. J'ai beaucoup apprécié nos échanges le jour de la soutenance.

Merci à tous les investigateurs de l'étude InnobioPark. Merci à Pr. Moro de mener ce beau projet et merci à tous les participants de l'étude pour votre contribution précieuse.

Je tiens aussi à remercier mes collègues de bureau : Fabien BOUX, Clément BROSSARD, Clément ACQUITTER et Chloé JAROSZYNSKI et tout l'équipe 5 du GIN/IRMaGe. Les bons moments passés ensemble ont rendue cette expérience d'autant plus belle.

Merci à mes copines, qui m'ont encouragé tout le long, une petite mention spéciale pour Anh-Dao VIRASSAMY et Ludivine VIGNARD.

Le plus grand merci au Dr. Thomas TJOUTIS. Merci d'être mon soutien. Il n'y a pas des mots pour exprimer toute la gratitude que je garde pour toi dans mon coeur. Merci d'être toi, merci d'être là.

...

Gracias a mis padres, como todo logro pasado y futuro, este doctorado es fruto de su esfuerzo. Los amo con todo mi corazón, gracias por dejarme volar tan alto y tan lejos... Siempre están en mi corazón.

Finalmente, gracias a mi hermano. Espero hacerte tan orgulloso como tú me haces sentir. Gracias por tus esfuerzos, gracias por acompañarme de lejos.

Contents

INTRODUCTION

On a common day, we are able to wake up, go to school or work, write a note, enjoy a meal, all without giving it a second thought. We cease our daily activities and after a good night of sleep we are prepared to do it all over again.

We seldom think about the 100 000 chemical reactions taking place in our brain every second, yet, every one of them is critical to the relay of information between the external world to our brain and from our brain to our body. When we eat for example, the visual system inspects the plate in front of us and determines how the fork should be grasped. This information is transmitted from the visual cortex to the frontal-lobe motor areas that plan and initiate the movement. The spinal cord carries these instructions to the motor neurons in the muscles of the arm and hand that proceed to pick up the fork with just the right amount of force and at a convenient speed thanks to the messages of the basal ganglia. Then, sensory receptors on the fingers send the signal that the fork is being held to the sensory cortex and finally to the motor cortex so that it can plan our next move. At the same time, we might be activating multiple other mechanisms by having a conversation, smelling our food or just by breathing. All within a split second.

In a complex network, neurons communicate with each other by releasing neurotransmitters, chemicals that act as messengers. Billions of nerve cells are arranged in specific patterns that coordinate thoughts, emotions, behaviors, movements and sensations, controlling everything from our heart rates to our mood.

The underproduction or overproduction of any of the 200 neurotransmitters produced by the brain alters the delicate balance necessary for proper functioning and can result in disease.

Parkinson's disease (PD) presents a clear example. The production of dopamine is progressively decreased due to damage in the substantia nigra, a basal ganglia structure located in the midbrain. This affects the inter-communication of the different structures in the basal ganglia but also, as dopamine levels drop, their interaction with the rest of the brain.

Neurodegenerative diseases like Parkinson's represent a public health preoccupation, indeed more than 6 million persons in the world live with this PD and due to the general aging of the population, the number of PD patients is expected to double by 2030.

PD is most well known for its motor symptoms, namely resting tremor, bradykinesia, rigidity and gait impairment. The damage to the substantia nigra affects movement control, for

this reason, patients may experience tremor and/or difficulty initiating voluntary movement.

The diagnosis of Parkinson's disease is made at the appearance of these clinical motor symptoms, however, when they occur, it has been estimated that 50% of dopamine producing neurons might have already been lost, dropping dopamine levels by 80%. Although there are effective symptomatic treatments, none of them are curative, so the patient's quality of life keeps on worsening over time, eventually leading to disability.

Reducing the gap between diagnosis and the onset of the neurodegenerative process is of paramount importance to identify treatments that would significantly slow its natural progression and this can only be achieved by studying the earliest possible pre-clinical stages of illness possible.

Indeed, non-motor symptoms have been found precede motor disturbances by up to a decade, indicating the presence of physiopathological changes from the onset of the disease. Some of the most common features include sleep and mood disorders, anxiety, depression, autonomic dysfunctions, visual and olfactory disturbances.

Neuroimaging plays a fundamental role in the diagnosis and assessment of neurodegenerative diseases. The combination of novel MRI (Magnetic Resonance Imaging) sequences and current developments in medical image analysis presents an excellent opportunity to study the complex physiopathology of Parkinson's disease. Multiple research studies have detected structural, functional and connectivity changes in newly diagnosed PD patients but heterogeneity in study cohorts, softwares and techniques have made it difficult to reach consensus. To this day there is no neuroimaging biomarkers that can diagnose PD with certainty.

In this context, the work presented in this thesis aims to detect and characterize abnormalities specific to the early stages of Parkinson's disease. For this purpose, structural and quantitative images have been studied through different medical imaging processing techniques going from classical computational methods to novel clustering and deep learning methods.

Motivation and strategy

We address the identification of early physiological changes as an anomaly detection problem. The goal of anomaly detection is to determine instances within a dataset that stand out as being dissimilar to all others [1, 2]. It is a very general approach that is employed in numerous fields, going from finance to health care.

Anomalies are also referred to as abnormalities, deviants or outliers in the data mining and statistics literature. They are not to be confused with novelties, which correspond to new or unobserved patterns in the data that can be added to the regular data model [2]. The

classification of a deviant point as a novelty or an anomaly is merely dependent on the abnormality threshold chosen. For this reason, the techniques used for novelty detection are often used for anomaly detection and vice versa.

There exists a plethora of novelty / anomaly detection methods, Pimentel & colleagues [3] classify them along five general categories:

1. **Probabilistic:** The density of the "normal" class is estimated and low density areas indicate a low probability of containing "normal" objects.
2. **Distance-based:** Includes the concepts of nearest-neighbor and clustering analysis. Novel data is assumed to be far from a tight "normal" cluster.
3. **Reconstruction-based:** A model is trained using only "normal" data. The mapping of "abnormal" data by the trained model is expected to have a high error with respect to the input. The model can be a regression model or a neural network such as an autoencoder.
4. **Domain-based:** A boundary around the "normal" class is defined, thus establishing the normal domain. Everything outside the domain is "abnormal".
5. **Information-theoretic techniques:** The information data is computed using information-theoretic measure. Novel data are supposed to alter significantly the information content of a dataset.

Each method is best suited to respond to a specific problem formulation. The survey presented by Chandola & colleagues [1] divides the problem formulation in four axes:

- **Nature of the data:** Univariate, multivariate, graph data, image data, etc.
- **Type of anomaly:** Point anomalies if every instance can be looked at individually, contextual if an instance is only considered as anomalous under a special context, or collective if several instances should be compared as a group to all others.
- **Degree of supervision:** Supervised if data labels are available for the normal and anomalous classes, semi-supervised if only normal labels are provided or unsupervised if no labels are necessary.
- **Output:** The anomalies can be reported as scores or binary labels.

MR imaging is able to extract information relevant to the structure and physiology of the brain. The comparison between the scans of healthy and recently diagnosed PD patients through anomaly detection techniques could bring to light relevant information about the changes that PD patients' brains undergo in the early stages of the disease.

In Part II of this manuscript, three very different unsupervised anomaly detection techniques are presented, applied and discussed.

The first one studies the morphometry of the brain and falls into the category of probabilistic approaches. The structural MR images of healthy individuals and PD patients are analyzed by group to measure differences in local concentrations of gray matter, detect signs of atrophy, cortical thinning or deformations of the cortical ribbon.

The second approach is reconstruction-based. Auto-encoder architectures are trained to reconstruct 2-D diffusion MR images of healthy subjects. The trained networks are then presented with both healthy and early PD diffusion MR slices and the reconstruction errors of the both populations are studied to find differences between the two.

Lastly, the third approach is distance-based. The quantitative multivariate MR data of healthy controls is first fitted by a mixture model of flexible distributions to constitute a reference model against which early PD data is positioned. Any instances with a low probability of being explained by the reference model are classified as outliers. Finally, the outliers are characterized as well by a second mixture model and their localization is investigated.

This multi-approach strategy has given the project an interesting point of view regarding the possible nature of anomalies in PD patients, the MR modalities that hold the most informative power, and the challenges faced ahead.

Challenges and contributions

The detection of anomalies at the early stages of Parkinson's disease is not an easy task. Changes at this phase are expected to be subtle and by consequence, not visible in routine MR scans. The use of novel MR contrasts and image analysis techniques is therefore necessary to study the physiopathology of this disease.

Since lesions are not apparent, all studies are exploratory and depend on the literature findings and clinical data to support their results. What is more, evidence suggests that there are different types of PD, so there is a possibility that abnormalities are not presented in the same regions or in the same manner across all patients.

Finally, we are exposed to common challenges in the medical field such as data availability and the heterogeneity in processing techniques and pipelines. This last point prevents the direct comparison of independent research studies. For example, when comparing local gray matter concentration, there are several techniques for brain tissue segmentation that can produce variable results. This is also true for the extraction of diffusion and perfusion parameter maps.

The three investigations presented in this manuscript employ state-of-the art processing methods that are easily available to the research community, also, techniques for results generalization were considered when possible.

To the best of our knowledge, we present the largest PD morphometry study investigating local gray matter changes, atrophy and cortical thinning. Bootstrap and family wise error correction were used to challenge the reproducibility of our results.

We found that no morphological differences and proposed quantitative MR data as a better alternative for the exploration of early PD biomarkers.

A pipeline was proposed for the use of auto-encoders as anomaly detection tools without any labels and applied it to diffusion images of PD patients. The pipeline could be easily adapted to other other types of medical images and/or pathologies.

Lastly, we adapted the mixture model pipeline specially designed for physiological data [4] to detect and characterize subtle anomalies on large datasets when no ground-truth is available.

Associated publications

In some part, the studies presented here have given place to several publications :

- Muñoz Ramírez V., Coupé P., Forbes F., Dojat M. (2019) « No structural differences are revealed using VBM in 'de novo' Parkinsonian patients ». Stud Health Technol Inform. 264: 268-272.
- Muñoz Ramírez V., Kmetzsch V., Forbes F., Dojat M. (2020) « Deep learning models to study the early stages of Parkinson Disease ». ISBI 2020 - IEEE International Symposium on Biomedical Imaging. Iowa City, USA.
- Muñoz Ramírez V., Kmetzsch V., Forbes F., Dojat, M. (2020) « Autoencoder-based anomaly detection for Parkinson's disease markers discovery ». **(in preparation)**
- Muñoz Ramírez V., Forbes F., Arbel J., Arnaud A., Dojat M. « Quantitative MRI characterization of brain abnormalities in de novo Parkinsonian patients ». ISBI 2019 - IEEE International Symposium on Biomedical Imaging. Venice, Italy.

They can be found as Annexes at the end of the manuscript.

Manuscript structure

In the first part of the manuscript "Medical and scientific context" a general overview of Parkinson's disease is presented in **Chapter 1**, including information about its incidence in the population, its physiopathology, symptoms, evolution and current treatment strategies. **Chapter 2** contains information about different structural and quantitative MRI techniques and how they have or could be harnessed for PD exploration. A state-of-the-art on the study

of Parkinson's Disease from MR imaging is included in **Chapter 3**, mainly focusing on techniques comparable to ours, notably studies on structural data, studies employing diffusion, perfusion or multivariate data and studies employing the outputs of unsupervised deep learning architectures.

In the second part of the manuscript "Anomalies in 'de novo' MR scans" three different approaches are introduced, applied and discussed. The first one (**Chapter 4**) is a comprehensive brain morphological analysis of Parkisonian patients. Local concentrations of gray matter, atrophy and cortical ribbon deformations are studied. The second one (**Chapter 5**) leverages the reconstruction errors produced by unsupervised deep learning architectures to discriminate between healthy and PD diffusion scans. The third method (**Chapter 6**) consists on clustering multivariate MR data with appropriate mixture distributions to differentiate between healthy and atypical instances at the voxel level. The notion of an MR multivariate signature is explored to discriminate between healthy subjects and PD patients.

Finally the employed methods and the physiological changes observed are discussed. We end the manuscript with a general discussion and perspectives for future work.

This introduction is directly followed by a summary of the manuscript in French.

Résumé en français

Le bon fonctionnement du cerveau dépend d'un équilibre délicat de plus de 200 neurotransmetteurs. La sous-production ou la surproduction de l'un de ces composants chimiques produits par le cerveau peut entraîner des troubles importants.

La maladie de Parkinson (MP) en présente un exemple clair. La production de dopamine est progressivement diminuée en raison de lésions de la substance noire, une structure des noyaux gris centraux située dans le mésencéphale. Cela affecte l'intercommunication des différentes structures dans les noyaux gris centraux mais aussi, à mesure que les niveaux de dopamine chutent, leur interaction avec le reste du cerveau.

Les maladies neurodégénératives comme la maladie de Parkinson représentent une préoccupation de santé publique, en effet plus de 6 millions de personnes dans le monde vivent avec cette maladie et en raison du vieillissement général de la population, le nombre de patients atteints de MP devrait doubler d'ici 2030.

La MP est surtout connue pour ses symptômes moteurs, à savoir les tremblements au repos, la bradykinésie, la rigidité et les troubles de la marche. Les dommages à la substance noire affectent le contrôle du mouvement, pour cette raison, les patients peuvent éprouver des tremblements et / ou des difficultés à initier un mouvement volontaire.

Le diagnostic de la maladie de Parkinson est posé à l'apparition de ces symptômes moteurs cliniques, cependant, lorsqu'ils surviennent, il a été estimé que 50% des neurones producteurs de dopamine auraient déjà été perdus, faisant chuter les niveaux de dopamine de 80%. Bien qu'il existe des traitements symptomatiques efficaces, aucun d'entre eux n'est curatif, de sorte que la qualité de vie du patient ne cesse de s'aggraver avec le temps, conduisant éventuellement à un handicap.

Réduire l'écart entre le diagnostic et le début du processus neurodégénératif est d'une importance capitale pour identifier les traitements qui ralentiraient considérablement sa progression naturelle et cela ne peut être réalisé qu'en étudiant les premiers stades précliniques possibles de la maladie.

En effet, des symptômes non moteurs ont été trouvés avant les troubles moteurs jusqu'à une décennie, indiquant la présence de changements physiopathologiques dès le début de la maladie. Certaines des caractéristiques les plus courantes comprennent les troubles du sommeil et de l'humeur, l'anxiété, la dépression, les dysfonctionnements autonomes, les troubles visuels et olfactifs.

La neuroimagerie joue un rôle fondamental dans le diagnostic et l'évaluation des maladies neurodégénératives. La combinaison de nouvelles séquences d'IRM (imagerie par résonance magnétique) et des développements actuels en analyse d'images médicales présente une excellente opportunité d'étudier la physiopathologie complexe de la maladie de Parkinson. De nombreuses études de recherche ont détecté des changements structurels, fonctionnels et de connectivité chez des patients atteints de MP nouvellement diagnostiqués, mais l'hétérogénéité des cohortes, des logiciels et des techniques d'étude a rendu difficile l'obtention d'un consensus. À ce jour, aucun biomarqueur de neuroimagerie ne permet de diagnostiquer la MP avec certitude.

Dans ce contexte, les travaux présentés dans cette thèse visent à détecter et caractériser des anomalies spécifiques aux stades précoces de la maladie de Parkinson. À cette fin, des images structurelles et quantitatives ont été étudiées à travers différentes techniques de traitement d'imagerie médicale allant des méthodes de calcul classiques aux nouvelles méthodes de clustering et d'apprentissage en profondeur.

Motivation et défis

Nous abordons l'identification des changements physiologiques précoces comme un problème de détection d'anomalies. Le but de la détection des anomalies est de déterminer les instances d'un ensemble de données qui se distinguent par leur différence avec toutes les autres [1, 2]. C'est une approche très générale qui est employée dans de nombreux domaines, allant de la finance à la santé.

L'imagerie par résonance magnétique est capable d'extraire des informations relatives à la structure et à la physiologie du cerveau. La comparaison entre les scans de patients atteints de MP sains et récemment diagnostiqués par des techniques de détection d'anomalies pourrait mettre en lumière des informations pertinentes sur les changements que le cerveau des patients atteints de MP subit aux premiers stades de la maladie.

Il existe une pléthore de méthodes de détection d'anomalies [3] (e.g. probabilistes, basées sur la distance, basées sur la reconstruction, basées sur le domaine) et chacune d'elles est la mieux adaptée pour répondre à une formulation de problème spécifique.

Dans ce manuscrit, trois études sont présentés. Ils utilisent des méthodes de traitement de pointe qui sont facilement accessibles à la communauté de recherche, ainsi que des techniques de généralisation des résultats ont été envisagées lorsque cela était possible.

Il est important de préciser que la détection d'anomalies aux premiers stades de la maladie de Parkinson n'est pas une tâche facile. On s'attend à ce que les changements à cette phase soient subtils et, par conséquent, non visibles dans les examens IRM de routine. L'utilisation de nouveaux contrastes IRM et de techniques d'analyse d'images est donc nécessaire pour étudier la physiopathologie de cette maladie.

Puisque les lésions ne sont pas apparentes, toutes les études sont exploratoires et dépendent des résultats de la littérature et des données cliniques pour appuyer leurs résultats. De plus, les preuves suggèrent qu'il existe différents types de MP, il est donc possible que les anomalies ne soient pas présentées dans les mêmes régions ou de la même manière chez tous les patients.

Enfin, nous sommes exposés à des défis communs dans le domaine médical tels que la disponibilité des données et l'hétérogénéité des techniques de traitement et des pipelines. Ce dernier point empêche la comparaison directe d'études de recherche indépendantes. Par exemple, lorsque l'on compare la concentration locale de matière grise, il existe plusieurs techniques de segmentation des tissus cérébraux qui peuvent produire des résultats variables. Ceci est également vrai pour l'extraction des cartes de paramètres de diffusion et de perfusion.

Contenu du manuscrit

Ce manuscrit est divisé en deux parties. La première partie, intitulée «Contexte médical et scientifique», est composée de trois chapitres. Un aperçu général de la maladie de Parkinson est présenté dans le **Chapitre 1**, y compris des informations sur son incidence dans la population, sa physiopathologie, ses symptômes, son évolution et les stratégies de traitement actuelles. **Chapitre 2** contient des informations sur les différentes techniques d'IRM structurelles et quantitatives et comment elles ont ou pourraient être exploitées pour l'exploration de la MP. Un état de l'art sur l'étude de la maladie de Parkinson à partir de l'imagerie par résonance magnétique est inclus dans **Chapitre 3**, principalement axé sur des techniques comparables aux nôtres, notamment des études sur des données structurelles, des études utilisant des données de diffusion, de perfusion ou multivariées et études utilisant les résultats d'architectures d'apprentissage en profondeur non supervisées.

Dans la deuxième partie du manuscrit, «Anomalies dans les données IRM des patients 'de novo'», trois approches non-supervisés sont présentées, appliquées et discutées. Un bref résumé de ces chapitres est présenté ci-dessous.

Étude des changements structurels (Chapitre 4)

Afin d'élucider la nature des différences morphologiques chez les patients Parkinsoniens de novo, on a réalisé une étude VBM (Morphométrie à base de voxel) à travers de deux pipelines: 1) la bien établie Computational Anatomy Toolbox (CAT12) (University of Jena) incluse dans la version actuelle du logiciel SPM (SPM12) et 2) une nouvelle plateforme en ligne: volBrain [5]. Les deux pipelines ont des atouts complémentaires qui sont exploités dans cette étude. volBrain effectue une segmentation de qualité des noyaux sous-corticaux

et CAT12 facilite l'analyse de groupe. En outre, nous avons recherché des différences quantitatives entre la classification des tissus effectuée par les deux approches, toutes deux incluant l'estimation du volume partiel.

Par la suite, des études DBM et SBM (Morphométrie à base de déformations et surface respectivement) ont été menées uniquement sur CAT12, car volBrain est un système axé sur la volumétrie.

La population de notre étude est composée 144 patients Parkinsoniens de novo (âge: 61,30 $\pm$ 9,06; sexe: 53 F, 91 H) et 66 témoins sains (âge: 60,12 $\pm$ 11,39; sexe: 23 F, 43 H) de la base de données PPMI [6] (www.ppmi-info.org/data). Nous avons pris un soin particulier à considérer une cohorte de sujets relativement importante, à considérer les effets d'un nombre déséquilibré de patients et de témoins et à corriger pour des comparaisons multiples.

Avec ces précautions, notre analyse VBM n'a pas trouvé de différences morphologiques chez les patients MP ($p < 0,05$ FWE), ni sur l'analyse du groupe de matière grise du cerveau entier ni sur l'analyse de plusieurs structures sous-corticales séparément. Les résultats des différences structurelles cérébrales rapportées dans la littérature ont tendance à être contradictoires, ce qui rend difficile la comparaison de nos résultats. Ce manque de consensus peut-être dû à une variété de facteurs, notamment l'étude des petites cohortes, les différences dans les techniques de segmentation et, plus important encore, le manque de correction pour les comparaisons multiples augmentant inévitablement les taux de faux positifs.

En ce qui concerne l'atrophie, notre étude DBM n'a pas établi de disparités significatives une fois la correction pour comparaisons multiples appliquée ($p < 0,05$ FWE). Nous notons qu'en omettant cette correction, plusieurs régions qui deviennent significatives font partie du réseau PD-ICA proposé par Zeighami et ses collègues [7] (e.g. amygdale, thalamus, hippocampe). Une raison possible pour les écarts entre nos résultats est due à la méthode employée, ils ont utilisé le pipeline CIVET et nous le pipeline CAT12, et il n'est pas clair si leur correction a été calculée pour chaque voxel dans l'ICA-ROI ou pour chaque ICA possible, auquel cas notre correction est beaucoup plus stricte.

Nos résultats de SBM coïncident avec ceux des dernières recherches sur les caractéristiques corticales des patients atteints de MP, il n'y a pas de différences significatives d'épaisseur corticale ni de gyrification chez les patients PD de novo. Cela confirme les capacités des méthodes CAT12 à estimer correctement le ruban cortical, et ce, en une fraction du temps par rapport à FreeSurfer. [8] a estimé 10-20 heures par sujet pour la reconstruction de la surface corticale sur FreeSurfer, alors que CAT12 effectue son traitement en moins d'une heure. À notre connaissance, l'étude SBM réalisée dans cette thèse est la plus importante de la littérature.

Cette étude renforce le message selon lequel les évaluations morphologiques sont des techniques délicates impliquant de nombreux paramètres qui doivent être manipulés avec précaution pour éviter que les faux positifs n'influencent les résultats finaux [9].

Au vu de l'absence de différences morphologiques, nous soupçonnons que les biomarqueurs précoces de la maladie de Parkinson reposent sur les propriétés physiologiques du cerveau parkinsonien et pourraient être étudiés par des techniques d'IRM quantitatives (relaxométrie, diffusion, perfusion).

Étude des changements de diffusion avec Neural Networks (Chapitre 5)

Dans ce chapitre, nous présentons un pipeline novateur pour la détection d'anomalies dans les données IRM de diffusion des patients récemment diagnostiqués avec la MP. Les auto-encodeurs entièrement convolutifs et les auto-encodeurs variationnels ont constitué les principaux éléments constitutifs de notre approche.

Les données utilisées dans ce travail proviennent également de la base de données PPMI [6]. Les scans IRM de diffusion de 56 contrôles sains et 129 patients atteints de DP *de novo* ont été regroupés et à partir de ces scans, deux mesures par voxel ont été calculées en utilisant MRtrix3.0 [10]: diffusivité moyenne (MD) et anisotropie fractionnelle (FA). Ces mesures décrivent la diffusion des molécules d'eau dans le cerveau, MD rend compte de leur déplacement global et FA indique l'orientation de la diffusion.

Même s'il est possible de former des auto-encodeurs avec des images 3D, 56 contrôles ne sont pas assez de données pour entraîner les modèles, d'autant plus qu'une partie des données doit être mise de côté pour les tests. C'est pourquoi nous avons décidé d'entrainer les auto-encodeurs avec des coupes axiales 2D de nos scans 3D. De plus, nous avons choisi de traiter chaque hémisphère comme un sujet différent.

L'ensemble de données de contrôle a été divisé en 41 contrôles de formation et 15 contrôles de test pour éviter les fuites de données. En conséquence, les modèles ont été entraînés avec un duo (FA & MD) de 1680 images (2 hémisphères $\times 40$ tranches $\times 41$ contrôles). D'autre part, l'ensemble de données des patients testés est constitué d'un duo de 10320 instances et l'ensemble de données des contrôles de test comprend un duo de 1200 images.

Nos résultats montrent que les reconstructions cérébrales des patients atteints de MP sont significativement différentes de celles des témoins sains pour les modèles sAE, sVAE et dVAE. Dans notre application, il n'y a pas de vérité terrain permettant de privilégier un modèle par rapport aux autres, néanmoins, les applications d'apprentissage supervisé des auto-encodeurs dans la littérature ont montré que les architectures dVAE ont du mal à reconstruire les détails fins dans les scans cérébraux, par exemple, les gyrifications [11].

Une procédure d'évaluation originale a été conçue pour comparer les profils d'erreur de reconstruction d'individus sains par rapport aux patients atteints de MP. La procédure était basée sur l'hypothèse que les reconstructions d'hémisphères de patients devraient présenter une plus grande quantité de points mal reconstruits et ainsi nous avons comparé les pourcentages de voxels anormaux dans les hémisphères cérébraux de nos sujets. En définissant le pourcentage qui présente le meilleur compromis entre sensibilité et spécificité comme

seuil pathologique, nous avons transformé notre dilemme en un problème de classification binaire dont nous pouvons calculer les performances discriminantes et prédictives. Cette procédure a été suivie pour tout l'hémisphère cérébral mais aussi pour des régions d'intérêt spécifiques telles que les lobes corticaux de matière grise, les structures sous-corticales et les structures de matière blanche.

Avec des ensembles de données déséquilibrés, nous avons pris un soin particulier à utiliser une métrique adaptée telle que la moyenne géométrique entre spécificité et sensibilité (g-mean), mais aussi à généraliser nos résultats en effectuant une validation croisée par 10 pour former et tester nos résultats. Cela s'est avéré être un élément clé de l'étude. En effet, sans raison évidente, la première sous-population réalise de bien meilleures performances que la moyenne de nos dix populations. A titre indicatif, le score g-mean sAE associé à l'hémisphère cérébral est de 74,3% alors que la moyenne est de 64,6%.

Étant donné que le DTI a été conçue pour étudier les caractéristiques de la substance blanche, il n'est pas surprenant que certaines des meilleures performances aient été obtenues par la substance blanche, cependant, en observant le seuil pathologique de la sAE déduit pour les petites structures sous-corticales comme la substantia nigra (20,96 %) et le STN (19,35 %) nous soupçonnons que l'auto-encodeur spatial était incapable de produire des reconstructions précises de ces régions.

Comme dans d'autres enquêtes, nous constatons que les patients présentant une déficience cognitive légère sont nettement plus faciles à classer que les patients conformes sur le plan cognitif en utilisant tout l'hémisphère cérébral, la combinaison de lobes de matière grise ou de substance blanche.

D'un point de vue concurrentiel, notre approche a obtenu des performances similaires à celles d'autres études de diffusion. Notamment, la procédure de validation croisée de Schuff et ses collègues [12] a obtenu une ROC AUC de 59% pour le segment rostral du SN qui est comparable à notre AUC ROC de 57,6% pour le SN complet. Ensuite, Correia et ses collègues [13] ont obtenu un score de précision moyen de 59,7% grâce à une procédure de validation croisée sur leur SVM et cette sélection de régions de substance blanche. Ceci est conforme à notre score de prédiction g-mean moyen de 61% pour 11 régions de matière blanche.

Dans leur ensemble, ces résultats offrent des preuves convaincantes que les modèles basés sur l'apprentissage profond sont utiles pour localiser les anomalies subtiles, comme on les trouve chez les patients Parkinsoniens *de novo*, même lorsqu'ils sont entrainés avec un nombre modéré d'images et uniquement des mesures FA et MD. De tels modèles pourraient être intéressants pour étudier d'autres troubles neurologiques lorsque de petites lésions sont suspectées et difficiles à localiser pour un observateur humain. On peut s'attendre à ce que l'insertion de mesures d'IRM quantitatives supplémentaires, telles que la perfusion, la teneur en fer et le temps de relaxation des tissus, améliorerait la détection des anomalies. De plus, la localisation spatiale d'altérations subtiles des modalités d'imagerie RM, sensibles à

différents paramètres physiologiques, pourrait apporter de nouvelles connaissances sur la physiopathologie de la maladie sous-jacente.

Caractérisation des données multivariées MR à l'aide de modèles de mélange (Chapitre 6)

Dans ce travail, nous proposons de tirer parti des fonctionnalités informatives fournies par l'IRM quantitative pour construire des modèles statistiques représentant des tissus cérébraux sains. Cela nous permet de détecter des valeurs atypiques pour ces caractéristiques dans le cerveau des patients parkinsoniens. Nous introduisons des modèles de mélange pour capturer la forme non standard de la distribution multivariée des données.

Les modèles de mélange sont des modèles probabilistes qui représentent des sous-populations au sein d'une population globale. Ils ne nécessitent pas de connaissance préalable de la sous-population à laquelle appartient un point de données, ce qui permet au modèle d'apprendre automatiquement les sous-populations de manière non supervisée.

Bien que les distributions gaussiennes soient les distributions statistiques les plus largement utilisées pour leur traitabilité et leur pouvoir de représentation, elles sont contraintes par des formes elliptiques. Les mélanges de gaussiens peuvent aider à modéliser des formes distributionnelles plus riches, mais ils ne sont toujours pas appropriés lorsque les données présentent des sous-groupes allongés et fortement non elliptiques. Comme alternative, nous considérons une famille de distributions plus riche basée sur des distributions de Student à échelles multiples (MSD). Ces distributions à queue lourde sont dotées de quantités marginales variables de poids de queue et leurs mélanges se sont avérés fournir une alternative efficace aux mélanges gaussiens [14, 4].

Dans l'approche proposée, les données quantitatives multivariées IRM des contrôles sains sont d'abord ajustées par mélange MSD pour constituer un modèle de référence par rapport auquel les données de PD précoces sont positionnées. Toutes les instances avec une faible probabilité d'être expliquées par le modèle de référence sont classées comme des valeurs aberrantes. Enfin, les valeurs aberrantes sont également caractérisées par un deuxième modèle de mélange et leur localisation est étudiée.

Three different applications of MSD mixture models are presented in this study:

- **Application 1**
  Données: Deux cartes paramétriques de diffusion (FA & MD) ainsi qu'une carte de flux sanguin cérébral (CBF) associées à 3 sujets sains (C1-C3) (Age: $\{28, 40, 50\}$; Sex: 2H, 1F) et 10 sujets Parkinsoniens (P1-P10) (Age $\in [36, 66]$; H&Y score ≤ 2; Sex: 7H, 3F). Seulement les valeurs se trouvant à l'intérieur des structures sous-corticales ont été considérées dans les modèles.
  Résultats: Le modèle de réference à 6 classes a été retenu. Malgré le fait que le modèle de référence a été estimé sans aucune information spatiale, le modèle de référence

est cohérent d'un point de vue anatomique et respecte la symétrie. 21.6% des voxels ont été classifiés comme atypiques, résultat sur un modèle à 4 classes. L'un des clusters, principalement situé sur la substantia nigra, le noyau rouge et le noyau sous-thalamique, est caractérisé par de faibles valeurs de CBF et de FA élevées, ce qui est conforme aux résultats de Fernandez-Seara & al. [15]. Un autre groupe dans le striatum coïncide avec des valeurs de DM plus élevées rapportées par Peran & al. [16]. De toute évidence, plus de sujets sont nécessaires pour mettre en évidence des biomarqueurs robustes de la MP. Cependant, ces résultats préliminaires montrent que l'application de modèles de mélange de distributions pertinentes est informative et prometteuse pour discriminer correctement la pathologie.

- **Application 2**
  Données: Cartes FA et MD (diffusion) issues de 56 sujets sains (Age: $\{61.1 \pm 9.8\}$; Sex: 34H, 23F) et 129 patients *de novo* (Age: $\{61.74 \pm 8.96\}$; Sex: 80H, 49F)
  Résultats: L'application de modèles MMST pour la détection d'anomalies de diffusion sur un ensemble de données considérablement volumineux (7,2 millions de voxels pour les sujets témoins et 35,9 millions de voxels pour les patients) a produit des résultats non concluants. La plupart des anomalies étaient localisées sur le corps calleux. Il s'agit de la plus grande structure de matière blanche du cerveau, constituée d'un faisceau plat de fibres commissionnelles reliant les hémisphères gauche et droit. L'implication de cette structure dans la maladie de Parkinson et a été liée au déclin cognitif cite Bledsoe2018, pourtant, une étude de cas récente rapportant un patient atteint de MP sans corps calleux (agénésie complète) cite Kho2019 suggère que les changements dégénératifs bilatéraux dans La MP peut survenir indépendamment de l'état du corps calleux. Il est fort probable que les données de diffusion à elles seules ne soient pas suffisantes pour distinguer les patients atteints de MP des témoins sains, du moins aux premiers stades de la maladie.

- **Application 3**
  Données: Nous avons 8 cartes paramétriques pour 2 sujets sains (Ages: 56,58, Sex: 2F) et 12 sujets Parkinsoniens de novo (Age: 62.9 $\pm$ 7.9, Sex: 6H, 6F, H&Y=2, MoCA>26). Ces cartes ont été extraites des scans de diffustion, perfusion (pCASL et DSC), relaxométrie (T1, T2*) acquises lors du protocole InnobioPark du CHU Grenoble Alpes. Les séquences avec un taux de crrélation au dessus de 0.5 aux autres paramètres. Les cartes de FA, pCBF, MTT, T1 et T2* ont été retenues.
  Résultats: Ces résultats sont à interpréter avec délicatesse car le modèle de référence n'a pas été tiré d'un échantillon représentatif de la population, les données disponibles appartiennent à deux témoins femelles de 58 et 56 ans.
  Néanmoins, les résultats préliminaires rapportés ici sont encourageants et plausibles au regard de la littérature. Dans l'ensemble de l'analyse du cerveau, les anomalies

caractérisées par des temps de relaxation T1 accrus ont été trouvées sur les structures sous-corticales, ce qui pourrait indiquer des changements microstructuraux dans les tissus [17].
Les temps de transit moyens sur les deux analyses étaient clairement plus élevés pour les patients atteints de MP, une explication pourrait être une diminution de la pression de perfusion cérébrale qui est compensée par la vasodilatation et produit à son tour une réduction du CBF et un MTT prolongé (théorie de Monro-Kellie).
Dans l'ensemble, la combinaison des temps de diffusion, de perfusion et de relaxation tissulaire semble être d'autant plus informative que les mesures séparées.

Les modèles MMST se sont avérés bien adaptés pour caractériser des tissus sains et pathologiques sur la base de paramètres physiologiques. Le pipeline développé pour la détection des anomalies a produit des résultats cohérents par rapport à la littérature dans les cas où plus d'une modalité RM était utilisée. En effet, aux stades précoces de la maladie de Parkinson, les anomalies sont susceptibles d'être subtiles et les propriétés de diffusion à elles seules semblent insuffisantes pour caractériser les patients atteints de MP.

Des développements informatiques sont en cours pour optimiser l'estimation des modèles MMST sur de grands ensembles de données.

Discussion

Trois méthodes très différentes pour étudier différents ensembles de données ont été explorées dans cette thèse.Cette stratégie multi-approches a donné au projet un point de vue intéressant sur la nature possible des anomalies chez les patients atteints de MP, les modalités de RM qui détiennent le pouvoir le plus informatif et la défis à relever.

Au meilleur de nos connaissances, nous présentons la plus grande étude de morphométrie PD portant sur les changements locaux de la matière grise, l'atrophie et l'amincissement cortical. Le bootstrap et la correction d'erreur au niveau de la famille ont été utilisés pour contester la reproductibilité de nos résultats.

Nous avons constaté qu'aucune différence morphologique et proposé des données quantitatives MR comme une meilleure alternative pour l'exploration des premiers biomarqueurs PD.

Un pipeline a été proposé pour l'utilisation d'auto-encodeurs comme outils de détection d'anomalies sans aucune étiquette et l'a appliqué à des images de diffusion de patients atteints de MP. Le pipeline pourrait être facilement adapté à d'autres types d'images médicales et / ou de pathologies.

Enfin, nous avons adapté le pipeline de modèles de mélanges spécialement conçu pour les données physiologiques cite Arnaud2017 pour détecter et caractériser des anomalies subtiles sur de grands ensembles de données lorsqu'aucune vérité terrain n'est disponible.

Nous pensons que les données quantitatives sont une alternative plus appropriée pour étudier la maladie de Parkinson, en particulier sur les patients «de novo» où tout changement devrait être subtil.

En résumé, nous avons trouvé des preuves d'anomalies de la substance blanche dans les premiers stades de la MP, mais des recherches supplémentaires devraient se concentrer sur la caractérisation de ces différences chez les patients sans déficience cognitive. De plus, la combinaison des paramètres de diffusion avec d'autres propriétés physiologiques devrait améliorer la classification. Bien qu'aucune étude indépendante de la perfusion n'ait été réalisée, l'intégration des paramètres de perfusion dans notre approche de détection d'anomalies basée sur le clustering a donné une indication claire de son potentiel à classer les patients atteints de MP des témoins, même aux premiers stades de la maladie.

De plus, son interaction avec d'autres propriétés physiologiques devrait être bénéfique pour les performances de classification.

Dans d'autres travaux, nous étudierons les changements de perfusion sur des cohortes plus importantes. En outre, la normalisation des cartes de paramètres relatifs DSC (rCBV, rCBF) devrait être explorée.

Pour conclure, l'intégration de mesures quantitatives de RM reflétant les multiples propriétés du cerveau est certainement essentielle pour détecter les anomalies dans le cerveau des patients atteints de MP dès les premiers stades possibles. Les mesures de diffusion, de perfusion et de relaxométrie peuvent brosser un tableau informatif de la physiologie normale du cerveau. De plus, les modèles de mélange et les auto-encodeurs sont d'excellents candidats pour détecter les anomalies liées à la physiopathologie de la MP à partir de ces données.

Part I

MEDICAL & SCIENTIFIC CONTEXT

Chapter 1

Parkinson's Disease Overview

Parkinson's disease (PD) is an ever-present neurodegenerative condition with an estimated global prevalence of 10 million cases [6]. It was first described by James Parkinson in 1817 as a movement disorder where patients have: *"Involuntary tremulous motion, with lessened muscular power, in parts not in action and even when supported; with a propensity to bend the trunk forward, and to pass from a walking to a running pace: the senses and intellects being uninjured"(Parkinson, 1817)* [18].

These features were later defined as the cardinal motor symptoms of PD: resting tremor, bradykinesia, rigidity, and gait impairment. They progress gradually over the years and their manifestation constitutes the main diagnostic tool for modern-day neurologists.

In the last two hundred years, important elements have been added to the clinical picture of PD. In the mid-1800, Jean-Martin Charcot refined the description of the disease, differentiating it from other disorders characterized by tremor [18], specifically from multiple sclerosis. Later on, he and his students defined atypical occurrences of classical Parkinson's disease that would later be termed as Parkinsonism-plus syndromes, such as progressive supranuclear palsy (PSP), corticobasal degeneration (CBD) and multiple system atrophy (MSA).

In this day and age it has been recognized that a multitude of non-motor features, such as cognitive impairment, autonomic dysfunction, disorders of sleep, depression, visual and olfactory disturbances, are part of the disease and add considerably to overall burden [19].

Although several treatments and symptom management techniques are available for PD patients, none stop the development of the disease, and so, while mortality has been drastically reduced (from a 3:1 ratio of deaths to expected deaths to a 1.52:1 ratio), the number of years lived with disability due to PD increased between 1990 and 2010 [20]. Furthermore, a progressive increase in the personal, societal and economic burden associated with the disease is expected in the future as the world population ages [21].

1.1 Epidemiology

PD is a progressive, adult-onset disease, and it gets more common with age. It is in fact the second most common neurodegenerative disease affecting the elderly, only behind Alzheimer's Disease.

PD incidence is usually comprised between 10 and 50 per 100,000 persons per year, and its prevalence between 100 and 300/100,000 population [22]. It is rare before age 50 years, and its incidence and prevalence both increase progressively after age 60.

Most studies on life expectancy found a significantly higher risk of mortality in patients with PD compared to the general population. Taking in consideration age and sex differences, their increased mortality risk ranges between 1.26 to 3.79 [22].

PD is usually more common in men than women, with a male-to-female ratio ranging from 1.3 to 2.0 [23]. Compared to men, women tend to have later age at disease onset; lower prevalence and incidence; higher rates of tremor phenotype; and a greater likelihood of dyskinesia, and motor and non-motor fluctuations.

Pre-clinical evidence has suggested a potential neuro-protective effect of oestrogen against dopaminergic damage through anti-inflammatory, anti-oxidative mechanisms. In support of this theory, it has been found that PD symptoms in women tend to worsen when oestogen levels are low (before menses) and are progressively alleviated as the levels rise (during ovulation) [23].

Most of the time, there's no known cause, but in a few cases, there might be a genetic explanation, like mutations in the PINK1, parkin, or alpha synuclein genes, indeed 15% of PD patients report a family history of PD among their first-degree relatives [22]. In rare cases, Parkinsonian symptoms may be caused by MPTP, a toxic impurity that can be found in the recreational drug MPPP or desmethylprodine, which is a synthetic opioid. In other people, one or more risk factors, rather than a single outright cause, might contribute to Parkinson's, for example pesticide exposure or DNA variants in genes like LRRK2.

In common with other complex traits, the pathogenesis in the large majority of cases of PD is expected to be multifactorial, involving a combination of genetic and environmental risk factors [24].

1.2 Physiopathology

While the specific pathophysiology underlying all of the symptoms observed in PD is not well understood, it has been clearly recognized that the main biochemical feature of PD is the progressive degeneration of dopamine-producing neurons in the substantia nigra pars compacta [21].

The dopaminergic neurons of the substantia nigra are projected onto the striatum. Both structures form part of the basal ganglia, a group of gray-matter structures situated at the base of the forebrain and top of the midbrain. The largest structure in the basal ganglia is the striatum which brings together the caudate nucleus, the putamen and the ventral striatum. The sub-thalamic nucleus (STN), the internal and external segments of the globus pallidus (GPi, GPe) and the substantia nigra (SN) are also included in the basal ganglia. The latter can be divided into pars compata (SNc) and pars reticulata (SNr).

In healthy subjects, the control of physical movements as well as important cognitive and limbic functions rely upon a series of parallel corticostriatal pathways that connect the basal ganglia with regions of the cerebral cortex. Dopamine plays a key role in the regulation of these functional circuits thus the loss of this neurotransmitter is likely to massively disrupt the basal ganglia circuitry, triggering a great part of the afflictions manifested by PD patients.

The motor pathway is one of the best documented examples of this disruption. In healthy individuals, circuits that originate from the motor and pre-motor cortical areas project to the striatum (putamen) through excitatory glutamatergic neurons. These striatal neurons use gamma-aminobutyric acid (GABA) as their primary neurotransmitter and are organised into two pathways that converge on the GPi and the SNr, the major ouput nuclei of the basal ganglia.

- The *direct pathway* connects the striatum to the GPi/SNr without deviation. At rest, neurons from the GPi and SNr project to the thalamus and maintain a steady release of the neurotransmitter GABA which acts to inhibit the thalamic neurons and keeps unwanted movements from occurring. However, when movement is desired, glutamate neurons from the corticostriatal pathway excite striatal neurons that release GABA in the GPi and the SNr, inhibiting the activity of these regions and stopping the inhibition of neurons in the thalamus that stimulates the premotor cortex and activates the muscles.
- The *indirect pathway* connects the striatum to the GPi/SNr by intermediary of the GPe and the STN. At rest, GABA neurons from the GPe exert an inhibitory effect on glutamate neurons in the STN, but when the indirect pathway is activated, GABA neurons in the striatum project to the GPe and inhibit the activity of neurons there, keeping them from being able to inhibit neurons in the STN. Simultaneausly, the STN neurons are activated by projections from the cortex, and they stimulate GABA neurons in the GPi and SNr. These GABA neurons in turn project to the thalamus, inhibiting thalamic neurons that travel to motor regions of the cerebral cortex to stimulate movement. The inhibition of these thalamic neurons thus inhibits movement.

The direct pathway promotes voluntary movement in targeted muscles while the indirect pathway simultaneously inhibits the movement in other muscles that do not contribute to the overall wanted movement the result is a coordinated smooth movement in which

those muscles necessary for the desired movement are recruited while the other muscles that might affect the flow off of the decided movement are inhibited.

The SNc is thought to modulate the activity of both pathways through dopamine release in the striatum targeting D1 dopamine receptors to further activate the direct pathway and D2 receptors to reduce the inhibitory effects of the indirect pathway. Both actions facilitate movement and it is believed the dynamic balance of dopaminergic stimulation exerted via both pathways regulates the amount of movement undertaken.

This model of the motor loop provides a physiopathological explanation of the akinetic features experienced by PD patients. The loss of D1 dopamine stimulation on striatal neurons reduces the degree of inhibition exerted in the GPi, leading to its over-activation. Likewise, the loss of inhibitory D2 stimulation of the striatal fibres in the indirect pathway leads to over-activation of the STN and consequent over-activation of the output nuclei of the basal ganglia. This unregulated activation of the GPi/SNr leads to an inhibition of the thalamus, which in turn impairs both ascending and descending pathways, resulting in the reduction of motor activity.

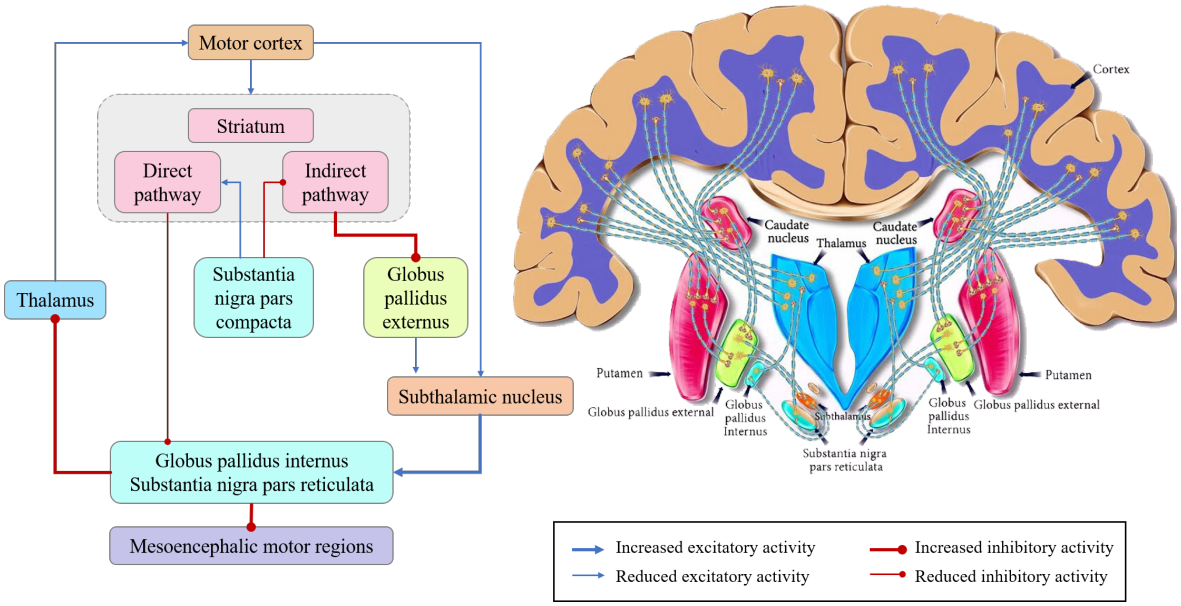


FIGURE 1.1: Motor cortex excitatory and inhibitory changes caused by PD.

Alongside the thalamus, the pedunculopontine nucleus (PPN) receives projections from the GPi and the SNr. This major brainstem nucleus is implicated in the process of locomotion and is likely the site that activates central pattern generators within the spinal cord. Recent studies suggest that, in PD patients, cellular loss in the PPN is independent from the cell loss in the nigrostriatal pathway which could explain gait disturbances and why such symptoms are not alleviated by dopaminergic treatment.

Indeed, cellular loss is not exclusive to neurons that express dopamine. Cholinergic, serotonergic and noradrenergic structures are impacted as well and doubtless contribute to several

disease features including cognitive, autonomic and affective disturbances.

Another important characteritic of Parkison's disease is the presence of Lewy bodies in neuronf the substantia nigra before they die. Lewy bodies are eosinophilic, round inclusions made of alpha-synuclein protein.

The function of alpha-synuclein is unknown, as well as the significance of Lewy bodies, and they are both found in other diseases like Lewy body dementia and multiple system atrophy.

1.3 Symptoms and evolution

Therefore, when substantia nigra pars compacta neurons die, the individual may be in a hypokinetic or low movement state which is commonly seen in Parkinson's. In addition to simply initiating movements, the substantia nigra helps to calibrate and fine tune the way that movements happen, which leads to the clinical features of Parkinson's.

First there's tremor, which is an involuntary shakiness most noticeable in the hands, characteristically called a "pill-rolling" tremor because it looks like someone rolling a pill between their thumb and index finger. This is a "resting tremor," meaning it is present at rest and diminishes with intentional movement.

Then there's rigidity, which refers to stiffness that can appear as "cogwheel" rigidity, which is when there are a series of catches or stalls as a person's arms or legs are passively moved by someone else. Rigidity is also responsible for the stooped posture and an almost expressionless face that some individuals with Parkinson's might have.

The patients can also experience difficulty initiating movements, this can induce bradykinesia, slow movement; hypokinesia, lessened movement or akinesia, the absence of movement. Examples of this are having the legs freeze up when trying to walk and also walking with a shuffling gait, or small steps.

Finally, a late feature of the disease is postural instability which causes problems with balance and can lead to falls. Despite these multiple effects on movement, Parkinson's Disease does not produce weakness. This helps differentiate it from diseases that affect the motor cortex or corticospinal pathway. In addition, the resting tremor of Parkinson's Disease helps to differentiate it from cerebellar diseases, which might result in an action or intention tremor, which is a tremor that's essentially the opposite of a resting tremor, where the tremor actually gets worse with movement.

Also, both bradykinesia and postural instability help differentiate Parkinson's from essential tremor of which an action tremor is also a hallmark feature.

In addition to the before-mentioned motor symptoms PD is also accompanied by numerous non-motor consequences that are already identifiable from the pre-motor stages of the

pathology. One of the most well-documented non-motor symptom of PD is olfactory loss [25]. Olfaction can be assessed with the Pennsylvania Smell Identification test and aid the diagnosis of pre-clinical and podromal subjects as the scores in smell discrimination have been robustly correlated to the risk of developing PD [26].

Autonomic dysfunctions such as constipation, urinary dysfunction, erectile dysfunction and orthostatic hypotension are eventually experienced by the majority of patients with PD. However, most autonomic symptoms are common in the general population and in other illness, notably Multiple System Atrophy (MSA) [25], making their specificity relatively low. Be as it may, a combination of autonomic dysfunction and another one of PD non-motor symptoms like RBD (Rapid Eye Movement Sleep Behavior Disorder) can accurately predict PD with a pronounced long lead time going up to 15 years [25].

RBD is characterized by the loss of the natural paralysis that occurs during REM sleep. Just over 1% of the general population over 50 has RBD which makes it by far the most specific/predictive marker [25]. There are other PD-related symptoms like visual disturbances, hallucinations, cognitive disorders, anxiety and depression, yet, the most substantial challenge for physicians is to recognize the non-motor profile of PD in order to proceed to further testing and eventually diagnosis.

The variety of non-motor symptoms of PD would indicate a multi-system neuronal degeneration underlying PD and while it clearly affects the quality of life of PD patients, it also multiplies the plausible biomarkers to characterize the evolution of PD.

Being a progressive disease, various measures have been used to outline disease severity at different stages in its course. Introduced in 1967, the Hoehn and Yahr (H&Y) scale has become the most commonly and widely used scale to estimate the severity of PD [27]. The scale consists on a 5 point staging assessment that evaluates the severity of over-all parkinsonism dysfunction based on bilateral motor involvement and the compromise of gait and balance.

- *Stage I*. Unilateral involvement only, usually with minimal or no functional impairment.
- *Stage II*. Bilateral or midline involvement, without impairment of balance.
- *Stage III*. First sign of impaired righting reflexes. This is evident by unsteadiness as the patient turns or is demonstrated when he is pushed from standing equilibrium with the feet together and eyes closed. Functionally the patient is somewhat restricted in his activities but may have some work potential depending upon the type of employment. Patients are physically capable of leading independent lives, and their disability is mild to moderate.
- *Stage IV*. Fully developed, severely disabling disease; the patient is still able to walk and stand unassisted but is markedly incapacitated.

- *Stage V*. Confinement to bed or wheelchair unless aided. [28]

Given its historic stature, the scale has been used as a gold standard for the testing of newly developed scales and good correlations have been established with standard PD ratings scales, e.g., Unified Parkinson's Disease Rating Scale (UPDRS), quality of life scales, and functional imaging studies [27].

While in recent years, several efforts have been made to study patients at risk of developing PD based on the emergence of non-motor symptoms, otherwise called *podromal* cohorts [29, 30, 25], there is a real challenge to find a balance between cohorts that are truly representative of idiopathic Parkinson's as a whole and cohorts that are sufficiently enriched for at-risk individuals to enable adequate numbers of positive results to be obtained [29, 31].

General consensus establishes the progression to stage III as a major milestone in the natural history of the illness [31, 19], hence, a cohort of patients in stages I and II of the disease, also named *de novo* patients, increases the probability of including different phenotypes of PD and should allow a deeper understanding of the most important non-motor symptoms.

1.4 Management

Fortunately, there are treatments that help with Parkinson's symptoms, although none stop the progressive neurodegeneration.

Most current therapies available focus on **pharmacological dopamine substitution** (l-DOPA treatment) [21]. Dopamine itself can't cross the blood-brain barrier, but its precursor levodopa can, and once in the brain, levodopa is converted to dopamine by dopa decarboxylase within the remaining nigrostriatal neurons.

Another strategy is using amantadine, which is also an antiviral medication that increases endogenous dopamine production, the complete interaction mechanisms are still being investigated.

A different strategy is to use dopamine agonists that can stimulate dopamine receptors and basically trick the brain into thinking there's more dopamine than there really is, like bromocriptine, which is an ergot or fungal derivative, as well as pramipexole and ropinirole, which are not ergot derivatives.

On occasions, inhibitors of COMT, catecholamine-O-methyltransferase, are prescribed with levadopa. COMT is an enzyme that degrades dopamine and levadopa. Inhibiting COMT prevents the enzyme from breaking levadopa down outside the central nervous system, thus allowing more of it to enter the brain. Very similarly, there are medications like selegiline which inhibits monoamine oxidase B, also known as MAO-B, which is another enzyme that metabolizes dopamine.

Since usually there's this balance of signaling between dopamine and acetylcholine, a loss of dopamine reaching the striatum increases the relative amount of acetylcholine signalling there. Therefore anticholinergics can be given to restore the balance of cholinergic and dopaminergic signaling, like benztropine, which improves the tremor of PD.

Besides pharmacological treatments, some patients can benefit from **deep brain stimulation (DBS)** therapy. This ground-breaking treatment involves an implantable device that directly sends electrical signals to the basal ganglia which counteracts the aberrant signaling in Parkinson's.

Good candidates are patients with PD with disabling motor fluctuations and/or medically intractable tremor without significant cognitive or psychiatric problems [32].

Significant improvements in quality of life and motor function have been obtained in the short-term and long-term with both subthalamic nucleus (STN) and globus pallidus internus (GPi) deep brain stimulation (DBS) in Parkinson's disease (PD). Both structures appear to be equally effective targets for treating motor symptoms with decreased benefits after 6 years. We note that while STN-DBS often allows for greater reduction in medication, it may be associated with worsening of non-motor symptoms and falls compared to GPi-DBS [33].

PD afflicts patients for as many as one to two decades of their lives and current treatments can only attenuate some motor manifestations [27]. Therefore, reducing the gap between diagnosis and the onset of the neurodegenerative process is of paramount importance to identify treatments that would significantly slow its natural progression and this can only be achieved by studying the earliest pre-clinical stages of illness possible [19].

Chapter 2

MR imaging as an exploration tool

2.1 Generalities

Neuroimaging plays a pivotal role in the research of PD by providing an *in vivo* opportunity to visualize the neuroanatomical and functional signatures of the disease, to aid differential diagnosis, to monitor disease progression and to measure the effects and complications of new therapies [34, 35].

MRI holds a special advantage over PET and SPECT imaging as it is non-invasive and more widely available than these nuclear neuroimaging techniques [16]. MRI uses hydrogen protons, which are abundant in the human body. All protons spin creating a small magnetic charge. When a strong magnetic field (B0) is applied, as is the case in an MRI machine, the protons align with that field. Radio-frequency pulses are then introduced to disrupt the protons (excitation), forcing them into a 90^o or 180^o realignment with the static magnetic field. Since the radio-frequency pulse pushed the proton against its nature, once this pulse is turned off, the protons realign with the magnetic field (relaxation), releasing electromagnetic energy along the way. The MRI is able to detect this energy and differentiate various tissues based on how quickly they release energy after the pulse is turned off.

Protons realign at different speeds and release different amounts of energy depending on the environment and the chemical nature of the molecules. These magnetic properties translate into different contrasts for different types of tissues, for example, gray and white matter. There exists about a hundred MR sequences, each one consisting on a subtle combination of radio-frequency pulses and magnetic field gradients that promote the signal of one tissue above another.

MRI has very much benefited from the technological advances of the last three decades, significantly improving the quality and rapidity of acquisitions, but also introducing the notion of quantitative MRI [36]. While qualitative analysis of MR images by physicians continues to be the norm, MR is increasingly used as a scientific instrument to make measurements of clinically relevant quantities, such as iron content, mean diffusivity of water and cerebral blood volume to name a few.

For many years, neuroimaging studies of PD were centered on the substantia nigra, but the recent technical and methodological advances have revolutionized the search with more than a dozen MRI-based techniques serving the characterization of physiological properties of the Parkinsonian brain beyond the dopaminergic system [35, 37].

2.2 Qualitative Imaging

As mentioned before, the design of an MR sequence affects the contrast of the images produced. The response of protons to particular excitations is tested to highlight different tissue classes. This allows physicians to visualize changes in **the structure of the brain** but also to detect abnormalities such as tumors and bleeding.

T1-weighted images differentiate tissues by their longitudinal relaxation times. Fat quickly realigns its longitudinal magnetization with B0, thus, it appears bright on a T1-weighted image. Conversely, water has a much slower longitudinal magnetization realignment and therefore appears dark.

T1-weighted sequences are used in automatic segmentation techniques like voxel-based, deformation-based and tensor-based morphometry. These methods measure differences in local concentrations of brain tissue and atrophic changes between two populations. T1-weighted sequences are also used to measure thickness and other characteristics of the cortical ribbon. Multiple studies have investigated morphometric changes in PD patients, a literature review is presented in the next chapter.

Conventional T1-weighted images present poor contrast in many structures of interest to the study of PD, like the substantia nigra, the subthalamic nucleus, the globus pallidus and the red nucleus [38]. These regions are better observed with sequences that are sensitive to iron-load such as T2-weighted, T2*-weighted and novel SWI (Susceptibility-weighted Imaging).

T2-weighted images differentiate tissues by their transverse relaxation times. In these scans, fat has an intermediate brightness while fluid appears bright. T2*-weighted images account not only for the transverse relaxation time of protons stimulated by a radio-frequency pulse, but also the effects of magnetic inhomogeneities.

When two tissue regions have different magnetic susceptibility, a spatially varying magnetic field is induced in the space surrounding the tissue interface or structure, resulting in T2* decay. As a result, paramagnetic elements such as iron and gadolinium have hypointense signals in T2-weighted that are further enhanced in T2*-weighted images. SWI scans use phase image information about local susceptibility changes between tissues to attenuate paramagnetic signals even more [39].

Researchers have employed T2, T2* and SWI sequences to study iron load in the substantia nigra [38] and specific T1-weighted sequences have been designed to visualize neuromelanin in the SN as well [31]. Neuromelanin is an iron chelator that could be linked to alpha-synuclein aggregation.

Alternative sequences currently being explored for the study of PD include proton density, short tau inversion recovery (STIR) and magnetic transfer (MT). Additionally, the increase in magnetic field provided by 7T human MRI, improved the visualization of basal ganglia contours and shapes with an increased spatial resolution and contrast [38, 29].

2.3 Quantitative Imaging

Quantitative MR data is considered, in particular, to be as independent as possible of the MRI scanner or the study center [36]. The scanner is no longer considered only as a camera but as a means of measurement.

Compared to qualitative analysis, quantitative analysis allows a direct interpretation of physiological parameters. This simplifies population comparisons, longitudinal analysis and interpretation of detected changes. What is more, multiple measures from different MR modalities can be combined to provide a complete picture of functional and structural changes caused by PD.

2.3.1 Relaxometry

The term ***relaxometry*** refers to the quantitative measurement of T1, T2 and T2* relaxation times. Instead of manipulating the acquisition parameters to obtain differences in contrast, the relaxation times are measured as biophysical properties of the tissue [36], describing the microstructure of the tissue. These measures are expected to be reproducible and thus allow comparisons across different scanners and subjects. They are displayed as maps in which every voxel value corresponds to a relaxation time, commonly expressed in milliseconds.

2.3.1.1 T1 mapping

The gold standard for T1 determination is to perform a series of independent single-point inversion recovery (IR) signal measurements at different inversion times (TI) [36, 17]. T1 can be calculated by mathematical fitting to the following equation :

$$S(TI) = S_o(1 - 2e^{-TI/T1} \tag{2.1}$$

where $S(TI)$ is the signal measured at time TI and S_o is the signal acquired at longitudinal magnetization equilibrium (M_o).

The single-point IR method is very time consuming (around 20 minutes), therefore other T1 mapping strategies have been devised.

The Look and Locker approach continuously samples the recovering magnetization with a series of small angle RF pulses. Thus, the T1 recovery curve is fully characterized following a single inversion pulse. Due to disturbing nature of the RF pulses, the magnetization recovery follows an apparent.

Another alternative is the method of variable flip angles. Several gradient echo sequences that combine low flip angle (θ) RF excitation with short repetition times (TR) (i.e. FLASH, SPGR) are acquired over a range of flip angles whilst the repetition time is held constant. T1 is calculated by resolving the following equation:

$$S = S_o \frac{1 - e^{-TR/T1}}{1 - cos\theta . e^{-TR/T1}} sin\theta \tag{2.2}$$

We note that several optimizations of the presented methods have been proposed in the literature [17].

2.3.1.2 T2/T2* mapping

T2 and T2*-mapping are typically performed using a balanced gradient echo sequence. Signals are sampled at several different echo times (TE) and fitted to an exponential decay curve, which in its simplest form can be described by:

$$S = S_o e^{-TE/T2} \tag{2.3}$$

...

Some of the most promising applications of relaxometry in the study of Parkinson's Disease are the estimation of iron load in subcortical structures which has been found to be increased in the Parkinsonian brain and the correlation of the average relaxation rate R2* (1/T2*) in the substantia nigra with the UPDRS motor score and/or the levodopa daily dose scores [16, 38].

2.3.2 Diffusion

Diffusion imaging monitors the displacement of water molecules in the brain. As they move, the molecules interact with macromolecules, proteins fibers and membranes. These obstacles vary depending on the tissue hence the rate of water diffusion gives indirect information on the structural composition of the brain at that particular location [40].

Generally three types of displacement can be observed: free diffusion, where water molecules can move in all directions (e.g. CSF); restricted isotrophic, where movements are limited by

obstacles but all directions are possible; and restricted anisotrophic, where movement is restricted in some directions and simplified in others, like in myelinated axons.

Diffusion sequences generally detect mobility differences by applying a initial the 180^o RF impulsion followed by two strong magnetic gradients in opposite directions, both symmetric to the RF current. If there is no displacement of water molecules between the two gradient applications, the molecules spins will be dephased by the first gradient and perfectly rephased by the second one; however, if the molecules move they will not recover the same phase with the second gradient, therefore signal loss will be detected.

To evaluate the overall displacement of molecules in the cartesian plane, these gradients are applied in the x, y and z planes. The signal from each of the three diffusion gradient planes is compared to a baseline T2-weighted images ($b = 0$). Free diffusion is identified when any of the three gradient directions displays signal loss compared to baseline, conversely restricted diffusion occurs when signal is preserved in all directions. In diffusion-weigthed images, areas of restrictive diffusion are bright.

The magnitude of diffusion is quantified through ADC (Apparent diffusion coefficient) maps that are no longer dependant on T2 signal. ADC is measured in s/mm^2, its calculation necessitates two diffusion images at known b, commonly $b_0 = 0s/mm^2$ and $b_1 = 1000s/mm^2$. The following equation is resolved:

$$ADC(x, y, z) = \frac{ln\frac{S_1(x,y,z)}{S_0(x,y,z)}}{b_0 - b_1} \tag{2.4}$$

ADC maps highlight changes in the mobility of water molecules without taking into account the direction they take.

Diffusion Tensor Imaging (DTI) characterizes the direction of displacement. DTI requires the acquisition of at least six diffusion weighted images, each obtained with a different orientation of the diffusion gradients. For every voxel diffusion is represented as an ellipsoid which characteristics are expressed in form of a tensor matrix. The axes of the ellipsoid are called the eigenvectors and the measures of their lengths eigenvalues.

Mean diffusivity (MD) is a sca

Fractional anisotropy (FA) is a scalar value ranging between 0 and 1 that describes the degree of anisotropy in water displacements. A value of zero corresponds to free diffusion and a value of one indicates that diffusion only occurs along one axis, probably a white matter fiber.

FA can be calculated as:

$$FA = \sqrt{\frac{1}{2}}\frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \tag{2.5}$$

where $\lambda_1, \lambda_2, \lambda_3$ correspond to the eigenvalues of the diffusion tensor.

Other measures of DTI include axial diffusion (AD) which is the value of the main diffusion direction (λ_1) and radial diffusivity (RD) defined as $(\lambda_2 + \lambda_3)/2$.

The particular micro structure of nerve fibers is at the origin of increased diffusion along the direction of white matter fibers and restricted diffusion in the perpendicular direction. The great anistrophy observed can be exploited to reconstruct tractography maps through a post-processing pipeline.

The use of strong gradients in diffusion imaging induces specific artifacts that can be corrected in post-processing. Great heterogeneity in diffusion acquisition parameters and post-processing pipelines hinders the reproducibility of diffusion studies. PD studies are no exception, as a matter of fact the literature review presented on the next chapter presents inconsistencies in the characterization of diffusion in the basal ganglia of PD patients.

2.3.3 Perfusion

Perfusion is physiologically defined as the steady-state delivery of blood to an element of tissue. This fundamental biological function is closely related to the arrival of oxygen and nutrients to tissue by means of blood flow [36]. In MRI, there exists two major approaches to measure cerebral perfusion: those with and without the use of an exogenous, intravascular, contrast agent.

Gadolinium-based contrast agents are the most common. In the presence of Gadolinium, the T1 and T2 relaxation times shorten, thereby causing reduced signal intensity on T2 and T2*-weighted images or increased signal intensity on T1-weighted images. **Dynamic susceptibility contrast (DSC)** MR perfusion exploits the regional susceptibility-induced signal loss caused by paramagnetic contrast agents in T2-weighted images. A bolus of gadolinium-containing contrast is injected intravenously and rapid repeated imaging of the tissue (most commonly brain) is performed during the first pass. This leads to a series of images with the signal in each voxel representing intrinsic tissue T2/T2* signal attenuated by susceptibility-induced signal loss proportional to the amount of contrast primarily in the microvasculature. Then a region's signal is interrogated over the time-course of the perfusion sequence, and a signal intensity-time curve is generated, from which various parameters can be calculated. We can name relative cerebral blood flow (rCBF), relative cerebral blood volume (rCBV) and mean transit time (MTT).

Although this technique can be performed with both T2 (e.g. spin echo) and T2* (e.g. gradient-echo echo-planar) sequences, the former requires higher doses of contrast, which is why T2* techniques are more commonly employed.

DSC relies upon detecting signal loss due to small amounts of contrast. One of its main pitfalls is the lack of reliable results in the presence of calcification of blood products that

generate a significant signal loss. This technique is also sensitive to artifacts from adjacent dense bone or aerated sinuses. Similarly, values in a region immediately adjacent to large vessels can also be affected.

Dynamic contrast-enhanced (DCE) MR perfusion, sometimes also referred to as permeability MRI, calculates perfusion parameters by evaluating T1 shortening induced by a gadolinium-based contrast bolus passing through tissue. The most commonly calculated parameter is k-trans.

DCE relies upon the T1-shortening effects of gadolinium-based contrast agents. An intravenous contrast bolus is injected and rapid repeated T1 imaging is obtained. Regional increased signal (T1 shortening) is due to gadolinium concentration which in turn will depend on a number of factors: intravascular gadolinium (i.e. true perfusion), and accumulation of gadolinium in the extravascular space (i.e. permeability).

Several local parameters can be extracted from DCE scans, notably, the k-transfer constant, the rate constant, the fractional volume of extravascular-extracellular space and the fractional volume of the plasma space.

More recently, the development of **arterial spin labelled (ASL)** MRI has enabled brain perfusion to be assessed non-invasively by tracing endogenous arterial blood water that is labelled electromagnetically using radiofrequency (RF) irradiation.

A number of magnetic labelling techniques have been proposed for ASL perfusion. We can count pulsed (PASL), continuous (CASL), pseudocontinuous (PCASL) and velocity-selective ASL (VS-ASL). The same general principle is applied to all these methods to quantify perfusion. Two images are acquired, a control image and a labeled image. The control image is subtracted from the labeled image to eliminate static signals and isolate the magnetization signals that are proportional to cerebral blood flow (CBF).

Bolus methods, with injections of a contrast agent, provide better sensitivity with higher spatial resolution, and are therefore more widely used in clinical applications. However, arterial spin-labeling methods provide a unique opportunity to measure cerebral blood flow without requiring an exogenous contrast agent and have better accuracy for quantification.

2.3.4 Functional

Lastly, ***functional MRI (fMRI)*** is nowadays widely used to study functional connectivity brain regions. Well known sequences include blood oxygen level-dependent (BOLD) fMRI and resting state fMRI (rs-fMRI).

Its application to study PD has suggested a remapping of cerebral connectivity resulting from dopamine depletion. This predominantly affects the sensorimotor circuit targeted by levodopa and differently associated with motor and non-motor symptoms [38].

Structural imaging (qualitative)	Characterizes brain morphology. • T1-w: gray matter concentration comparison, atrophy detection and cortical ribbon characterization. • T2-w, T2*-w & SWI: basal ganglia delineation and visualization of iron accumulation.
Relaxometry	Characterizes tissues microstructure • T1-mapping: Micro-structure characterization • T2, T2*-mapping: Estimate iron-load concentration. • MTR: Transfer of energy between highly bound protons, myelination and axonal density.
Diffusion imaging	Estimates the overall displacement of water molecules and their orientation. • ADC: Overall displacement of water. • FA: Orientation of diffusion. • AD: Characterize diffusion along its main direction. • RD: Characterize diffusion on the perpendicular direction.
Brain perfusion	Tracing of electromagnetically labeled arterial blood. • DSC (Gd) - rCBF, rCBV, MTT and others • DCE (Gd) - fractional volume of extravascular-extracellular space and of the plasma space • ASL - CBF
Functional MRI	Functional connectivity of brain regions. - *BOLD* - Blood oxygenated level-dependent f-MRI - *rs-fMRI* - Resting state f-MRI

TABLE 2.1: MRI techniques and their role in the characterization of Parkinson's Disease. *Adapted from Pyatigorskaya and colleagues [38].*

Chapter 3

State-of-the art in the study of PD from MR data

Although routine MR scans are generally considered as normal in PD, the quest for a prognosis biomarker continues. Indeed, the broad availability of MR imaging and the standardization of acquisition parameters would confer a potential for widespread application to any such biomarker. For this reason numerous methodological improvements have been explored to build robust imaging biomarkers.

3.1 Studies on structural data

3.1.1 VBM: Volume-based morphometry studies

The loss of dopaminergic neurons (amongst others) in specific regions of the brain logically induces their progressive shrinkage. That being said, currently there is no recognized biomarker of morphological changes in the brain at the early stages of PD.

In the literature, several studies have reported structural brain differences in PD patients compared to controls. However, these findings tend to be contradictory. Using locally gathered databases, Summerfield and colleagues detected gray matter loss on the right hipocampus, the left anterior cingulate region and the left superior temporal gyrus (p=0.001 uncorrected) in PD patients (n= 13) compared to controls (n=13) [41]. Nyberg and colleagues found an augmentation in the volume of the hipocampus (p=0.03 uncorrected) of PD patients (n=21) and shape deformations of the right accumbens nucleus (p=0.005 uncorrected) compared to controls (n=20) [42]. Radziunas and colleagues observed that PD patients (n=28) with sleep disturbances had bigger ventricules and smaller hipocampus (p-FDR<0.05) than healthy controls (n=28) [43]. Using an globally available database called PPMI, Jia and colleagues noted gray matter losses (p-FWE<0.001) in the fronto-parietal areas and the caudate nucleus, as well as an increase in the size of the limbic and paralimbic areas, the globus pallidus and the putamen of PD patients (n=89) versus controls (n=55) using SPM8 [44].

Also using the PPMI, this time for a structural limbic gray matter analysis, Li and colleagues from the University of Nottingham [45], assessed the gray matter densities and cross-sectional estimates of age-related gray matter change from 366 PD patients and 172 healthy controls. Brain structural differences were compared via voxel-based morphometry (VBM) then, structural brain networks were obtained using covariance analysis seeded in regions showing gray matter abnormalities in the PD subject group, notably the amygdala. Last, local age-related gray matter density change was estimated in subjects with Parkinson's and compared with those estimated in healthy controls. They found gray matter deficits in the right amygdala and reduced cerebellar but increased temporal connectivity. In addition, there was an increase of age-related gray matter intensity changes in the right limbic and paralimbic system.

The latter study supports the early affection of the limbic network in Parkinson's and a role in autonomic dysfunction and early cognitive impairment.

Authors	Features	Dataset	Findings
Summerfield & al. (2005)	VBM	13 HC 13 PD	GM loss on the R hippocampus, L anterior cingulate and L superior temporal gyrus (p=0.001 uncorr.)
Nyberg & al. (2015)	VBM	20 HC 21 PD	↗ hippocampus (p=0.03 uncorr.) and shape deformations of R accumbens (p=0.005 uncorr.)
Radzunias & al. (2018)	VBM	20 HC 28 PD with sleep disturbances	Bigger ventricules and smaller hippocampus (p-FDR<0.05)
Jia & al. (2015)	VBM	55 HC 89 PD *PPMI*	↘ GM in the fronto-parietal and caudate ↗ limbic and paralimbic, GP and putamen (p-FWE<0.001)
Li & al. (2017)	VBM	172 HC 366 PD *PPMI*	GM ↘ R amygdala GM ↗ R limbic and paralimbic

TABLE 3.1: Summarized findings of morphological volumetric PD biomarker studies

3.1.2 DBM: Deformation-based morphometry studies

Aside from in size, changes in shape (i.e. deformations) of specific brain structures can constitute more sensitive types of morphometric biomarkers. Deformation-based morphometry studies have been carried out in several studies with somewhat inconsistent results.

Borghammer and colleagues [46] investigated brain deformations in a group of 24 early stage PD patients versus 26 age-matched control subjects and found a significant contraction in the left cerebellum of PD patients. They also observed a high correlation between the unified PD rating scores and the local expansions in or near sulci bordering on frontal and temporal cortex.

Later, Tessa and colleagues [47] effectuated a longitudinal analysis where they measured the atrophic changes of 22 de novo PD patients without cognitive impairment and 17 age-matched control subjects at baseline and at a three-year follow-up. At baseline, no significant differences were found between controls and patients but the longitudinal evaluation revealed yearly atrophy rates in regions mainly related to cognitive function, notably the prefrontal cortex, the anterior cingulum, the head of the caudate nucleus and the thalamus. This study highlighted the importance of differentiating between cognitively intact, PD-MCI (Mild Cognitive Impairment) besides demented patients.

Zeighami and colleagues [7], combined deformation-based morphometry (DBM) and tensor probabilistic independent component analysis (ICA) to identify brain regions demonstrating atrophy in early PD. They analyzed 232 early PD patients versus 117 age-matched controls and found significantly negative differences in DBM values (p=0.003 Bonferroni corrected) in one and only one of 30 estimated components. This component, subsequently named the PD-ICA network, included all components of the basal ganglia (substantia nigra, subthalamic nucleus, nucleus accumbens, putamen, caudate nucleus, and internal and external globus pallidus), the pedunculopontine nucleus, basal forebrain, including bed nucleus of the stria terminalis and an area containing the nucleus basalis of Meynert, the hypothalamus, amygdala, hippocampus, parahippocampal gyrus, and two thalamic regions, the ventrolateral nucleus and pulvinar. Cortical regions in this network are the insula, occipital cortex Brodmann area 19, superior temporal gyrus, rostral anterior cingulate cortex, premotor and supplementary areas, and parts of lateral prefrontal cortex. These regions correspond to a normal brain network thus support the idea of a network-spread mechanism in PD.

On a follow-up study [48], they studied the correlation between atrophy in the PD-ICA network and several clinical measurements following a Partial Least Squares approach. Without controlling for it, age has the biggest impact for atrophy. After removing the age effect, highlighted features include gender (males worse), memory-specific cognitive impairment, RBD and certain affective behavioral scores.

On a another follow-up study [49], Zeighami and colleagues investigated whether this PD-network atrophy pattern could predict the rate of progression of motor and non-motor symptoms. The PD-network biomarker outperformed UPDRS, SPECT and PIGD scores as a prognostic biomarker with a modest ROC-AUC of 63% on average. Even though its predictive power is very limited and thus not a viable biomarker in clinic the authors propose it as a better alternative to other biomarkers during clinical trials.

3.1.3 SBM: Surface-based morphometry studies

The characterization of the brain's cortical ribbon has been useful to detect cortical changes in patients with Alzheimer's disease and Huntington's disease but a robust biomarker of cortical thickness or complexity is yet to be elucidated for Parkinson's disease.

Authors	Features	Dataset	Findings
Borghammer & al. (2010)	DBM	26 HC 24 early PD	↘ L cerbellum ↗ frontal and temporal cortex
Tessa & al. (2014)	DBM	17 HC 22 early PD (longitudinal)	Atrophy over time on the prefrontal cortex, anterior cingulum, caudate and thalamus often related to cognitive decline.
Zeighami & al. (2015)	DBM + ICA	117 HC 232 early PD *PPMI*	Definition of the PD-ICA network of atrophy (p=0.003 corrected) including the basal ganglia, the forebrain, the hippocampus, parahippocampus and other neighboring regions.
Zeighami & al. (2017)	DBM + PLS	117 HC 229 early PD *PPMI*	Important correlations between the PD-ICA network and the age, sex, cognitive impairment, RBD and affective behavioral scores.
Zeighami & al. (2019)	DBM	57 HC 105 early PD *PPMI* (longitudinal)	The PD-ICA network provides a better progression prediction than other clinical test scores with a ROC-AUC=63%.

TABLE 3.2: Summarized findings of morphological deformation PD biomarker studies.

Tinaz and colleagues [50] reported cortical thinning in the oribtofrontal cortex, the ventrolateral prefrontal cortex and the occipito-parietal areas of a de novo PD patient group made up of fifteen volunteers when compared to fifteen age-matched healthy individuals. Their images were obtained at 1.5T and the surface reconstruction method employed appeared to be in-house developed.

Later, Pereira and colleagues [51] assessed cortical degeneration in 20 PD patients at several stages of the disease versus 20 HC with FreeSurfer. They found widespread cortical thinning in the left lateral occipital cortex that extended to inferior and superior parietal areas, as well as in the right hemisphere, concretely in the inferior parietal cortex, extending into various regions of the lateral occipital, supramarginal, inferior, middle, and superior temporal cortex, but also in the right frontal cortex, comprising the pars opercularis, triangularis, precentral, and postcentral areas in PD patients. In their research, cortical thickness correlated with disease progression.

Ibarretxe-Bilbao and colleagues [52] studied the progression of cortical thinning in early PD. They followed 16 PD patients (H&Y ≤ 2) and 16 age-matched controls after 35 months. While they failed to find relevant differences at baseline, they discovered a greater rate of cortical thinning in PD patients in the right superior frontal gyrus, extending to caudal middle frontal gyrus, precentral sulcus, and precentral gyrus; right frontal pars opercularis and precentral gyrus; right superior temporal gyrus extending to the adjacent temporal sulcus; left caudal middle frontal gyrus extending to rostral middle frontal gyrus and precentral region; and left middle temporal gyrus extending to parietal cortex.

These last two studies concluded that surface-based analysis may be more sensitive to morphological changes on PD.

Studying a larger population, Malek and colleagues [24] imaged 105 newly diagnosed PD patients (39 MCI and 66 with no cognitive impairment) and 37 HC at baseline and 18 months later. They reconstructed their cortical surfaces using FreeSurfer. There were no significant differences ($p < 0.05$ FWE) between the cognitively compliant patients and the controls at baseline but a significantly greater percentage of cortical thinning in the caudal middle frontal cortex at the 18-month mark. In the case of PD-MCI patients, at baseline, cortical thickness was significantly reduced in the frontal, parietal and occipital cortices (left supramarginal cortex, bilateral rostral middle frontal cortex, left isthmus cingulate and right posterior cingulate cortices, and the right lateral occipital cortex). Over 18 months, PD-MCI had significantly increased percentage of cortical thinning in the frontal and parietal cortices (left superior frontal cortex, left supramarginal cortex, and right precuneus). They also established a significant positive assosiation with baseline cortical thickness and higher MoCA scores, in frontal and temporo-parietal cortices (left fusiform gyrus, left superior frontal cortex, left inferior parietal cortex, left orbitofrontal cortex and right parahippocampal gyrus).

Recently, Sampedro and colleagues [53] characterized the relationship of dopaminergic degeneration and cortical thickness of 87 de novo PD patients and 38 healthy controls from the PPMI database. They employed striatal DAT uptake to assess dopaminergic degeneration and FreeSurfer to measure cortical thickness from T1-w images. Their results show a significant link between dopaminergic integrity and the reduction of cortical thickness in frontal and posterior-cortical regions that in turn affect cognitive performance.

Aside from cortical thickness, less attention has been paid to features such as gyrification that assess the degree of folding of the cortical ribbon, that is the percentage of the cortex that is visible from the exterior of the brain compared to the portion that is buried in the circonvolutions.

Zhang and colleagues [54] effectuated a study including 37 PD patients at different stages of the disease and 34 matched HC where they compared the local gyrification indexes of the two populations. They employed FreeSurfer and found significant reductions in the inferior parietal cortex, the parahippocampal gyrus, the entorhinal cortex, the lingual and fusiform gyri, the orbitofrontal cortex and the right superior and middle temporal gyri.

Sterling and colleagues [55] realized a longitudinal cohort study with 70 PD patients (17 de novo) and 70 HC. They measured the local gyrification index with FreeSurfer and reported that patients with less than 4 years of illness did not have any significant loss in gyrification at baseline but over time presented reductions in the postcentral, precentral, superior frontal and supramarginal areas.

Authors	Features	Dataset	Findings
Tinaz & al. (2011)	SBM (in-house)	15 HC 15 early PD	Cortical thinning in the oribtofrontal cortex, the ventrolateral prefrontal cortex and the occipito-parietal areas.
Pereira & al. (2012)	SBM *FreeSurfer*	20 HC 20 PD	Cortical thinning in sections of the occipital and parietal cortex on both sides and the occipital and temporal cortex in the right hemisphere.
Ibarretxe-Bilbao & al. (2012)	SBM	16 HC 16 early PD (longitudinal)	No relevant differences at baseline. ↗ rate of widespread cortical thinning.
Malek & al. (2015)	SBM *FreeSurfer*	37 HC 66 early PD 39 early PD-MCI (longitudinal)	At baseline: No differences between controls and PD patients with normal cognition but cortical thinning in the frontal, parietal and occipital cortices of PD-MCI. At 18 months, significant differences of cortical
Sampedro & al. (2018)	SBM + DAT *FreeSurfer*	38 HC 87 early PD *PPMI*	Significant link between dopaminergic degeneration and cortical thickness in frontal and posterior-cortical regions.
Zhang & al. (2013)	SBM *FreeSurfer*	34 HC 37 PD	Significant gyrification reductions in the inferior parietal cortex, the parahippocampal, lingual, fusiform and the right superior and middle temporal gyri.
Sterling & al. (2016)	SBM *FreeSurfer*	70 HC 70 PD	Patients with less than 4 years of illness do not present significant losses in gyrification.

TABLE 3.3: Summarized findings of morphological surface PD biomarker studies.

3.2 Studies on quantitative data

A biomarker is defined as a quantitative characteristic which that can be employed as an indicator of biological or pathological states [38]. As seen in the previous section, changes in thickness, volumes and densities of gray matter can be assessed from structural images. Quantitative MR data are considered, in particular, to be as independent as possible of the MRI scanner or the study center [36]. The scanner is no longer considered only as a camera but as a means of measurement. They can be employed in a straight-forward manner to study physiopathological changes at the voxel level, notably relaxometry, diffusion and perfusion imaging.

3.2.1 Diffusion studies

Undeniably diffusion tensor imaging (DTI) has been the subject of many studies, Schwarz and colleagues at the Queen's Medical Centre in UK assessed the diagnostic value of nigral FA and MD measures based on a cohort of 32 PD patients and 27 healthy controls (HC) [56].

In contrast with Du and colleagues [57], they didn't find any differences in SN FA values between patients with PD and controls but a significant increase of MD in the SN (P <0.005).

Schwarz and colleagues also performed a systematic review and meta-analysis to estimate the disease effect size (D_{ES}) related to nigral DTI changes. The results indicate there is no significant disease effect nigral MD changes (10 studies, D_{ES}=+0.26, P=0.17, I^2=30%) and in the case of nigral FA, the values were significantly reduced in PD patients when taking into account data from 11 studies with a high variability (I^2=86%); however, after exclusion of five studies with unusual high values of nigral FA in the control group, an acceptable heterogeneity was reached, but there was non-significant disease effect (D_{ES}=-0.5, P=0.22, I^2=28%). The general conclusion being that the available published reports do not support nigral DTI metrics as useful diagnostic marker of PD at this point in time [56].

Cochrane and Ebmier [58] scoured the literature on diffusion studies from 1946 to 2012 and conducted a meta-analysis with 43 studies totaling 764 controls and 958 patients affected by a parkinsonian syndrome. The heterogeneity of the pathological population with PD was not negligible, notably regarding acquisition parameters, analysis methods and the introduction of medication. Most studies focused on the SN as ROI and often reported FA reductions in different segments with a slight tendency towards the caudal segment. However, no significant association was detected between disease severity and FA values.

The principal flaw of meta-analyses is inter-study heterogeneity. The available MRI data is unlikely to be acquired with the same scanner, the extraction of parameters could be done with different software and there can be important age differences. Large datasets like the Parkinson's Progression Markers Initiative (PPMI) from the Michael J.Fox Foundation allow to conduct studies on a large number of patients for whom the age, sex, clinical information and scanner conditions are known.

In another SN study, Schuff and colleagues [12] studied FA, radial and axial diffusivity on a cohort made up of 67 HC and 153 newly diagnosed PD patients from the PPMI database. They defined manually six ROIs within the SN, the rostral, medial and caudal segment of both hemispheres. In an interesting choice they decided to consider the laterality of the pathology and so, instead of separating the hemispheres into left and right, they employed the contra and ipsilateral labels, where the contralateral corresponds to the brain side opposite to the body side presenting greater symptoms. They found a significant reduction (p=0.04) of FA in the rostral aspect of the contralateral SN of PD patients. Furthermore, they reported significant relationships between the dopaminergic deficits displayed in DAT scan images and the FA values of the rostral and caudal regions of the contralateral SN. In spite of these results, FA contralateral rostral values only achieved a classification ROC AUC of 59% using a bootstrapped half-split cross-validation proceadure.

The variability in reported locations with altered diffusion in the SN is not surprising given that the SN is a small heterogeneous structure [12], nonetheless other studies have continued to explored DTI alterations beyond the SN.

One example lies in the meta-analysis of 39 publications conducted by Atkinson-Clement and colleagues at Aix-Marseille Université [59]. In the first part of the analysis, regions of interest (ROI) were classified anatomically (subcortical structures; white matter; cortical areas; cerebellum) [27 ROI, 1087 PD and 768 HC] . Then, a statistical analysis was effectuated considering the disease effect size (D_{ES}) as the main variable; fiber degradation in PD was expected to be associated to negative FA-D_{ES} scores and positive MD-D_{ES} scores. The results highlight five regions demonstrating significant differences between pathological and healthy subjects. Four of these regions presented reduced values of FA-D_{ES} and increased MD-D_{ES}: the SN, the corpus callosum, the cingular and temporal cortices, while the corticospinal tract displayed increased FA-D_{ES} and decreased MD-D_{ES}.

On the collaborative research of Martin Cousineau and colleagues [60], the DTI scans of 412 PD patients and 179 controls from the PPMI database were processed, this time to dissect 50 white matter (WM) fascicles using a method called high angular diffusion imaging (HARDI) crossing fiber modeling and tractography. The chosen fascicles connect deep nuclei (thalamus, putamen, pallidum) to different cortical functional areas (associative, motor, sensorimotor, limbic), basal forebrain and substantia nigra; only the ones that passed a test-retest reproducibility procedure qualified for further tractometry analysis. Welch's unequal variances t-test identified the sections of those tract profiles that were significantly different between PD patients and controls. The analysis found statistically significant differences in tract profiles along the subcortico-cortical pathways between PD patients and healthy controls. In particular, significant increases in FA, apparent fiber density, tract-density and generalized FA were detected in some locations of the SN-STN-Putamen-Thalamus-Cortex pathway, which is one of the major motor circuits balancing the coordination of motor output.

Talai and colleagues [61] demonstrated that diffusion alterations could also be helpful to differentiate PD from other parkinsonian syndromes such as progressive supranuclear palsy (PSP). Using a feature selection method the FA and MD average values of 17 ROIs were chosen to train a SVM in a leave-one-out cross-validation procedure. The classifier was able to differentiate PD (52) and PSP (21) patients with an accuracy of 87.7% (sens = 88.5% , spe = 82.3%) . Their results show that PSP patients present more sever and widespread diffusion alterations than PD patients. The main affected regions were the brainstem, putamen, palladium, thalamus and some areas of the frontal cortex.

In a similar study, Correia and colleagues [13] went farther proposed to employ SVMs to classify between controls (43), PD (32), PSP (33) and corticobasal syndrome (CBS) (26) patients with FA and MD measures in selected regions of white matter. They were very conscientious in the quality control of the data and the set up of their classification tests. Leave-two-out cross-validation (CV) and independent sample test were employed in every experiment and feature selection was effectuated for both atlas-based anatomical regions and PCA components. The poorest classification performances were obtained when separating controls from patients. The mean accuracy for the leave-two-out cross-validation was of 61.3%

for the atlas ROIs and 85.4% for the PCA components, whereas for the independent sample the ROIs obtained a score of 59.7% and the PCA components 57.6%. In general PCA features provided higher classification accuracies in the cross-validation approach but suffered significantly when an independent sample was used for validation. This could be an indication that PCA components are more prone to over-fitting. The authors did not specify the names of the most relevant WM ROIs for differentiating between control and PD groups but from their illustrations we can observe that the left thalamic radiation, the left retrolenticular segment of the internal capsule, the left splendium and body of the corps callosum, the right posterior limb of the internal capsule and the bilateral middle cerebellar peduncle played an important role in said discrimination.

Some studies concentrated on white matter fiber tracts, Guimaraes and colleagues [62] compared FSL tract-based spatial statistics (TBSS) and tractography to study diffusion abnormalities in a population of 137 HC and 132 PD patients (24 mild, 60 moderate and 13 severe). While they did find significant lower FA in patients in genu, body and splendium of corpus callosum, internal and external capsule, corona radiata, posterior thalamic radiation, sagittal stratum, cingulum and superior longitudinal fasciculus, these differences were only corroborated in part for the severe patients in the tractography study. What is more, major abnormalities were found in the corpus callosum and could be a direct result of cognitive decline in advanced PD patients. The authors went over to conclude that TBSS and generally DTI analysis might not be the best tool to assess early alterations in PD but could be employed to differentiate disease stages.

More recently, Wang and colleagues [63] made use of TBSS again, this time with a fully automatic pipeline that reduces both the bias of coregistration in voxel based methods and that of manual selection of fibers in tractography. The fiber tracts are identified and segmented into several clusters via an established pipeline, then the clusters are resampled into 200 nodes (points) each. The diffusion properties and a weight measure based on distance are calculated for each node and the mean value and standard deviation of each cluster are used to recognize abnormal tracts. Their study, while small (28 HC and 30 early PD from the PPMI) was able to find significant group differences in the cingulum, the inferior occipital fascilicus, the corpus callosum, the external capsule, the uncinate fascilicus, the superior longitudinal fascilicus and the thalamo frontal fascilicus.

As time advances, the techniques to study diffusion in the brain have seen major improvements, however, there is still a large heterogeneity in methodologies, with newer ones appearing every couple of years. This poses a problem when trying to reach a consensus from several papers results. Furthermore, the physiopathological complexity of PD suggests that, while diffusion holds interesting information regarding PD, DTI would be used in combination with other imaging modalities to construct a robust biomarker [58].

Authors	Features	Dataset	Findings
Schwarz & al. (2013)	FA and MD D_{ES} *(SN)*	32 PD 27 HC Meta-analysis	No differences in SN FA values but increased MD. Non-significant disease effect size in the meta-analysis for DTI measures.
Cochrane & Ebmier (2013)	FA and MD	764 HC 958 PD Meta-analysis	FA reductions in the substantia nigra. No significant association between disease severity and FA values.
Schuff & al. (2015)	FA, RD, AD	67 HC 153 early PD *PPMI*	Significant reduction (p=0.04) of FA in the rostral segment of the controlateral SN of PD patients. Classification score ROC-AUC=59%.
Atkinson-Clement & al. (2017)	FA and MD D_{ES}	768 HC 1087 PD Meta-analysis	Reduced FA-D_{ES} and increased MD-D_{ES} in the SN, the corpus callosum, the cingular and temporal cortex. Increased FA-D_{ES} and decreased MD-D_{ES} in the corticospinal tract.
Cousineau & al. (2017)	HARDI crossing fiber modeling and tractography	179 HC 412 PD *PPMI*	Significant differences in the tract profiles along the subcortico-cortical pathway. Increases in FA and fiber density in the SN-STN-Putamen-Thalamus-Cortex tract.
Talai & al. (2018)	FA and MD	52 PD 21 PSP	The mean values of 17 ROIs including the brainstem, putamen, palladium and thalamus were selected to train a SVM that achieved a classification accuracy of 87.7%.
Correia & al. (2020)	FA and MD	43 HC 32 PD 33 PSP 36 CBS	The mean accuracy for the leave-two-out cross-validation was of 61.3% for the atlas ROIs and 85.4% for the PCA components, whereas for the independent sample the ROIs obtained a score of 59.7% and the PCA components 57.6%.
Guimaraes & al. (2018)	TBSS and tractography	137 HC 132 PD	$\searrow$ FA in severe patients in the corpus callosum, internal and external capsule, corona radiata, posterior thalamic radiation, sagittal stratum, cingulum and superior longitudinal fasciculus. Relationship to cognitive decline in advanced PD patients.
Wang & al. (2020)	TBSS	28 HC 30 early PD *PPMI*	Significant group differences in the cingulum, the inferior occipital fascilicus, the corpus callosum, the external capsule, the uncinate fascilicus, the superior longitudinal fascilicus and the thalamo frontal fascilicus.

TABLE 3.4: Summarized findings of diffusion PD biomarker studies.

3.2.2 Perfusion studies

Nuclear imaging techniques have been often used to assess pathological perfusion changes in Parkinson's disease [64, 65, 66] and other neurodegenerative diseases. Besides the need for a radioactive tracer PET imaging is a demanding and time-consuming proceadure and SPECT has a poor spatial resolution.

The use of MRI techniques to study perfusion are relatively new but have obtained good results. What is more, the introduction of Arterial spin-labelling (ASL) has made it possible to assess perfusion in a non-invasive manner.

Melzer and colleagues [67] effectuated the first sizeable study to derive a PD related perfusion network using ASL MRI. They examined the scans of 29 HC and 61 patients at different stages of the disease (26 drug naive and 35 under medication) with a PCA framework focusing on gray matter perfusion. The PD group exhibited reduced GM perfusion (ROC AUC = 0.66). The found PD-network describing this heterogenous population was characterized by a decreased cortical (posterior parieto-occipital cortex, precuneus, cuneus and middle frontal gyri) and preserved subcortical and sensorimotor cortical perfusion.

Taking direct advantage of the quantitative nature of ASL, Fernandez-Seara and colleagues [15] compared the ASL scans of 25 early to moderate PD patients to 34 healthy controls through statistical parametric mapping. PD patients were characterized by widespread cortical hypoperfusion possibly related to the loss of cognitive functions. Within the subcortical structures, decreased perfusion in the caudate nucleus was detected. This study also discusses the dangers of global mean normalization on populations with generalized differences.

Teune and colleagues [68] conducted one of the first investigations comparing pCASL images of healthy controls (17) to those of PD patients (14). Employing PCA analysis the authors compared the PD-related perfusion covariance pattern to the metabolic covariance pattern obtained from PET images. They reported that both patterns correlated positively (0.5) but did not overlap completely. This being said the obtained perfusion pattern was in accordance with the literature, showing decreased activity in the cortical regions, including the insula and relative increases in activity in the cerebellum and pons, the right thalamus and palladium, the sensorimotor cortex, the paracentral lobule and the supplementary motor area. This study presented pCASL as a promising tool for the early diagnosis of PD.

Studying neurovascular changes in PD patients with ASL, Al-Bachari and colleagues [69] investigated the CBF and ATT (arterial arrival time) values of 52 controls (18 with a history of cerebrovascular disease and 34 without) and 51 PD patients at different stages of the disease (H&Y: 2.6 ± 1.0). They found evidence of hyperperfusion in the globus pallidus and of hypoperfusion in the occipital cortex predominantly. Additionally, whole brain ATT revealed to be significantly longer PD group compared to the two groups of control subjects.

On a another ASL study, Erro and colleagues [70] attempted to differentiate perfusion patterns in the subcortical structures of healthy (34), PD (30) and MSA (30) patients. A perfusion reduction was observed in several cerebellar areas, the right caudate and the bilateral thalamus for both MSA and PD patients. Comparing PD and MSA, PD patients had significantly lower perfusion values in the right caudate and thalami.

Authors	Features	Dataset	Findings
Melzer & al. (2011)	ASL PCA	29 HC 61 PD (early to severe)	↘ GM perfusion in the cortex (posterior parieto-occipital cortex, precuneus, cuneus and middle frontal gyri)
Fernandez-Seara & al. (2012)	ASL	34 HC 25 PD (early to moderate)	↘ perfusion in the cortex, related to the loss of cognitive functions. Within the subcortical structures, decreased perfusion in the caudate nucleus was detected.
Teune & al. (2014)	pCASL + PET PCA	17 HC 14 PD	Positive correlation between the PD-related perfusion covariance pattern and the metabolic covariance pattern obtained from PET images. Decreased perfusion in the cortical regions, including the insula and relative increases in activity in the cerebellum and pons, the right thalamus and palladium, the sensorimotor cortex, the paracentral lobule and the supplementary motor area.
Al-Bachari & al. & (2017)	ASL (CBF + ATT)	52 HC 51 PD	Hyperperfusion in the globus pallidus and of hypoperfusion in the occipital cortex. Significantly longer ATT in the PD group.
Erro & al. (2020)	ASL	34 HC 30 PD 30 MSA	CBF ↘ in several cerebellar areas, the right caudate and the bilateral thalamus for both MSA and PD patients. ↘ perfusion in the right caudate and thalami of PD compared to MSA.

TABLE 3.5: Summarized findings of perfusion PD biomarker studies.

3.2.3 Multivariate MR studies

Several studies in recent years have made use of more than one MRI feature to bring about a deeper understanding of PD as summarized in Table 3.6. Probably one of the oldest ones was carried on by Péran and colleagues in 2010 [16]. T2*-w, T1-w and DTI scans were acquired from 30 PD patients and 22 control subjects as to extract transverse relaxation rate (R2*), mean diffusivity (MD) and fractional anisotropy (FA) values of the in the pallidum, putamen, caudate nucleus, thalamus, SN and red nucleus. Region-based and voxel-based analysis showed that compared to control subjects, patients with PD had significantly higher R2* values in the SN, lower FA values in the SN and thalamus, and higher MD values in the thalamus and striatum. Furthermore, logistic regression analysis calculated that the combination of three different markers was sufficient to obtain >95% discrimination between patients with PD and controls.

One year afterwards, a complementary study by Du and colleagues at Pennsylvania State University [57] observed the changes in R2* and FA in the SN of 16 PD patients compared to 16 healthy volunteers. The association between R2* and DTI measures was dismissed via Pearson correlation suggesting that R2* and FA changes may reflect different pathological aspects of PD. Their results for regional covariance analysis showed both significantly

increased R2* and reduced FA values in the SN. Using logistic regression and receiver operating characteristic (ROC) curves to compare the sensitivity and specificity of R2* and FA they found significant discrimination between the groups using either R2* or FA alone (R2*, Area under ROC = 0.930, P < .0001; FA, Area under ROC = 0.742, P < .0001). The combination of R2* and FA in the SN, however, yielded even greater discrimination (Area under ROC = 0.996, P < .0001).

Another network that arouses interest is the intracranial visual system. In their study, Arrigo and colleagues [71] investigated this system based on a cohort of 20 *de novo* patients and 20 controls. They combined DTI to assess white matter diffusion with VBM to explore constrained spherical deconvolution-based connectivity and to investigate alteration in occipital gray matter. Significant alterations were found in optic radiation connectivity distribution, with decreased lateral geniculate nucleus V2 density ($P < 0.05$), a significant increase in optic radiation mean diffusivity ($P = 0.014$), and a significant reduction in white matter concentration. VBM analysis also showed a significant reduction in visual cortical volumes ($P < 0.05$). Moreover, the chiasmatic area and volume were significantly reduced ($P < 0.05$). Their conclusion states that visual system alterations can be detected in early stages of PD and that visual impairment affecting discrimination of color and contrast may be a sensitive marker, in accordance with a previous study by Diederich and colleagues in 2010 [72].

Functional connectivity (FC) in the basal ganglia network (BGN) has also been suggested as a promising biomarker for PD on its own [73, 74].

Including functional connectivity features in a multimodal approach is not without challenge. In their study, Bowman and colleagues [75] utilize elastic net, an advanced statistical learning technique to detect features from T1, rs-fMRI and DTI that reliably distinguish 28 PD patients from 14 healthy controls. Based on the Automated Anatomical Labeling (AAL) atlas, the brain was parcellated into 290 regions for which they calculated the volume (VBM), fractional amplitude of low frequency fluctuation (fALFF), FC, FA and structural connectivity (SC) measures. To reduce the number of features, a bootstrap procedure was employed and the remaining variables entered the elastic net method from which 24 parameters were chosen, corresponding to the 10% with the greatest predictive power. 21 of the parameters came from FC, one from SC and two from VBM. Considering the importance of parsimony, the authors brought forward three models that using three of the 24 mentioned features achieve a perfect distinction of the two groups (PD and HC). One of these models combines the FC of the left frontal superior medial region with the left anterior cingulum, the FC of the left thalamus with the left mid-temporal pole and the VBM of the right inferior frontal orbit.

This is one of the first approaches that truly combine the predictive power of different parameters and not only adds them, yet, 42 subjects may not be enough to validate the technique and adding more subjects is likely to be computationally expensive.

Also focusing on brain connectivity, Amoroso and colleagues [76] pooled T1 images (169 HC and 243 de novo PD) from the PPMI database and proceeded to extract connectivity measures with an approach borrowed from graph theory. All images were divided into an equal number of patches, every patch was considered as a node in the graph having as weights the pairwise Pearson's correlation measured between each pair of nodes throughout the whole brain. To classification between controls and patients maps was achieved by a combination of two Random Forest (RF) and a Support Vector Machine (SVM) classifiers. The first RF was dedicated to feature selection and the second one to subject scoring. The SVM combined these scores and other clinical features to discriminate between PD and HC scans and achieved an accuracy score of 93% corresponding to an specificity of 93% and a sensitivity of 92%.

More recently, Laganà and colleagues [77] studied the interaction between aberrant functional connectivity and perfusion alteration in a cohort of 26 early PD patients and 18 age and sex matched controls. rs-fMRI images were processed by an independant component analysis from which 11 FC components were retained, additionally CBF maps were extracted from pCASL MR images. FC was reported to be significantly reduced in PD patients within a sensory-motor network (p-FWE=0.01) and the primary (p-FWE=0.022) and lateral (p-FWE=0.01) visual networks. Perfusion in these visual networks was found to be reduced as well, pointing towards alterations of other systems besides the dopaminergic system. No gray matter atrophy (assessed from T1 scans) was related to these changes.

Authors	Features	Dataset	Findings
Péran & al. (2010)	R2* FA, MD *(basal ganglia)*	30 PD 22 HC	Higher R2* in the SN, lower FA in the SN and the thalamus, higher MD in the thalamus and striatum. 95% discrimination of PD and HC using the 3 values combined.
Du & al. (2011)	R2* and FA *(SN)*	16 PD 16 HC	R2* and FA values of the SN are uncorrelated and significantly discriminatory on their own; combined they yield greater sensitivity and specificity.
Arrigo & al. (2017)	DTI VBM *(Visual system)*	20 PD 20 HC	Alterations in optic radiation connectivity. Reduction in visual cortical and chiasmatic volumes.
Bowman & al. (2016)	VBM fALFF FC FA SC	28 PD 14 HC	24 region specific parameters, mostly issued from functional connectivity. The combination of any three features discriminates PD from HC.
Laganà & al. (2020)	FC CBF VBM	26 early PD 18 HC	↘ FC in the SMN (p-FWE=0.01) ↘ CBF and FC in primary (p-FWE=0.022) and lateral (p-FWE=0.01) visual networks.

TABLE 3.6: Summarized findings of multiparametric PD biomarker studies

3.3 Deep learning studies

Deep Learning (DL) is a growing trend in the general data analysis. Recent breakthroughs like the explosion of the available computer power and the availability of data have allowed artificial neural networks to obtain state-of-the art performances in several computer vision challenges (i.e. ILSVRC2012). These developments have great potential to exploit the valuable information generated in the medical field.

Concerning Parkinson's Disease, several strategies have been developed to analyze physiological signals. For example, to analyze data recorded by wearable devices and characterize symptoms such as bradykinesia [78] or freezing of gait [79]. They have also been effective to detect particularities in PD patients handwriting [80] and even to identify PD specific vocal features [81].

Combinations of all of these features have been studied in [82] and included in smartphone applications [83, 84] including one that is now a registered Class I medical device: cloudUPDRS [83] that utilize a previously supervised Feed Forward Artificial Neural Networks (FFANN) to analyze the subjects results.

Medical imaging represents the largest percentage of data produced in healthcare and thus a particular interest has been taken in DL methods to create support tools for radiologists to analyze images, segment lesions and even detect subtle pathological changes that even an expert eye can miss.

Over the years different architectures have emerged and within the MRI community several public databases have been curated with annotated images to compare their performances. Some of the most prominent databases are the MSSEG, for MS lesion segmentation; BRATS, for brain tumor segmentation; ISLES, for ischemic stroke lession segmentation; and mTOP for mild traumatic injury outcome prediction.

Challenges are organized regularly to showcase the latest technological advancements and push the community towards better performances in these important tasks.

In the case of PD, this kind of challenges are unheard of as any remarkable features are only visible on MR scans once the disease is at an advanced stage. Moreover, any outcome prediction would necessitate a clear picture of the possible outcomes whereas at the moment the phenotypical progression paths of the disease are still up for debate. Nevertheless, convolutional neural networks have been proposed to discriminate between healthy and pathological data and unsupervised deep learning models can be used in an exploratory manner to identify subtle anomalies in the MR images of PD patients.

In the literature, most state-of-the art DL anomaly detection techniques use Generative Adversarial Networks (GAN) or Auto-encoders (AE) and their variations. Both types of architectures are trained to model "normal" data with two complementary networks. GANs are composed of a generator and a discriminator that learn simultaneously to generate "normal"

data from samples in a latent space and to discriminate between true and generated data. The overall goal is to train a generator that produces candidates susceptible to fool the discriminator. Differently, AEs trains an encoder and a decoder networks that can effectively reconstruct "normal" inputs. The encoder summarizes the data into a smaller latent space representation from which the decoder can construct an image. Once the models are trained, different anomaly scores can be defined to measure how far new observations fall from the normality model.

One of the precursor applications of unsupervised DL methods for anomaly detection on medical data was proposed by Schlegl and colleagues [85] to identify anomalous regions in spectral-domain OCT scans of the retina. Their architecture, named AnoGAN, consisted on a convolutional GAN architecture trained with 2D patches (64×64). During training, the generator learned to map 1D vectors of uniformly distributed input noise to a 2D image in the image space manifold χ, which is populated by healthy examples exclusively. During testing, for every query image, first the latent vector corresponding to the most similar image in the manifold was found via backpropagation, then, an anomaly score was computed. This score was defined as the combination of two losses : the dissimilarity between the generated and the query image on one side and the sigmoid cross-entropy of the generated image with respect to χ. AnoGAN achieved a ROC AUC of 89%.

As presented by Varghese and colleagues [86], another possibility is to use the discriminator of the trained GAN architecture to output, for each query image, a probability map that gives an indication of the likelihood for every point of belonging to the learnt "normal" data distribution. The authors applied the proposed method to detect brain tumors from the multimodal MR images (FLAIR, T2, T1, T1 post contrast) of the BRATS dataset. The model was trained with 13 000 healthy patches of 64×64 patches and obtained a DICE score of 69 % for the whole tumor segmentation task. This corresponds to a sensitivity score of 92% and a specificity score of 59%.

In an unpublished study, Ha Son and colleagues [87] proposed to increase the training stability of GANs by employing two autoencoders as generator and discriminator. Their model, called ADAE (Adversarial Dual Autoencoders), was also trained on BRATS data and achieved a ROC AUC score of 89.2%.

Baur and colleagues also propose to combine GAN and AE architectures. They devised the AnoVAEGAN model, which is constituted by a Variational Autoencoder that generates reconstructions of input images with a discriminator network at the end to assesses every point in the reconstructions as "real" or "fake", otherwise "healthy" or "pathological".

To evaluate the model, the authors conducted experiments on a private dataset containing FLAIR and T1 image pairs from 83 healthy controls and 49 patients with multiple sclerosis (MS) lesions. For each brain volume, 20 consecutive axial slices of 256×256 pixels were extracted. Experiments compared the proposed model with AnoGAN [85] and with dense and spatial variants of VAEs and AEs. Results measured by the DICE score in the task of MS

lesion segmentation showed that AnoVAEGAN beats AnoGAN (0.605 vs. 0.375) as well as dense and spatial variations of VAEs and AEs. However, spatial VAE (0.592) and spatial AE (0.585) without any adversarial training performed only slightly worse than the proposed model.

While a DICE score of 60.5% may seem poor compared to the previous DL results here exposed, MS lesions are generally more challenging to segment than tumors. Additionally, the DICE score evaluates the accuracy of the automatic segmentation, whereas if the task was merely to detect the presence of a lesion, the above mentioned techniques would have achieved much better performances in MS.

Authors	Architecture	Dataset	Performances
Schlegl & al. (2017)	GAN AnoGAN	OCT scans of the retina 64×64 patches	*Classification* ROC AUC: 89%
Varghese & al. (2017)	GAN	BRATS 13 000 64×64 patches Multimodal data	*Segmentation* DICE: 69 % sens: 92%, spe: 59%
Ha Son & al.	GAN + AE ADAE	BRATS	ROC AUC: 89.2%
Baur & al.	GAN + AE AnoVAEGAN	MS dataset 83 HC & 49 MS 256×256 slices	*Segmentation* DICE AnoVAEGAN: 60.5% VAE: 59.2% AE: 58.5%

TABLE 3.7: A brief sample of some medical DL biomarker studies

In the scale of difficulty, early PD anomaly detection is much higher than MS lesion detection. Indeed, as mentioned before, early symptoms in PD have not yet been translated to identifiable characteristics in MR imaging. Regardless, in the last year some innovative methods have been devised to apply unsupervised deep learning to PD anomaly detection in MR imaging.

Li and colleagues[88] set out to employ a stacked sparse autoencoder (SSAE) to classify between controls (62) and PD patients (142) from longitudinal data pooled from the PPMI database. T1 and DTI maps were obtained for every subject at baseline, after 12 months and over 24 months. Gray matter, white matter and mean diffusivity features were extracted for every one of the 116 ROIs in the AAL atlas and serve as input for the SSAE. The outputs of the architecture at the three time-points are classified with an SVM. The proposed method achieved a ROC AUC score of 86% at baseline and 97% at 24 months. They compared their results with a simple sAE, a CNN and a DBN architecture. The sAE achieved the second best results with a ROC AUC of 82% at baseline and 92% at 24 months.

Other studies prefered to employ CNNs to discriminate PD from HC scans. Shinde and colleagues [89] adapted the well know ResNet50 design to study 2D NMS-MRI images of the substantia nigra (45 PD, 35 HC) from a local database. The average Hoehn and Yahr

score for the patients was of 1.70 ± 0.54. They achieved a good classification performance (ROC-AUC = 90%) taking into account clinical laterality.

Sivaranjini and colleagues [90] proposed to employ the popular AlexNet CNN to discriminate T2-weighted MR scans of 82 HC and 100 PD patients from the PPMI at various stages of the disease. One particularity of this investigation is the utilization of transfer learning, that is, the CNN architecture had been pre-trained with natural images before medical images were input. Their results show an accuracy of 88.9% corresponding to 89.3% in sensitivity and 88.4% in specificity.

Authors	Architecture	Dataset	Performances
Li & al. (2019)	SSAE + SVM	PPMI: 62 HC & 142 PD T1 and DTI maps divided by 116 ROIs at baseline	SSAE ROC AUC: 86% sAE ROC AUC: 82%
Shinde & al. (2019)	CNN: ResNet50 adaptation	35 HC and 45 PD 2D NMS-MRI images of the SN at H&Y= 1.7 ± 0.54	ROC AUC: 90%
Sivaranjini & al. (2020)	CNN: AlexNet with transfer learning	82 HC and 100 PD (PPMI) T2-w MR scans	*Classification* sens: 89.3% spe: 88.4%

TABLE 3.8: Summarized findings of DL PD biomarker studies

Part II

ANOMALIES IN ’DE NOVO’ MR SCANS

Chapter 4

Study of structural changes

4.1 A word on morphometric analysis

Anatomical MR images provide detailed information about the shape and size of several brain regions *in vivo*. In a world were 100 years ago any knowledge on the brain's anatomy came from autopsy studies, this has been quite revolutionary.

Scanners having field strengths of 3 Tesla or higher render images of great spatial resolution where brain tissues can be accurately differentiated and investigators can delineate brain structures. The quantification of their shapes and sizes is named morphometry [91].

All morphometric studies meet two requirements: (1) the availability of measurable features of interest belonging to at least two populations and (2) appropriate statistical methods to compare the measurements quantitatively. Given the complex anatomy of the brain, it comes without surprise that multiple techniques have been developed for feature specific purposes. The most popular method is know as *Voxel-based morphometry (VBM)* [92]. In VBM, the images are spatially normalized into a template space and then compared at a voxel level to disclose any regionally specific differences. Other approaches have been developed to employ the deformation fields used to normalize the neuroimaging data in a straightforward fashion, this is the case of *Deformation-based morphometry (DBM)* [93, 94]. Another option is to use brain segmentation to reconstruct the boundaries between different classes of tissue as a surface and study features like cortical thickness and gyrification. This method is best known as *Surface-based morphometry (SBM)* [95].

We note that fiber tracking based on diffusion-weighted imaging is also considered as a morphometric analysis but it will not be discussed in this chapter.

The use of structural T1-weighted images may be the most common in clinical MR imaging. While the signal in these images is generally inversely related to the longitudinal relaxation time (T1) in a nonlinear manner, many factors can alter their calibration and thus, they cannot be used to produce quantitative results. However, they are adequately employed in VBM, DBM and SBM techniques after some standard pre-processing steps.

Several automated non-operator dependent pipelines have emerged that involve all steps from pre-processing and registration to the segmentation of tissues and structures. They are supported by sensitive methods that help detect statistical deviations of the anatomical features from the normal brain. We can cite famous open-source software packages like SPM, FSL and FreeSurfer. There are also several universities that have developed their own tools like the CIVET pipeline of the McGill University and in the latest developments the Polytechnic University of Valencia and Bordeaux University provide not only access to their software pipeline called volBrain [5] but also share the computational resources of their institutions.

The work presented on this chapter is the fruit of a collaboration with Pierrick Coupé's team at Université de Bordeaux, who developed volBrain alongside José V. Manjon's team at the Polytechnic University of Valencia.

4.2 Pipeline description

The study of morphological brain differences between pathological and healthy groups can potentially identify key regions affected during the PD prodromal phase to better understand PD pathophysiology and its treatment. As seen in Chapter 3 several studies have already focused on this subject however, population size is often reduced and results reproducibility is seldom considered. Here, we explore this question with up-to-date methods applied to a large and openly available population.

In order to elucidate the nature of morphological differences in de novo PD patients, performed a VBM study through both 1) the well-established Computational Anatomy Toolbox (CAT12) (University of Jena) included in the current version of the Statistical Parametric Mapping (SPM12) software and 2) a new online platform: volBrain [5]. Both pipelines have complementary strengths that are exploited in this study: volBrain performs quality segmentation of subcortical nuclei and CAT12 facilitates group analysis. Furthermore, we looked for quantitative differences between the tissue classification performed by the two approaches, both including partial volume estimation.

Subsequently, DBM and SBM studies were carried out on CAT12 uniquely, as volBrain is a volumetry focused system. A representation of the whole pipeline is available in Figure 4.1

Our study included 144 de novo PD patients (age: 61.30 ± 9.06; sex: 53 F, 91 M) and 66 healthy controls (age: 60.12 ± 11.39; sex: 23 F, 43 M) from the PPMI database [6] (www.ppmi-info.org/data). The PPMI (Parkinson's Progression Markers Initiative) is a longitudinal study that follows *de novo* PD patients of 35 centers for five years. The database is openly available for researchers and contains, among other clinical and behavioral assessments, structural and DTI MR scans.

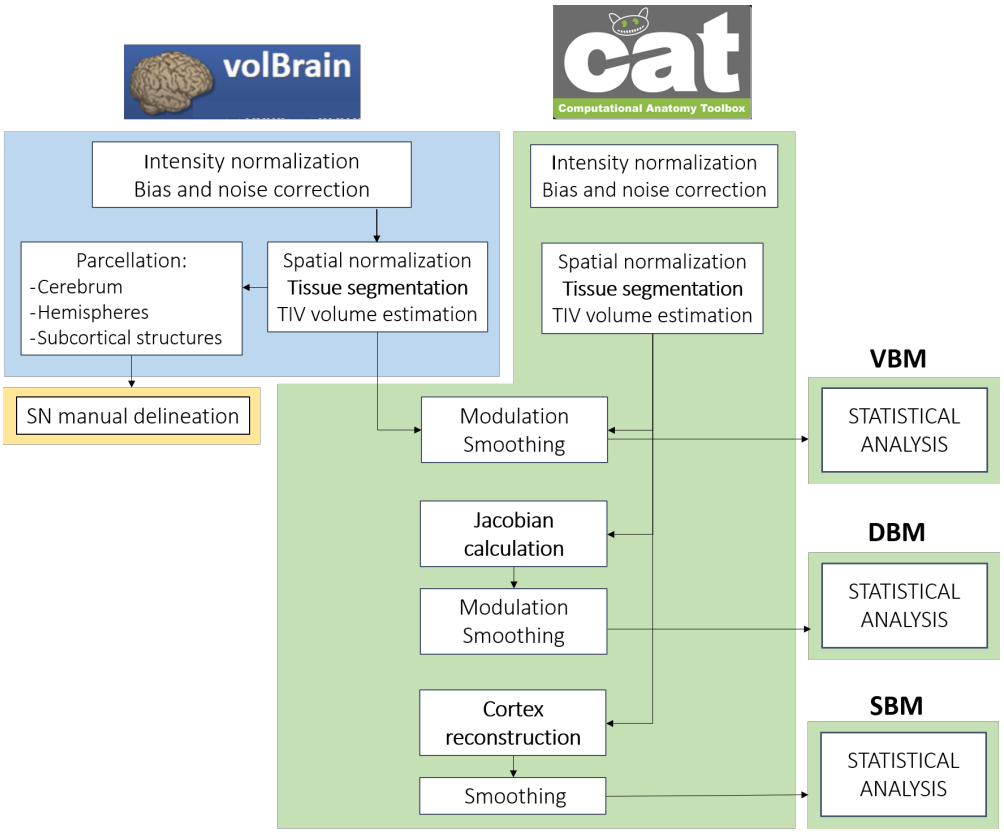


FIGURE 4.1: Graphical description of the morphometric analysis pipeline employed. Tissue segmentation maps were estimated with both volBrain and CAT12 for VBM analysis. A manual delineation of the substantia nigra (SN) was added to the parcellated regions extracted with volBrain for ROI analysis. DBM and SBM analysis were carried out with CAT12 exclusively.

The structural T1-weighted MRI images extracted were acquired with a 3T Siemens Trio Tim scanner with repetition time (TR) = 2300ms; echo time (TE) = 3ms; flip angle = 9^o; thickness = 1mm. We chose to pool data acquired with the same acquisition parameters, including magnetic field and scanner manufacturer, to eliminate any additional sources of bias.

4.2.1 Data pre-processing

While algorithms may progress over time, the pre-processing steps for morphometric studies are very similar and typically include:

- Noise reduction. The images are smoothed with a filter that preserves boundaries between brain tissues.
- Intensity correction. The variations introduced by inhomogeneities in the magnetic field B0 or fluctuations in the radiofrequency pulses are adjusted.
- Spatial normalization. The images are registered to a reference template brain.

- Brain tissue segmentation. Delimitation of the brain, the CSF, white matter and gray matter, for instance.

CAT12 and volBrain follow these steps with some methodological differences. They both employ a Spatially Adaptive Non-Local Means (SANLM) filter [96] for noise correction, however CAT12 registers the images to a reference brain using an affine and non-linear method (DARTEL and Geodesic Shooting), whereas volBrain utilizes a linear affine transformation with ANTs software. Concerning intensity correction, volBrains opts for the N4 method [97] for coarse inhomogeneities and SPM8's method [98] for fine inhomogeneities once the images are registered. CAT12 only corrects inhomogeneities before registration with SPM8's method which consists of a parametric bias correction where the intensities of different tissues are modeled as a mixture of Gaussians and the bias are estimated as a smooth intensity variation multiplying the true signal.

For tissue segmentation, CAT12 integrates a classical Markov Random Field and the Adaptive Maximum Posterior (AMAP) technique to reduce the dependency on Tissue Probability Maps. In addition, partial volumes are estimated using two additional mixed classes: GM-WM and GM-CSF. volBrain employs the Trimmed Mean Segmentation (TMS) method [5], robustly estimating the mean values of the different tissues by excluding partial volume voxels from the estimation jointly with the use of an unbiased robust mean estimator. Partial volume coefficients are computed as well.

CAT12 does not propose tools for tissue parcellation within its pipeline, yet several atlases are available for region-based statistical analysis. volBrain goes a step further and segments each individual brain into cerebrum, cerebellum, brainstem, hemispheres, and seven subcortical structures (putamen, caudate, globus pallidus, thalamus, hipocampus, amygdala, and accumbens) with a multi-template method that considers non-local label fusion schemes [96]. Additionally substantia nigra was manually delineated as can be seen in Figure 4.2.

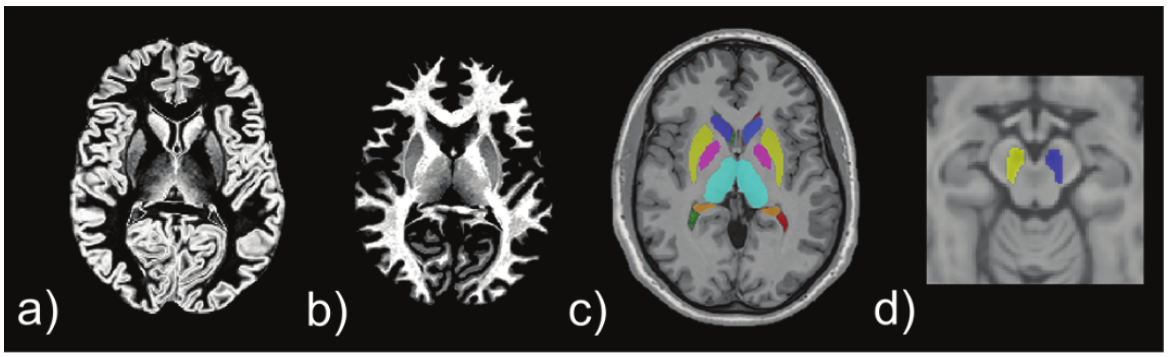


FIGURE 4.2: volBrain segmentation: a) GM, b) WM and c) subcortical structures. d) Manually segmented substantia nigra.

4.2.2 VBM : Voxel-based morphometry

VBM is most efficient when comparing the local concentration of gray matter between two groups of subjects. It can also be used to examine white matter but other methods such as DTI are more sensible to white matter effects.

4.2.2.1 CAT12

Before performing a statistical analysis some additional steps are necessary. The segmented images from both pipelines were modulated by scaling with the amount of volume changes due to spatial registration, so that the total amount of grey matter in the modulated image remained the same as it would be in the original image. The resulting images were smoothed with an isotropic Gaussian kernel of 8mm and finally the Total Intracranial Volume (TIV) was estimated so that it could be used as a covariate during statistical analysis.

We chose to employ a two-sample T-test to compare the CAT12 modulated tissue maps (GM and WM) of patients versus controls with a general linear model (GLM) where age, sex, and TIV were entered as covariates.

4.2.2.2 volBrain

The volBrain volumetric analysis is deployed through a web interface (http://volbrain.upv.es) and utilizes the resources of the Polytechnic University of Valencia. It does not require installation, configuration or training. The web interface takes a single anonymized compressed MRI T1-weighted Nifti file as input and outputs a pdf and csv file summarizing the volumes and asymmetry ratios estimated from the data. If the age and sex of the subject were submitted, population-based normal volumes and asymmetry bounds for all structures are added to the report for reference purposes. These normality bounds are automatically estimated from the IXI dataset (http://www.brain-development.org) which contains almost 600 normal subjects covering most of adult lifespan. Nifti files containing the gray matter, white matter, and subcortical structures segmentations in the native and MNI spaces are also available for download.

For large studies it is possible to contact the team to input the T1w images in batches instead of one-by-one. To perform a group analysis on the calculated tissue maps CAT12's procedure was utilized. The GM and WM were deformed to the same template space as CAT12's outputs by applying a forward deformation with the corresponding DARTEL fields, then they were spatially smoothed with a 8mm kernel.

A two-sample T-test was employed again to compare volBrain's tissue maps with sex, age and TIV as covariates.

The subcortial structures segmentations were also deformed into the template space to perform some complementary region-based morphometric analysis.

4.2.3 DBM : Deformation-based morphometry

DBM is a valuable tool to detect morphological variations in the entire brain since it analyses positional differences between every voxel and a standard brain.

As mentioned before, morphometric analysis require images to be in the same stereotactic space. Spatial normalization is achieved by registering each of the brain scans to the same template image, thus minimizing the residual sum of squared differences between them. This procedure does not attempt to match every cortical feature exactly, but rather to correct for global brain shape differences.

Each individual voxel undergoes a transformation during the registration process. These transformations are encoded as displacement vectors and the map of all displacement vectors in the scan constitutes a deformation field.

Deformations reveal information about the type and localization of the structural differences between the brains and can undergo subsequent DBM analysis [93, 94]. One option is to compare the deformation fields through a multivariate test using the parameters that describe the deformations and employ a Hotelling's T^2 statistic for simple comparisons or a Wilk's λ statistic for more complex analysis. A second and probably the most popular option is to compute the Jacobian determinant of the displacement in each voxel to produce a map of local volume changes that can be quantified. This approach is often called Tensor-based morphometry (TBM) because the Jacobian determinant represents such a tensor. It is the strategy hereby adopted.

The resulting Jacobian determinant maps can be compared between the healthy and pathological groups using statistical tests like in the VBM strategy. A two-sample t-test was chosen here too to examine the previously smoothed maps (8mm) of both populations with age and sex as covariates.

As mild cognitive impairment has been strongly correlated to patterns in atrophy [48], a small study was also carried out on a sub-population of 24 PD patients that scored below 26 points on the MoCA test for cognitive assessment. The potential atrophy of those 24 PD subjects (age: 64.25 ± 6.52; sex: 6F, 18M) was analyzed against 24 randomly selected control subjects (age: 64.25 ± 12.34; sex: 8F, 16M) using a two-sample t-test as well.

4.2.4 SBM : Surface-based morphometry

The cerebral cortex is a highly folded sheet of gray matter with ridges (gyri) and faults (sulci) that is located between the cerebrospinal fluid on the exterior and a core of white matter in

the interior. It has an average thickness of around 2.5mm varying between a range of 1 and 4.5mm across different brain regions [99].

Cortical thickness and complexity are two additional features that convey relevant information on morphological changes of the brain. To effectuate these measurements there are two classes of methods available: surface and volume-based approaches [8].

Surface-based approaches require an explicit reconstruction of the brain's cortical surface delivering highly accurate measurements of the cortical surface at the expense of extensive processing times. FreeSurfer's software suite has positioned this method as the gold standard in cortical thickness measurements.

Volume-based methods work direcly from the MRI volumes drastically reducing computation time. However they are more prone to deliver inaccurate measurements due to the presence of noise and erratic partial volume effects (PVE) on the outer boundary. To overcome these issues, CAT12 introduced a projection-based thickness (PBT) technique [95] that uses a projection scheme considering buried sulci, an otherwise problematic region.

The PBT method estimates the distance between every GM voxel and the WM inner boundary to then project the local maxima to other GM voxels using a neighbor relationship defined by this distance map. PBT has been proven to handle partial volume information, sulcal blurring, and sulcal asymmetries without explicit sulcus reconstruction [8]. Additionally, PBT enables the easy reconstruction of the central surface, defined as the average of the two GM boundaries. This surface, passing exactly at the middle of the CSF-GM and GM-WM boundaries, is thought to be a better representation of the cortical sheet, less error-prone to noise.

Next, to compare the cortical surface of multiple subjects, a spherical coordinate system is adopted as a common space, but before being inflated (mapped to a sphere) the subjects surfaces should be free of any topological defects. The CAT12 SBM pipeline performs topological correction of defects using spherical harmonics [100]. In short, the defects are automatically labeled as 'fill' or 'cut' operations using the original T1 image intensities so that the brain surface mesh can be modified in a way that the defective points are no longer on the surface of the sphere. Then, a regularly sampled grid is overlaid on top of the modified mesh during reparametrization. The points outside the sphere are highly unlikely to intersect the grid and thus the defects are minimized. Finally, the areas that contained defects are replaced by a low-pass filtered reconstruction of the surface based on spherical harmonics. Indeed, a spherical harmonic can be described as Fourrier transform on a sphere that decomposes the brain surface data into frequency components, therefore applying an inverse Fourrier transform to the smoothed (low-pass filtered) harmonic produces the coordinates needed to reconstruct the surface mesh where defects laid.

Once all the subjects can be properly mapped into a sphere, CAT12 employs an adapted version of the volume-based diffeomorphic DARTEL algorithm to the surface [101] to work with spherical maps [102]. It consists on a multi-grid approach that uses reparameterized

values of sulcal depth and shape index defined on the sphere to estimate a flow field that allows deforming a spherical grid.

All these steps, from PBT to registration to mesh surface registration, are fully automated and invisible for the user. At this point, cortical thickness has been estimated and the central cortical surface has been created for the left and right hemisphere. It is also possible to extract additional parameters such as the gyrification index [103] and/or measures of cortical complexity and sulcal depth.

In this study we concentrate on cortical thickness and gyrification indices for our statistical analysis. Gyrification is the term employed to designate the ratio between deep (sulci) and superficial (gyri) in a region of the brain. This information can be projected onto the central surface for comparison.

The last action needed before analysis can proceed is to smooth the (central) surfaces. A filtering kernel of 15mm FWHM (Full Width at Half Maximum) is employed for thickness data and one of 20mm for gyrification following the standard software recommendations.

Cortical thickness and gyrification measures were also extracted for individual all the regions in the Desikan-Killiany 40 atlas [104] for supplementary region-based analysis.

All comparisons were effectuated using a two sample t-test with age and sex as covariates.

4.3 Results

4.3.1 VBM

When choosing a p-value of 0.05 with Family Wise Error (FWE) correction for multiple comparisons, no voxels survive the difference analysis between PD patient and control groups with tissue map computed with CAT12 or volBrain. In order to replicate some literature results (exploratory study), we decreased the statistical threshold to $p<0.001$ uncorrected for multiple comparisons. Several clusters were then found showing a volume decrease in PD patients mainly located in the bilateral posterior cingulate gyrus, the cerebellum, the right supramarginal gyrus and the left parietal operculum for CAT12 and for volBrain there is one cluster that almost perfectly coincides with the substantia nigra and other small clusters in the right occipital pole, the left superior frontal gyrus and white matter.

Also, two-sample T-test comparisons of each subcortical structure independently (computed by volBrain) failed to detect any differences in GM contents $p<0.05$ FWE corrected. In an exploratory test, differences were found in the caudate nucleus, the hipocampus and the putamen for an uncorrected p-value of 0.001.

"Small volume" analysis in SPM12 was used as well to study possible morphometric changes in the substantia nigra, key structure in PD research, using volBrain maps. We observed

that differences were only present for an uncorrected p-value of 0.001 and did not survived multiple comparison correction.

A recent study investigating the high rate of false positive present in VBM studies recommends the use of the same group size to detect morphological differences between two groups [105]. Following this recommendation, we repeated our analysis five times to compare the tissue maps of 66 controls versus 66 randomly selected patients using sampling with replacement technique. Their age and sex characteristics are summarized on Table 4.1

	Age	**Sex**
Controls	60.1 ± 11.4	43 M, 23 F
PD Patients	61.3 ± 9.1	91 M, 53 F
PD sample1	59.2 ± 9.2	45 M, 21 F
PD sample2	62.3 ± 7.9	36 M, 30 F
PD sample3	60.5 ± 9.0	39 M, 27 F
PD sample4	60.9 ± 9.0	38 M, 29 F
PD sample5	60.3 ± 9.4	42 M, 24 F

TABLE 4.1: Characteristics of the original study population and the 5 sub-samples of patients equal in size to the control group.

For all of the 5 new equal size sub-populations (see Table 1), no differences were found in the CAT12 tissue maps for $p < 0.05$ FWE corrected, whilst several significant clusters appeared for an uncorrected p-value of 0.001, specifically in the bilateral cerebellum, precuneus, and supramarginal gyrus, as well as the right posterior cingulate gyrus, the right posterior orbital gyrus and the left subcallosal area.

Concerning volBrain, there were some clusters at $p < 0.001$ uncorrected in the right supplementary motor cortex and the substantia nigra but most shared clusters appeared in the left superior frontal gyrus and the right occipital pole aside from smaller ones dispersed withing the white matter.

4.3.2 DBM

The two-sample t-test on the Jacobian deformation maps with age and sex as covariates failed to find any significant differences ($p < 0.05$) once corrected for family wise error for the general population and the five same-size sample populations.

At the exploratory theshold $p < 0.001$ uncorrected some clusters appeared in the left fusiform gyrus, the left inferior temporal gyrus and the left thalamus.

The only shared cluster amongst the five same-size samples at $p < 0.001$ uncorrected enveloped the amygdala, including a section of white matter and the hippocampus. Other clusters included the right putamen and a region between the left fusiform gyrus and the left inferior temporal gyrus.

The test on mildly cognitively impaired patients compared to controls did not bring to light any significant differences at $p<0.05$ FWE corrected and at $p < 0.001$ uncorrected a cluster appeared in the middle of the left putamen and the left anterior insula.

4.3.3 SBM

Concerning cortical characteristics, there were no significant thickness nor gyrification differences for a $p < 0.05$ FWE-corrected nor $p < 0.001$ uncorrected. This was also true for the five same-size populations.

The most notable results resided on the gyrification index of the bilateral pars triangularis with $p = 0.006$ for the right hemisphere and $p=0.008$ for the left hemisphere in the whole population study and an average of $p = 0.014$ in the right hemispheres of the same-size population.

4.4 Discussion

Using two recent approaches for accurate segmentation of tissues (CAT12 and volBrain) and subcortical structures (volBrain) we failed to detect robust structural differences in de novo PD patients and healthy controls. We took special care to consider a relatively large cohort of subjects, consider the effects of an unbalanced number of patients and controls and correct for multiple comparisons. Following these precautions, no morphological differences were found in PD patients, neither on whole brain GM group analysis or on the analysis of several subcortical structures separately.

In the literature, there is an overall lack of consensus on the presence of structural brain differences in de novo PD patients. These discrepancies may be due to a variety of factors. Some studies were carried out on small cohorts, no more than 60 subjects in total [41, 42, 43], so one may argue that the inconsistencies could be resolved with a larger cohort more representative of the population. Although, in [105], it was brought to light that sample size does not appear to influence false positive rate, a small sample may incorrectly represent a pathological population, hindering the reproductibility of results.

There is a wide variety of softwares for pre-processing MRI images (i.e. SPM, FSL), all using different techniques that will inevitably influence the final statistical results as proven by Popescu and colleagues [106] on the study of Multiple Sclerosis. By combining the latest improvements on VBM analysis present in CAT12 (notably denoising and partial volume

estimation) with the state of the art segmentations of volBrain [5] we sought to reduce estimation bias considerably. Finally, correction for multiple comparison is vital to reduce the false positive rate and thefore to produce robust and reproducible results [9]. Exploratory studies, which use lenient statistical thresholds, could be interesting to indicate some trends in the observed population, that should be confirmed by more robust studies. In the case of [44], the tests were FWE corrected, but the the study was effectuated on VBM8 while, according to [107], the CAT12 toolbox can contribute to more robust detection compared to VBM8.

Regarding the differences we observed between the tissue classification with CAT12 versus volBrain, raw volBrain's tissue maps seem to better distinguish the presence of gray and white matter in the subcortical nuclei. However, as in this study no morphological robust differences were found between PD patients and controls, a more in depth investigation would be necessary to pertinently test the performances of both methods of partial volume estimation.

The study took a step further to study any early signs of atrophy in the PD population using CAT12 but no significant disparities were established once family wise error correction was applied. Atrophy is seldom reported in early PD patients [47, 48] and there is a possibility that small investigations reporting differences may be biased.

We note that by omiting multiple comparison correction, several regions that become significant are part of the PD-ICA network proposed by Zeighami and colleagues [7] (i.e amygdala, thalamus, hippocampus). A possible reason for the discrepancies between our results are due to the method employed, they employed the CIVET pipeline and us the CAT12 pipeline, plus it is not clear if their correction was calculated for every voxel in the ICA-ROI or for every possible ICA, in which case our correction is far more strict. What is more, in [48] it was brought to light that the strongest correlates to the PD-ICA network were age and gender. While this is in accordance with the literature, aging is the largest risk factor for the development and progression of PD [108] and older males tend to have poorer prognosis [109], we stand by our decision of introducing age and gender as covariates in our statistical analysis to draw-out PD-specific patterns of atrophy.

Lastly, our study included less subjects from the PPMI database (PD:144, HC:66 vs. PD:232, HC:117) seeing that we chose to pool subjects imaged with the same acquisition parameters to reduce the sources of bias. If their study included more patients with a specific set of shared clinical traits such as MCI or anxiety, that were highly correlated to the PD-ICA network, the statistical power of their test was increased.

Our SBM results coincide with those of the latest research on cortical characteristics of PD patients, there are no significant differences in cortical thickness nor gyrification in de novo PD patients. This confirms the capabilities of CAT12 methods to correctly estimate the cortical ribbon, and this, in a fraction of the time compared to FreeSurfer. [8] estimated 10-20 hours per subject for the reconstruction of the cortical surface on FreeSurfer, whereas CAT12

effectuates its processing in less than an hour. To our knowledge, the SBM study carried out in this thesis is the largest one in the literature.

4.5 Conclusion

In conclusion, using two recent approaches for accurate tissue classification (CAT12 and volBrain) we failed to detect robust structural differences between de novo PD patients and healthy controls. Findings of structural brain differences reported in the literature tend to be contradictory. This lack of consensus may be due to a variety of factors, notably the study of small cohorts, differences in segmentation techniques, and more importantly, the lack of correction for multiple comparison inevitably increasing false positive rates. We reinforce the message that morphological assessments are delicate techniques involving many parameters that should be handled with care to avoid false positive influencing the final results [9].

In sight of the lack of morphological differences, we suspect that early PD biomarkers may lie on the physiological properties of the Parkinsonian brain and could be investigated through quantitative MRI techniques (relaxometry, diffusion, perfusion).

Chapter 5

Investigating diffusion changes with Neural Networks

5.1 A word on anomaly detection using auto-encoders

Deep learning based architectures have been employed with great success to detect pathological characteristics in medical imaging as illustrated in Chapter 3. Unsupervised methods enjoy a particular success as they by-pass the time-consuming task of annotating images and rather learn a model of normality from healthy images to later detect any outliers as pathological. Auto-encoders are deep architectures that play a fundamental role in unsupervised anomaly detection [110].

Introduced in the 1980s by Hinton and the PDP group [111], an auto-encoder is an unsupervised neural network that is trained to attempt to copy its input into its output. They work by compressing the input into a compressed representation, and then reconstructing the output from this representation. They can be viewed as consisting of two parts: an encoder and a decoder.

The encoder is a mapping that is applied on a input $\mathbf{x} \in \mathcal{R}^d$ in order to transform it to a latent representation $\mathbf{z} \in \mathcal{R}^{d'}$ where d' is typically smaller that d (Figure 5.1-A). It is generally modeled with a non-linear function (i.e. sigmoid, ReLu, etc.) applied to the affine transformation of the input:

$$\mathbf{z} = f(W\mathbf{x} + \mathbf{b}) \tag{5.1}$$

where W is a $d' \times d$ weight matrix, $\mathbf{b} \in \mathcal{R}^{d'}$ is a bias vector associated with the mapping and f is the (non)-linear function of choice.

The decoder is the inverse mapping from the hidden representation space to the original space with a similarly configured transformation:

$$\hat{\mathbf{x}} = g(W'\mathbf{z} + \mathbf{b}') \tag{5.2}$$

The difference between the original input data and its reconstruction is the reconstruction error that has to be minimized during training. The parameter sets $\theta = \{W, b\}$ and $\phi = \{W', b'\}$ are optimized through back-propagation during a series of epochs as to minimize the reconstruction error across all the instances of a given data set D composed of N examples. The loss function can be defined as:

$$\mathcal{L}(\mathbf{D}; \theta) = \frac{1}{N} \sum_{i=1}^{N} c(\mathbf{x}_i, \hat{\mathbf{x}}_i) \tag{5.3}$$

where $\mathbf{x}_i$ is the i-th example, $\hat{\mathbf{x}}_i$ is its reconstruction and c is the chosen measure of the reconstruction error i.e. the absolute error or the mean squared error.

Auto-encoders can be ***stacked*** in a layer-wise matter where a sequence of layers performs the encoding and another sequence decodes the middle-layer representation (Figure 5.1-B). In a stacked auto-encoder, the latent representation of one layer serves as input of the next one :

$$\mathbf{z}^k = f(W^k \mathbf{z}^{k-1} + \mathbf{b}^k)$$

This allows the model to learn more complex mappings and, hence, more abstract representations in the middle layer.

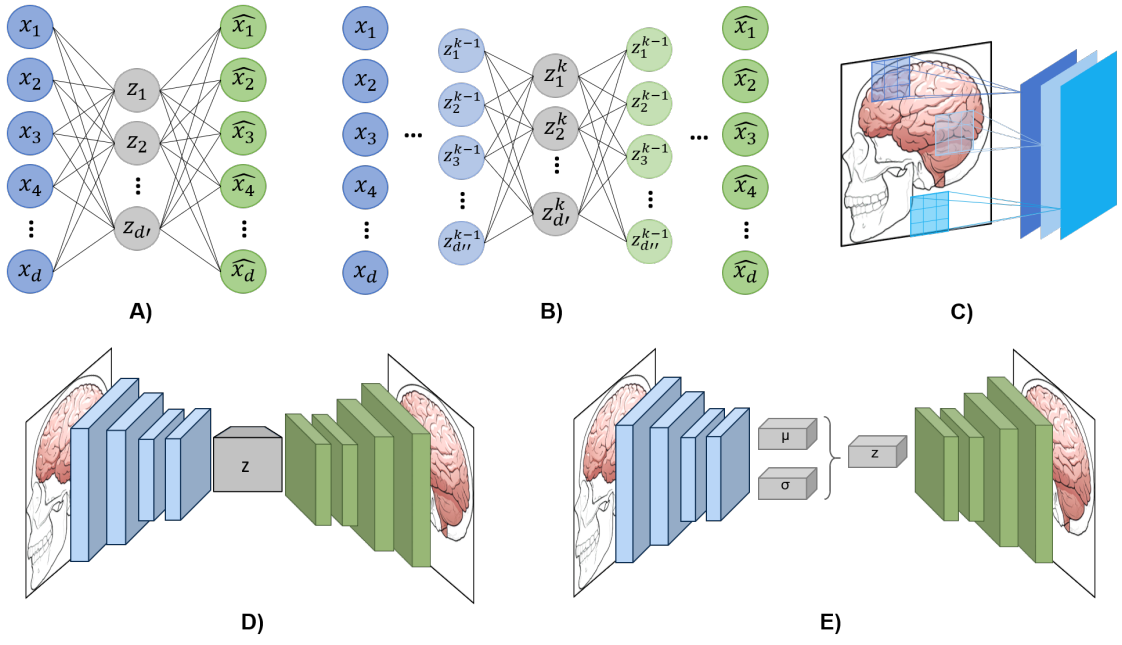


FIGURE 5.1: Different types of auto-encoders: A) Simple, fully-connected; B) Stacked, fully-connected; D) Stacked, convolutional; E) Stacked, convolutional, variational. A representation on the sliding kernel technique employed in convolutional layers is presented in C).

The main disadvantage of fully-connected auto-encoders is that they ignore the spatial relationships present in 2D or 3D images and are prone to learning redundant features. This

issue can be overcome by considering ***convolutional auto-encoders*** instead (Figure 5.1-C-E). This type of auto-encoders exploit the convolution operator and shared weights so as to learn features that are common in various locations of the image. More precisely, the input image x is mapped to a number of feature maps and each feature map $\mathbf{z}^k$ is computed as:

$$\mathbf{z}^k = f(\mathbf{x} * W^k + \mathbf{b}^k),$$

where W^k is the shared weight matrix, $\mathbf{b}^k$ is the shared bias, f is a chosen activation function and $*$ denotes the convolution operation. The reverse mapping to the input space is achieved through a "deconvolution" or more appropriately a transposed convolution. The parameters are tuned to optimize, as in the case of basic auto-encoders, a chosen loss function. Naturally, convolutional auto-encoders too can be stacked to form more complex architectures.

The latent space dimension for the hidden representation is crucial: too few dimensions may cause important information loss, whereas too many dimensions will prevent the model from filtering out redundant information. It is desired that the hidden units capture features that well represent the input data. As a matter of fact, Baur & colleagues [11] showed that using a fully convolutional encoder-decoder architecture in order to preserve spatial information in the latent space, i.e. having $z \in \mathcal{R}^{h \times w \times c}$, improves the anomaly detection performance.

Another point that deserves special attention is the regularity of the latent space in the auto-encoder. For the moment, the auto-encoder is solely trained to encode and decode with as few loss as possible, without any regard to how the latent space is organised. This enables the network to take advantage of any overfitting possibilities to achieve its task, implying that there are some points in the latent space that are "overused" and others that give meaningless content once decoded.

So, in order to be able to benefit from the generative properties of auto-encoders, explicit regularisation can be introduced during the training process to ensure that the latent space is regular enough. This constitutes a ***variational auto-encoder (VAE).***

Based on an auto-encoder, a VAE is trained to minimise the reconstruction error between the produced and the original data. However, instead of producing a single point, the encoder portion in the VAE produces a distribution over the latent space. Subsequently, a point from the latent space is sampled from that distribution to be decoded. The resulting reconstruction error is backpropagated through the network.

In other words, given training data $\mathbf{x}$, the model is trained to learn a distribution $p_\theta(\mathbf{x})$ using a latent representation model :

$$p_\theta(\mathbf{x}) = \int p_\theta(\mathbf{x}|\mathbf{z}) p_\theta(\mathbf{z}) dz$$

where $p_\theta(\mathbf{x})$ is also called the marginal likelihood, $\mathbf{z}$ is a latent variable of lower dimension than $\mathbf{x}$, $p_\theta(\mathbf{z})$ is its prior distribution and θ denotes the parameters of the encoder.

Since the posterior distribution $p_\theta(\mathbf{z}|\mathbf{x})$ is intractable, the encoder is rather learning an approximation $q_\phi(\mathbf{z}|\mathbf{x})$ while the decoder is generating data from $p_\theta(\mathbf{x}|\mathbf{z})$ where ϕ represents the parameters of the decoder.

The objective of the VAE is to maximize the log-likelihood of a given dataset, where for any observation $\log p_\theta(\mathbf{x})$ can be written as :

$$\begin{aligned}
\log p(x) &= \mathbb{E}_z \log \frac{p_\theta(\mathbf{x}|\mathbf{z})p_\theta(\mathbf{z})}{p_\theta(\mathbf{z}|\mathbf{x})} \\
&= \mathbb{E}_z \log \left(\frac{p_\theta(\mathbf{x}|\mathbf{z})p_\theta(\mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})} \frac{q_\phi(\mathbf{z}|\mathbf{x})}{p_\theta(\mathbf{z}|\mathbf{x})} \right) \\
&= \mathbb{E}_z \log p_\theta(\mathbf{x}|\mathbf{z}) - \mathbb{E}_z \log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{p_\theta(\mathbf{z})} + \mathbb{E}_z \log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{p_\theta(\mathbf{z}|\mathbf{x})} \\
&= \underbrace{\mathbb{E}_z \log p_\theta(\mathbf{x}|\mathbf{z}) - D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z}))}_{(1) ELBO} + \underbrace{D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z}|\mathbf{x}))}_{(2) D_{KL} \geq 0}
\end{aligned} \tag{5.4}$$

where the first term is known as the variational lower bound of the marginal log-likelihood, also called ELBO (evidence lower bound) and the second term is the Kullback-Leibler divergence between the approximated posterior $q_\phi(\mathbf{z}|\mathbf{x})$ from the true posterior $p_\theta(\mathbf{z}|\mathbf{x})$.

Because $D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z}|\mathbf{x}))$ is non-negative and only zero if $q_\phi(\mathbf{z}|\mathbf{x})$ equals the true posterior distribution, maximizing $\log p(x)$ boils down to maximizing ELBO which by consequence corresponds to the loss function of VAE:

$$\mathcal{L}_{vae}(x) = \mathbb{E}_z \log p_\theta(\mathbf{x}|\mathbf{z}) - D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z})) \tag{5.5}$$

One common setting is to consider that $p_\theta(\mathbf{x}|\mathbf{z})$ is a Gaussian distribution and have the VAE encoder output its mean μ and standard deviation Σ which are then used to draw $\mathbf{z} = \mu + \Sigma \cdot \epsilon$ of dimensionality $M \in \mathbb{R}$. In this case, the KL-divergence has the following closed form :

$$D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z})) = -\frac{1}{2} \sum_{m=1}^{M} [1 + \log(\Sigma_m^2) - \mu_m^2 - \Sigma_m^2]$$

As it is not possible to backpropagate through a sampling operation, the introduction of ϵ is a fundamental characteristic of VAE, called the "re-parameterization trick". The sampling operation is effectuated on ϵ $\mathcal{N}(0, 1)$ which liberates μ and Σ for training. We note that $p_\theta(\mathbf{x}|\mathbf{z})$ can take the form of a multivariate Gaussian distribution and even a mixture of Gaussians, depending on the application.

As demonstrated, autoencoders are very flexible structures, besides being stacked, convolutional and variational, they can take on very simple and very complex data of many dimensions (as long as there is enough data available) and they can even be equipped with a discriminator, a classifier or other statistical tools.

Regardless of the autoencoder architecture, the anomaly detection procedure is virtually the same: First, the model is trained solely on data instances considered "normal". Then, the model is tested with a "normal" and "abnormal" instances. The individual reconstruction error at voxel level $|x_i - \hat{x}_i|$ may be used to compute an anomaly score and thus any sample with a reconstruction error above a certain threshold (t), it is considered to be anomalous. Indeed, large reconstruction errors are expected when trying to reconstruct anomalous data samples while all other samples should be well reconstructed. Finally, different metrics can be used to evaluate the effectiveness and efficiency of the implemented autoencoder for in its anomaly detection application.

The effectiveness in anomaly detection can be evaluated according to the quantity of anomalous data points that are correctly identified (true negative rate) and the number of normal data that are misclassified as anomalous (false negative rate). Notably, receiver operating characteristic (ROC) curves can be used to represent the trade-off between specificity and specificity.

5.2 Pipeline description

The application presented on this chapter searches to identify abnormalities in the diffusion MR data of newly diagnosed PD patients. Several studies have already looked for biomarkers in diffusion MR data, however, the presence of inconsistent findings, as shown by [56] and [59], indicates the need of a robust methodology to resolve such discrepancies.

Here, we propose the use of fully convolutional auto-encoders and variational auto-encoders to provide a meaningful representation of quantitative MRI data from healthy brains. Once these models are trained, they are used to identify unusual patterns in data from *de novo* PD patients.

This is not without challenge as we do not dispose for imaging data of a ground truth that identifies, at the voxel level, values specific to PD, especially for *de novo* patients. This complicates model evaluation and selection. The only available information is the classification of each individual as PD patient or healthy control. In addition, PD related abnormalities may be very local and particularly subtle to detect in *de novo* patients.

Despite the rising interest in uncovering biomarkers of PD and the advent of Deep Learning for anomaly detection, no studies have been published, to the best of our knowledge, concerning this application.

5.2.1 Data pre-processing

Like in Chapter 4, the data used in this work comes from the PPMI database [6], only this time we extracted DTI MR scans.

DTI data are particularly sensitive to various sources of bias [112]. For this reason, we pooled DTI data from 57 healthy controls and 129 *de novo* PD patients acquired with the same MR scanner model (3T Siemens Trio Tim) and configured with the same acquisition parameters (90° flip angle, 64 gradient directions, EP Pulse sequence, TE=88ms).

From these images, two measures per voxel were computed using MRtrix3.0 [10]: mean diffusivity (MD) and fractional anisotropy (FA). As mentioned in Chapter 2, these measures describe the diffusion of water molecules in the brain, MD accounts for the their overall displacement and FA indicates the orientation of diffusion.

Each MR feature map was obtained in the NIfTI format as a 3-dimensional array. The images were normalized into the MNI space with a non-linear deformation. Only one healthy control was taken out of the study due to important artifacts in the images.

Even though it is possible to train auto-encoders with 3D images, 56 controls is not enough data to train the models, specially since part of the data needs to be set aside for testing. This is why we decided to train the auto-encoders with 2D axial slices of our 3-D scans, keeping only 40 slices around the center of the brain to avoid training the models with empty or almost-empty images. Additionally, we opted to treat each hemisphere as different subject. In practice, the right and left hemisphere were segmented separately in two volumes, then the left hemisphere was flipped to resemble the right hemisphere and both volumes were cropped. The process, was done in the exact same way for FA and MD images as can be seen in Figure 5.2.

The control dataset was divided into 41 training controls and 15 testing controls to avoid data leakage. As a result the models were trained with a duo (FA & MD) of 1680 images (2 hemispheres $\times$ 40 slices $\times$ 41 controls). On the other hand, the test patients dataset is constituted by a duo of 10320 instances and the test controls dataset includes a duo of 1200 images. All hemisphere slices measure 75×145 voxels.

To accrue the generalization of our results to other populations we implemented a 10-fold cross-validation to train and test our results. That is, formed 10 different combinations of train and test controls, taking special care to maintain similar age profiles and 40 % of females in both datasets and along the ten samples as summarized in Table 5.1. For every sample, the values of FA and MD were normalized into the range $[0, 1]$ using the minimum and maximum values within the corresponding training samples.

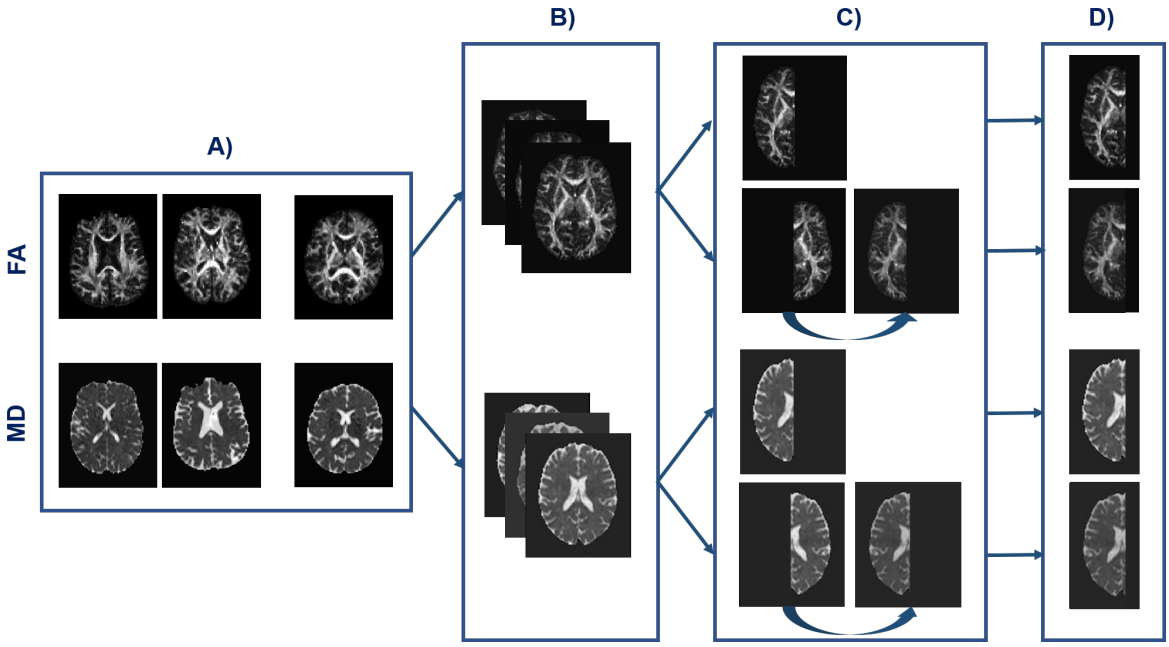


FIGURE 5.2: Pre-processing procedure : A) MD and FA parameter maps are extracted from DTI scans; B) Normalization of the volumes into the MNI space and selection of 40 axial slices around the center of the brain; C) Separation of the two brain hemispheres, the left hemisphere is flipped to resemble right hemisphere; D) All images are cropped to a 75 $\times$ 145 image size.

	Train controls	**Test controls**
Sample 1	60.9 ± 10.5	61.5 ± 8.2
Sample 2	61.3 ± 9.2	60.7 ± 11.9
Sample 3	61.2 ± 10.2	60.9 ± 9.3
Sample 4	61.0 ± 10.0	61.3 ± 9.8
Sample 5	61.1 ± 9.5	61.1 ± 11.2
Sample 6	61.1 ± 9.5	61.0 ± 11.2
Sample 7	61.1 ± 10.2	61.1 ± 9.2
Sample 8	61.1 ± 10.6	61.1 ± 7.8
Sample 9	61.3 ± 10.1	60.4 ± 9.5
Sample 10	61.1 ± 10.3	61.1 ± 9.0

TABLE 5.1: Ages of the control train and test sub-populations across the 10 different samples.

5.2.2 Auto-encoder architectures design

Three autoencoder-based models were developed and evaluated: a spatial autoencoder (sAE), a spatial variational autoencoder (sVAE) and a dense variational autoencoder (dVAE).

We extended the work of [11] by using multiple quantitative MRI measures simultaneously as input. Thus, every input can be expressed as $x \in \mathbb{R}^{H \times W \times C}$, where H is the height, W the width and C the number of channels (FA and MD).

The design of the architectures was inspired from the first pathway of the DeepMedic [113]

network and other 5 layers deep networks [114]. As seen in Figure 5.3, depending on the network architecture, the latent code may be a simple vector ($z \in \mathbb{R}^d$) or a third-order tensor ($z \in \mathbb{R}^{h \times w \times c}$). The former is referred as a dense bottleneck and the latter as a spatial bottleneck.

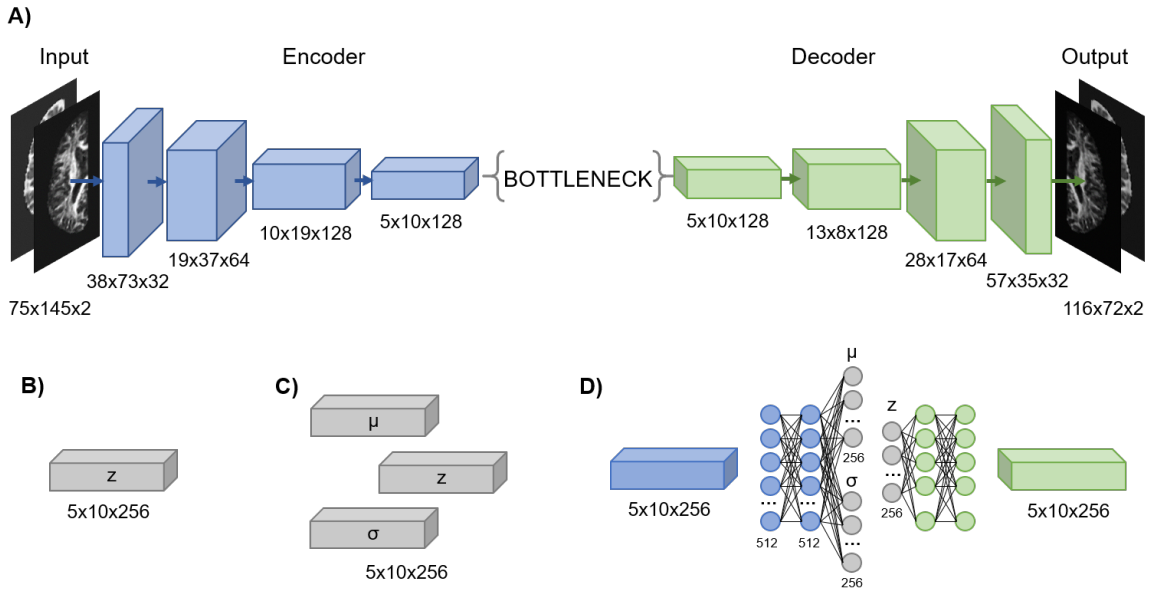


FIGURE 5.3: A) The general architecture of the implemented autoencoders with an unspecified bottleneck; B) sAE spatial bottleneck; C) sVAE spatial bottleneck and D) fourth convolutional layer of the dVAE along with its fully connected layers and dense bottleneck. μ and σ describe the approximate posterior of the latent variable, z is obtained by a sampling operation.

All models were implemented using Python 3.6.8, PyTorch 1.0.1, CUDA 10.0.130 and trained on a NVIDIA GeForce RTX 2080 Ti GPU with batches of 40 images. After each convolutional layer, batch normalization [115] was applied for its regularization properties. The nonlinear activation function in each layer was the rectified linear unit (ReLU), except for the last layer where a sigmoid was employed in order to have output pixels normalized between $[0, 1]$. The loss functions were optimized using Adam [116], a popular optimization algorithm for training deep neural networks.

In order to choose the best network architecture and tune the corresponding hyper-parameters, the three models were trained and assessed with a 7-fold cross-validation employing 35 MR volumes used for training and 6 MR volumes for validation in each fold. We note that we did not perform a nested cross-validation, and thus, this parameter tuning step was only effectuated for the fist sample population. Nevertheless, the learning behavior of the models remained consistent across all the different samples for the three models.

5.2.2.1 Spatial Auto-encoder (sAE)

The spatial autoencoder model is fully convolutional, 5 convolutional layers go from input to bottleneck and 5 transposed convolutional layers from bottleneck to output. As depicted in Figure 5.3-B, the output of the encoder network is directly the latent vector z and loss function is simply the absolute difference between output and input:

$$\mathcal{L} = \|x - \hat{x}\|_1 \tag{5.6}$$

This model was trained for 200 epochs, with a learning rate of 0.9×10^{-3}. 3×3 kernels were convolved using padding of 1 pixel and a stride of (2, 2) for the first 4 layers and a stride of (1,1) for the last one. There were no pooling layers.

5.2.2.2 Spatial Variational Auto-encoder (sVAE)

Our spatial variational autoencoder model is shown in Figure 5.3-C. Similar to sAE, the model is fully convolutional, however, the encoder generated the parameters of the approximate posterior of the latent variable given the input, constrained to follow a multivariate normal distribution. A sampling operation was needed to obtain an actual value for z.

Training lasted for 200 epochs using a learning rate of 0.3×10^{-3}. A 3×3 kernel was chosen as filter, along with a padding of 1 and stride of (2, 2) for the first 4 layers and a stride of (1,1) for the last one. No pooling layers were used. The loss function was computed as follows:

$$\mathcal{L} = \lambda\|x - \hat{x}\|_1 + (1 - \lambda)\Big[-\frac{1}{2}\sum_{j=1}^{J}(1 + \log((\sigma_j)^2) - (\mu_j)^2 - (\sigma_j)^2)\Big] \tag{5.7}$$

where μ and σ denote the mean and the variance of the approximate posterior, J is the number of dimensions of the latent space and λ controls the proportions between the two terms. The first term is the reconstruction error and the second term is the Kullback-Leibler (KL) divergence between the approximate posterior and the prior of the latent variable, for the Gaussian case [117]. To favor good reconstructions over a Gaussian-like distribution of the latent variables, we put more weight (90%) in the reconstruction term and less weight (10%) in the KL divergence term.

5.2.2.3 Dense Variational Auto-encoder (dVAE)

The main difference of the dense variational autoencoder when compared to the sVAE is its dense bottleneck. Encoder and decoder also have fully connected layers in addition to the convolutional layers, as shown in Figure 5.3-D.

For regularization purposes, the dropout [118] technique was used to turn off 30% of the units in fully-connected layers during training. This model was trained for 100 epochs with a learning rate of 0.3×10^{-3}. There were no pooling layers. Kernels for all convolutional layers were 3×3, and convolutions were performed with a padding of 1 and stride of (2, 2) for all layers except the last one. The dVAE shared the same loss function as the sVAE and we kept the same 90/10 proportion between the reconstruction term and the KL divergence term.

5.2.3 Evaluation procedure

Since the networks were solely trained on healthy subjects, we hypothesized that the MR scans of the PD population would have greater reconstruction errors in some regions. The idea is therefore to use the reconstruction error as an anomaly score.

The voxel-wise reconstruction errors in one image can be computed as $|x_i - \hat{x}_i|$. Seeing that our decoders output two images ($\hat{\text{FA}}, \hat{\text{MD}}$), we defined the joint reconstruction error of every voxel as:

$$e_i = \sqrt{(\text{FA}_i - \hat{\text{FA}}_i)^2 + (\text{MD}_i - \hat{\text{MD}}_i)^2} \tag{5.8}$$

We identified four sources of reconstruction errors :

1. Noise in the input data.
2. Loss of information due to dimension reduction in the latent space.
3. Variability of healthy controls not captured by the model.
4. PD-related anomalies.

Because we were only interested in anomalies caused by PD, the best way to evaluate and compare the models is by measuring their ability to discriminate between controls and PD patients, based on the intensity and localization of the reconstruction errors.

We investigated extreme reconstruction errors with the idea that PD patients should exhibit very abnormal voxels in larger quantities. Accordingly, we considered an extreme quantile (eg. the 98% quantile) of the distribution of errors in the control population as threshold value to decide whether or not a given voxel was considered as abnormal, hereafter called the ***abnormality threshold***.

Next, for each control and PD subject, we counted the number of extreme abnormalities detected in every structure as a percentage. The idea being to classify a subject as PD or healthy when this number was above a certain value. The critical choice of this value was investigated using a Receiver Operator Curve (ROC) of sensitivity and specificity.

Truly, the imbalanced nature of the test set (30 healthy and 258 PD brain hemispheres) demanded an evaluation criterion that is well adapted to evaluate performance in each class, independently from class prevalence. The trade-off between sensitivity and specificity depicted by the ROC curve deemed itself the best adapted. What is more, the area under the ROC curve (AUC) can be directly used as a measure of classification performance.

However, to classify any future patients or controls, the following question must be answered: how much abnormal voxels are typical in the healthy population ? As every point in the ROC curve corresponds to the sensitivity and specificity values obtained by a given threshold, the point closest to a perfect sensitivity and specificity can be considered as the percentage of abnormality threshold, hereafter called, the ***pathological threshold***. This is illustrated on Figure 5.4.

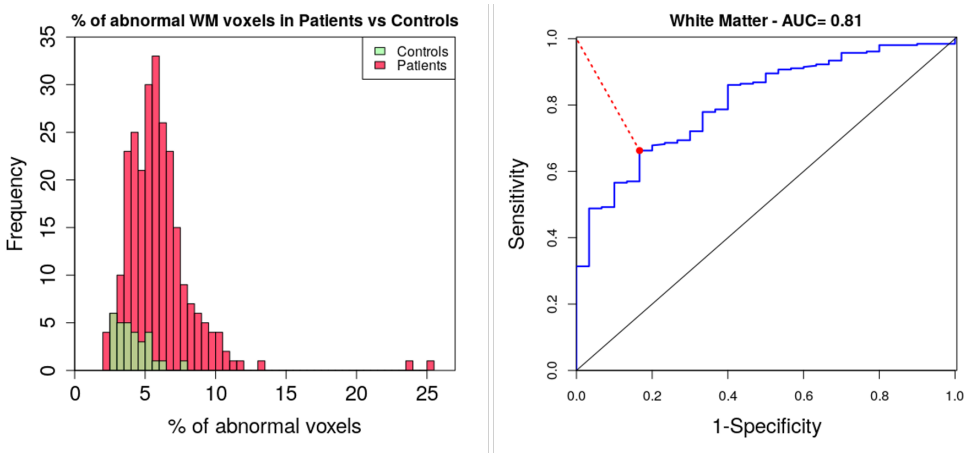


FIGURE 5.4: Histogram of the percentage of WM voxels considered as abnormal for the controls and the patients (left) and the corresponding ROC curve (right). The pathological threshold is indicated by a red point.

And so, the ***abnormality threshold*** serves for classification at the voxel-level, while the ***pathological threshold*** differentiates subjects at the hemisphere-level or ROI-level. The classification performance of the pathological threshold can be evaluated using the geometric mean (g-mean) between sensitivity and specificity :

$$G - mean = \sqrt{Sensibility \times Specificity} \tag{5.9}$$

Additionally, to help evaluate the **localization of anomalies**, two atlases were applied: the Neuromorphometrics atlas [119] and the MNI PD25 atlas [120]. The first is used to segment the brain into 8 macro-regions: Subcortical structures, White matter and the 5 gray matter lobes (Frontal, Temporal, Parietal, Occipital, Cingulaire/Insulaire). The latter was specifically designed for PD patient exploration. It contains 8 regions: substantia nigra (SN), red nucleus (RN), subthalamic nucleus (STN), globus pallidus interna and externa (GPi, GPe), thalamus, putamen and caudate nucleus. Both atlases are depicted in Figure 5.5

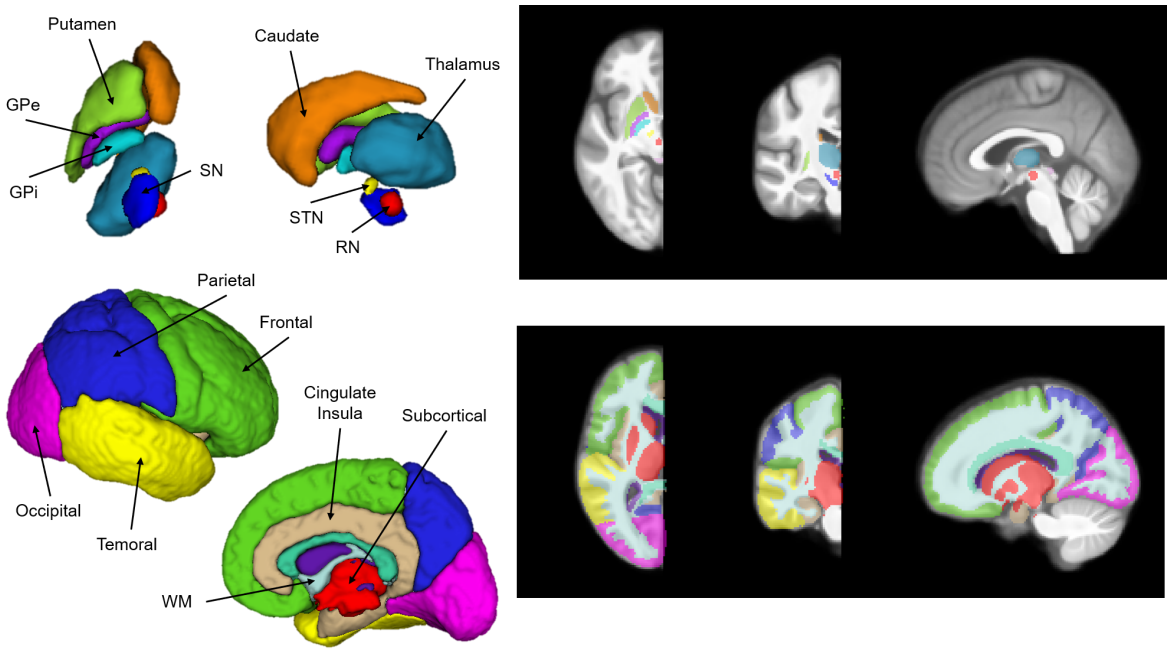


FIGURE 5.5: MNIPD25 (top) and Neuromorphometrics (bottom) modified atlases to segment subcortical and brain regions on the subject's hemispheres.

For all of the before-mentioned regions of interest (ROI), we calculated the ROC-AUC and the g-mean of the associated pathological threshold. The results across all sample populations for all models are featured in the following section.

Next arises the question of how would this pipeline perform when classifying new incoming subjects. In an effort to begin to answer this question, we made use of the test datasets of every sample, and employed a logistic regression onto the percentage of abnormal voxels in the individuals for the whole hemisphere and in every ROI studied. In a 5-fold stratified cross-validation the logistic regression was fitted onto the 80% of the dataset and predicted the type of the remaining 20% (control or patient). The g-mean score was once again chosen to measure the quality of the predictions. Different tests were realized to examine the performance of different ROI combinations.

5.3 Results

As it can be observed on Figure 5.6 , the best reconstructions are achieved by the sAE, which is expected as there are not any regularization constraints for the reconstructions. The sVAE generates quite good reconstructions as well, on the contrary, the dVAE is unable to reconstruct fine details in the images, specially around the circonvolutions and the subcortical structures. This may be caused by the loss of information in its fully connected layers.

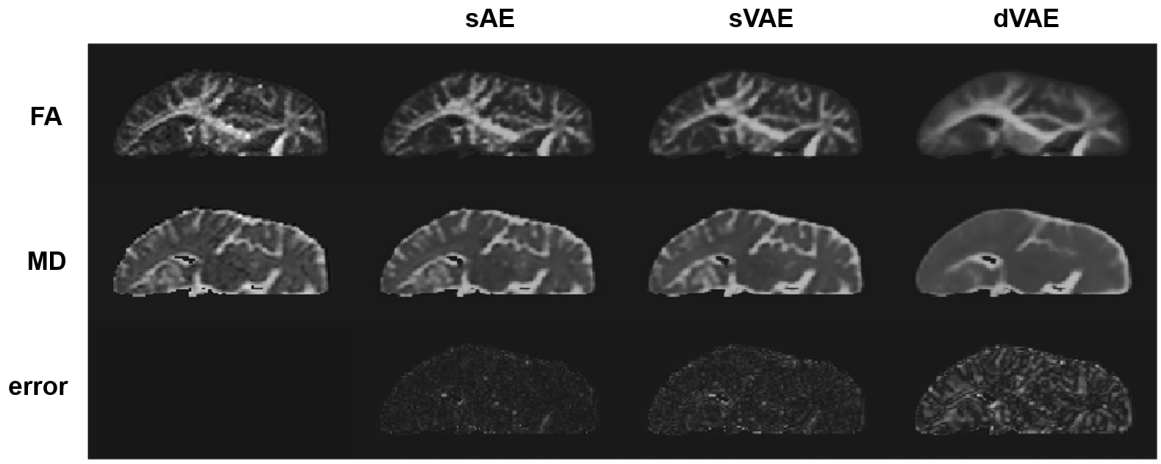


FIGURE 5.6: Example of an FA (top) and MD (middle) reconstruction of a brain hemisphere (first column) from the sAE, sVAE and dVAE trained architectures. The compound reconstruction error of the two feature maps is also featured (bottom).

The first performance evaluation corresponds to the ROC-AUC obtained when comparing, for each ROI, the percentage of values above the 98% quantile of the distribution of reconstruction errors originated by healthy controls. The results displayed in Figure 5.7 show that comparing the quantities of abnormal voxels in the whole hemisphere yields satisfactory results with an average AUC of 67.6%, 67.5% and 63% for the sAE, sVAE and dVAE respectively.

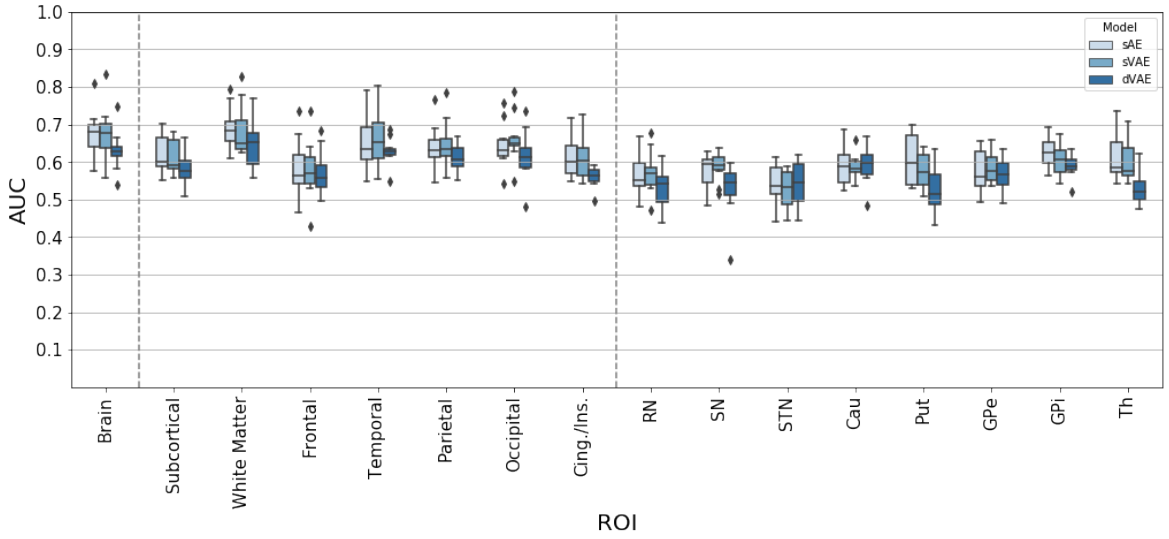


FIGURE 5.7: Obtained ROC-AUC scores for the whole brain hemisphere (first) and several ROIs.

Due to the lack of ground truth, we cannot say that a model is superior to one another but for the rest of the study we focus on the scores obtained by the sAE and take those of the sVAE as a sign of robustness.

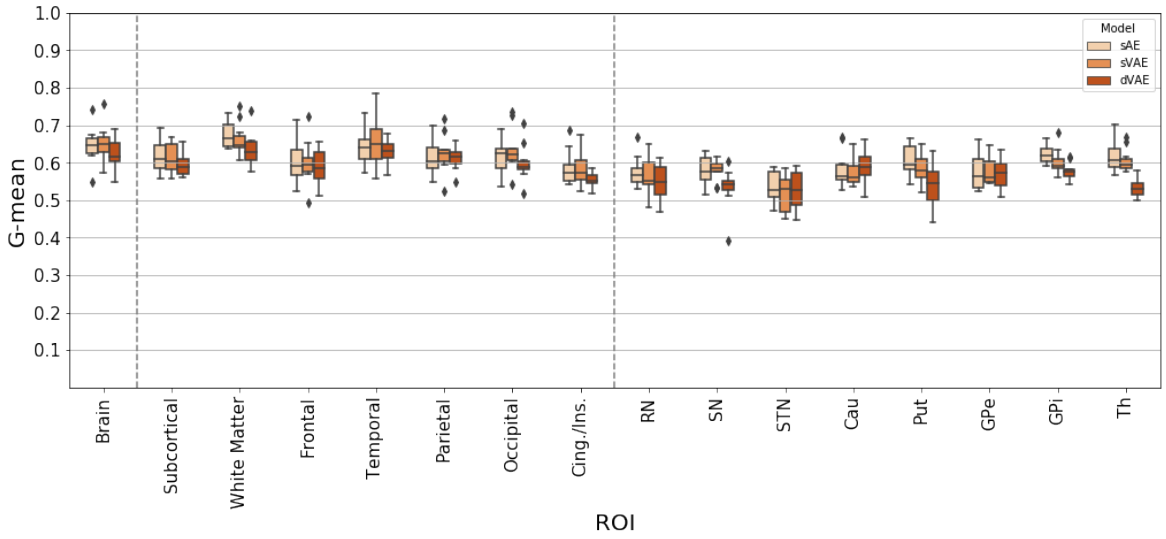


FIGURE 5.8: Obtained G-mean scores for the whole brain hemisphere (first) and several ROIs.

The ROC curves belonging to very population sample throw an specific pathological threshold from which the associated g-mean score can be computed. For example, in average, we expect that any hemisphere sAE reconstruction with more than 2.20% of abnormal voxels (see 5.2.3), most likely belongs to a PD patient. The averages of the ROC-AUC scores, the retained pathological thresholds and their associated g-mean scores are presented on Table 5.2.

Stucture	ROC-AUC (%)	P. threshold (%)	G-mean (%)
Brain	67.6	2.18	64.6
Subcortical	62.0	1.02	61.2
White Matter	69.0	4.24	67.4
Frontal	58.4	1.02	60.4
Temporal	65.1	1.70	64.4
Parietal	64.7	0.80	61.1
Occipital	64.4	1.98	61.7
Cing./Ins.	61.0	1.66	58.5
RN	56.7	9.21	57.5
SN	57.6	20.96	57.7
STN	54.1	19.35	53.6
Cau	59.1	1.03	58.3
Put	60.5	1.08	60.4
GPe	57.6	2.91	57.6
GPi	62.5	5.11	62.2
Th	60.9	1.54	61.7

TABLE 5.2: ROC-AUC scores obtained from sAE reconstructions, the pathological threshold retained from the ROC curves and the associated g-mean scores. All in percentages

Still, it is interesting to observe the variation of the g-mean classification scores obtained for every ROI as presented in Figure 5.8

We notice that the macro-structures that obtained the highest scores for the sAE reconstructions are the White Matter and the Temporal lobe with a g-mean average of 67.4% (p.t. 4.24%) and 64.4% (p.t. 1.7%) respectively and that within the subcortical structures, it is the internal part of the Globus Pallidus (GPi) and the Thalamus that achieved the best performances in average with a g-mean score of 62% for both. This being said, the variation of the temporal lobe scores is quite important (std: 5.6%) compared to the other four ROIs. What is more, the g-mean score variation associated to the Temporal lobe sVAE reconstructions is the highest one out of all ROIs with a value of $65.1 \pm 6.6\%$. For comparison, the White Matter obtained a g-mean score (sVAE) of $66.2 \pm 4.3\%$, the GPi, $60.1 \pm 3.5\%$ and the Thalamus, $60.5 \pm 3.2\%$.

In order further exploit the good classification results of the white matter, we used the John Hopkins University (JHU) WM atlas [121] to search for a specific ROI were abnormalities were clustered. The regions of the atlas that were mainly concentrated on the midline of the brain were suppressed to keep only regions that were well represented in both hemispheres (Figure 5.9), there remained: the three segments of the corps callosum (genu, body and splenium); the cerebral peduncle; the anterior, posterior and retrolenticular limb of the internal capsule (AIC, PIC, RlIC); the external capsule; the anterior, superior and posterior part of the corona radiata (ACR, SCR, PCR); the posterior thalamic radiation (PTR) including the optic radiation; the cigulate gyrus; the hippocampus; the fornix; the tapetum; the superior longitudinal fasciliculus (SLF); the uncinate fascilicus (UF); the superior fronto-occipital fasciliculus (SFOF); and the sagittal stratum, which includes the inferior longitidinal fasciculus and inferior fronto-occipital fasciculus.

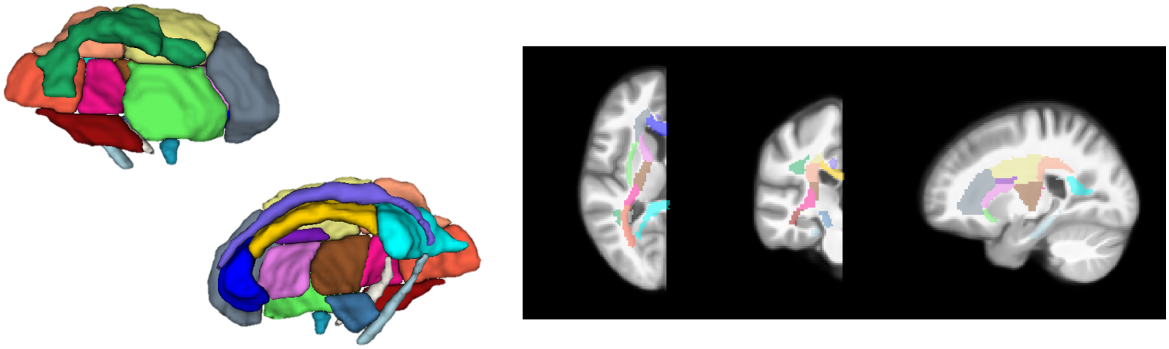


FIGURE 5.9: White Matter John Hopkins atlas in one hemisphere

We found that the sagittal stratum, the SLF and the PTR had the best sAE g-mean results with averages of 63, 62 and 62% accordingly.

5.3.1 Performance predictions

The results of our 5-fold stratified cross-validation employed to predict the performances that our pipeline could produce on new data showed that the threshold on the whole hemisphere has the best performance with and average g-mean of 61%, closely followed by white matter (60.7%) and the combination of the 5 gray-matter lobes (59.3%). The combination of several smaller ROIs yielded similar results as shown in Figure 5.10. ROI 1 is the combination of all regions in the JHU atlas, ROI 2 is composed of 11 white matter regions found near the subcortical structures of the brain (AlIC, PlIC, RlIC, ACR, SCR, PCR, PTR, sagittal stratum, hippocampus, SLF, SFOF), ROI 3 is constituted by 3 subcortical structures : the GPi, the thalamus and the SN, ROI 4 includes ROI 3 plus the AlIC, PlIC, SCR and PTR, these 4 regions surround the three subcortical structures. Finally, ROI 5 is composed of the sagittal stratum, the SLF and the PTR.

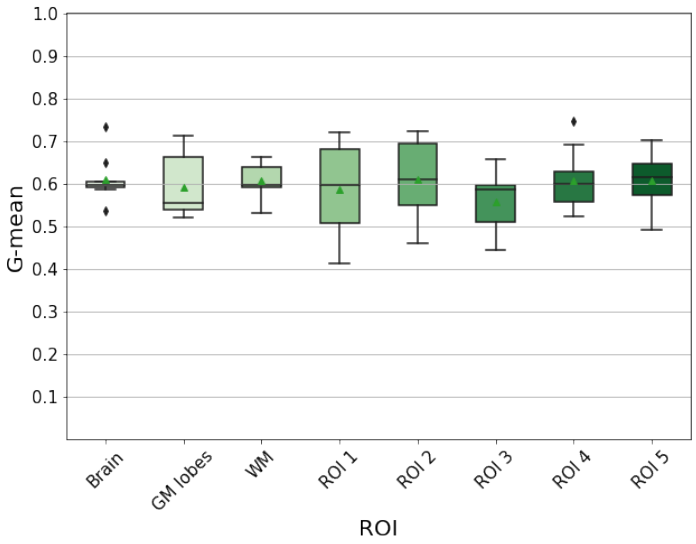


FIGURE 5.10: G-mean prediction scores for the whole brain hemispheres, the GM lobes, the White Matter and 5 composed ROIs.

While on average, all g-mean scores oscillate around 60%, ROI 3 exhibits the worst score (55.7 ± 6.5 %), ROI 1 has the greatest standard deviation (58.7 ± 9.9 %) and the whole brain hemisphere the smallest (61.0 ± 4.8 %).

As seen in Chapter 3 some classification methods are more successful in detecting anomalies in PD patients that present mild cognitive impairment (MCI) versus patients without cognitive symptoms [47, 48, 24, 62]. To test for this eventuality in our dataset, we divided the patients into the two categories based on their MoCA test scores. The patients with a score below 26 points were classified as MCI and the rest as "normal cognition".

A secondary classification was made based on their motor symptoms. Indeed, PD often presents unilateral symptoms in the early stages of the disease, and thus, following the inspiration of other studies [12, 89] we studied the prediction differences on the ipsilateral (the side where symptoms are presented) and contralateral (the opposite side) hemispheres

based on the UPDRS III test scores. For many patients, however, motor symptoms are presented on both sides and so an hemisphere was considered as symptomatic if the sum of the UPDRS III scores related to the hemisphere is equal or above 10 points.

The logistic regression cross-validation procedure was repeated for these four subsets of patients alongside with the combination of patients that have MCI and important motor symptoms. The results are displayed on Figure 5.11.

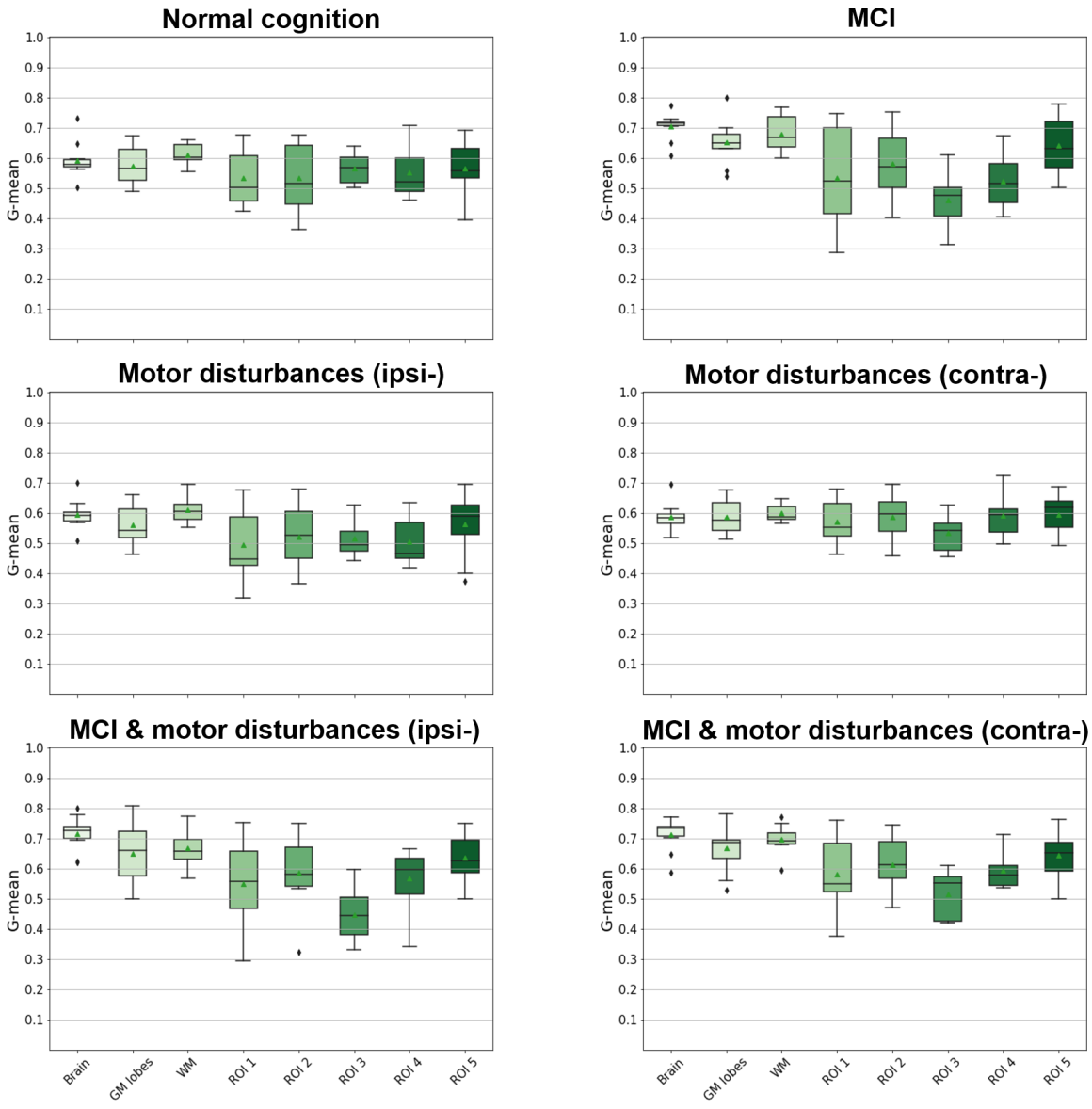


FIGURE 5.11: Logistic regression prediction scores for different subsets of patients hemispheres regarding their cognitive symptoms (MCI or normal) and their motor symptoms.

5.3.2 Assessment at 48 months

While we are particularly interested on early diffusivity changes in PD patients, we chose to assess the efficacy of our methodology on discriminating between healthy controls and

PD patients 48 months after their diagnosis. In consequence, we extracted 41 advanced PD patients (Age: 63 $\pm$ 8.2, Sex: 26M, 16F) from the PPMI database. The FA and MD parameter maps per hemisphere were extracted with the same pipeline as the baseline patients and healthy controls (see Section 5.2.1).

The parameter maps of the new patients were reconstructed by the pre-trained autoencoder architectures and abnormal voxels were identified based on the same reconstruction error criteria as for the rest of the subjects (see Section 5.2.3). The discrimination between patients and controls is dependant on the percentage of abnormal voxels in different ROIs. For the ten train controls sub-populations, a logistic regression was trained with the information from healthy controls and PD patients at baseline, then tested on the data of the advanced PD patients.

The effect of cognitive symptoms was briefly explored as well, by training and testing the logistic regressions only with patients in the MCI group in one instance and only with cognitively compliant patients in another. The results are presented in Figure 5.12.

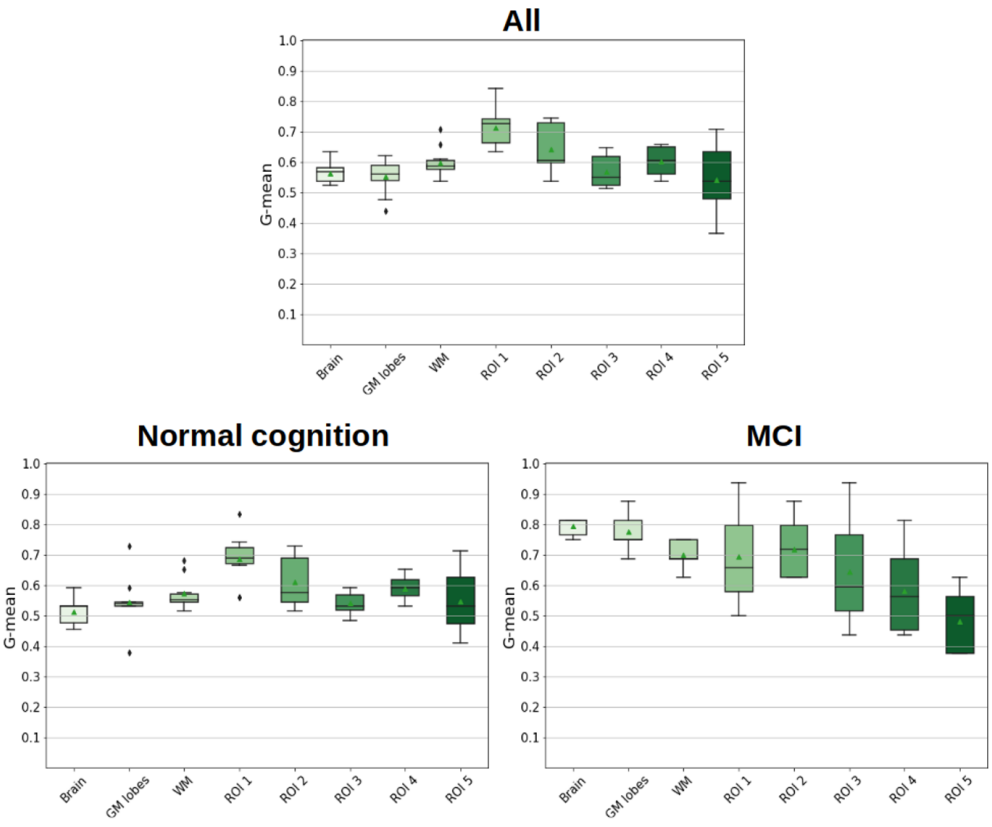


FIGURE 5.12: G-mean prediction scores for 41 patients 48 months after diagnosis.

5.4 Discussion

In this Chapter we designed and evaluated a pipeline for the detection of anomalies in the diffusion MR data of newly diagnosed PD patients. Fully convolutional auto-encoders and variational auto-encoders constituted the main building blocks our our approach.

We were able to appropriately train the architectures with a medium-size collection of samples (41 controls) and while preserving spatial information by employing 2D slices instead of 3D volumes and considering brain hemispheres instead of whole brain images. Indeed, classical techniques of data augmentation such as flipping, cropping and rotating images are irrelevant to our application as the goal is to learn the normal distribution of quantitative data. What is more, we found that the models were better suited to learn to reconstruct normalized images.

Our results show that the brain reconstructions of PD patients are significantly different than those of healthy controls for the sAE, sVAE and dVAE models. In our application, there is not any ground truth permitting to favor one model over the others, nevertheless, supervised learning applications of auto-encoders in the literature have shown that dVAE architectures struggle to reconstruct fine details in brain scans, for example, gyri [11].

An original evaluation procedure was conceived to compare the reconstruction error profiles of healthy individuals versus PD patients. The procedure was based on the hypothesis that the reconstructions of patients hemispheres should exhibit a largest quantity of ill reconstructed points and thus we compared the percentages of abnormal voxels in the brain hemispheres of our subjects. By defining the percentage that presents the best compromise between sensitivity and specificity as a pathological threshold, we transformed our dilemma into a binary classification problem for which we can calculate the discriminating and predictive performances. This procedure was followed for the whole brain hemisphere but also for specific regions of interest such as gray matter cortical lobes, subcortical structures and white matter structures.

With an unbalanced datasets we took special care in employing convenient metrics such as the geometric mean between specificity and sensibility, but also in generalizing our results by effectuating a 10-fold cross-validation to train and test our results. This turned out to be a key part in the study. As a matter of fact, for no obvious reason the first sub-population achieves much better performances than the average of our ten populations. For indication, the sAE g-mean score associated to the brain hemisphere is of 74.3% while the average is 64.6%.

We were also particularly cautious with our pre-processing. Indeed, on a previous study [122] we opted to extract our FA and MD tensor metrics using an older pipeline and obtained considerably better results in the subcortical structures. Nevertheless the parameter maps calculated with the current state-of-the-art software, MRtrix3, are visually better while the older maps displayed small artifacts that could have biased the study.

Seeing that DTI was designed to study white matter characteristics, it did not come as a surprise that some of the best performances were attained by the hemisphere's white matter and white matter structures, however, by observing the sAE pathological threshold inferred for small subcortical structures like the substantia nigra (20.96%) and the STN (19.35%) we suspect that the spatial auto-encoder was unable to produce accurate reconstructions of these regions and it would be beneficial to train the networks with patch images concentrating on basal ganglia in default to including more control subjects in the training population in order to better study diffusion abnormalities in the subcortical structures of PD patients with this method.

Like in other investigations, we find that patients displaying mild cognitive impairment are significantly easier to classify than cognitively compliant patients using the whole brain hemisphere, the combination of gray matter lobes or white matter. However, the combination of all JHU white matter structures or the 11 white matter regions in ROI 2 do not improve their performances, which may be an indication that the white matter anomalies in MCI patients are often toward the sagittal stratum, SLF and PTR (ROI 5) but generally diffuse. It is also remarkable that the g-mean prediction scores of the GPi, thalamus and SN (ROI 3) are better for the patients without cognitive symptoms that those with MCI. This may be an indication of the presence of two types of patients with different diffusion anomaly locations.

Our research also hints that motor disturbances mainly manifests as diffusion anomalies in the white matter of the contralateral hemisphere.

The predictive assessment on patients 48 months after diagnosis is difficult to interpret. We could expect that after two years some diffusion properties of the brain would had continued to degrade, however, all but one of these patients take at least one PD related medication. We can see an indication of degradation in the white matter structures surrounding the basal ganglia from the results of ROI1, ROI2 and ROI4. The whole hemisphere of patients with MCI ($8 \longrightarrow 16$ hemispheres) have the best g-mean scores from the whole study with an average of 79.6%. As the cognitive symptoms progress, this may be due to cortical atrophy [47, 24].

From a competitive standpoint, our approach achieved similar performances than other diffusion studies. Notably, the cross-validation procedure of Schuff and colleagues [12] obtained a ROC AUC of 59% for the rostral segment of the SN which is comparable to our ROC AUC of 57.6% for the complete SN. Then, Correia and colleagues [13] obtained a mean accuracy score of 59.7% through a cross-validation procedure on their SVM and this selected regions of white matter. This is in accordance to our average g-mean prediction score of 61% for 11 white matter regions (ROI2).

Oppositely, our results fall below those achieved by Li and colleagues [88] employing a stacked spatial autoencoder with GM, WM and mean diffusivity features for 116 ROIs as input. Indeed they got a ROC AUC of 86% but they did not indicate any cross-validation

or generalization procedure. Still, it is possible that the relationship between white matter quantity, gray matter quantity and mean diffusion brings out anomalies that are not significant when studying mophometric and diffusion properties separately.

Taken together, these results offer compelling evidence that deep learning-based models are useful to locate subtle anomalies, as found in *de novo* PD patients, even when trained with a moderate number of images and only FA and MD measures. Such models could be of interest for studying other neurological disorders when small lesions are suspected (such as mild traumatic brain injury) and difficult to localize for a human observer. We may expect that the insertion of additional quantitative MRI measures, such as perfusion, iron content and tissue relaxation time, would improve abnormalities detection. Moreover, the spatial localization of subtle alterations in MR imaging modalities, sensitive to different physiological parameters, could bring new knowledge about the physio-pathology of the underlying disease.

In future research we will seek to train a classifier with the latent vectors produced by the encoder part of the architectures to decrease the level of post-processing necessary. It would also be interesting to include clinical scores in the classification as a means to study phenotypes of Parkinson's disease.

5.5 Conclusion

Auto-encoders, and unsupervised deep anomaly detection techniques in general, are cost effective techniques that do not require annotated data for training. Instead, they learn the inherent data characteristics to define normal data behavior from which outlier detection is possible.

The main limitation of these methods in the medical field is the requirement of large amounts of data, specially when attempting to learn complex data, as they are sensitive to noise and data corruption. The presented pipeline for anomaly detection employing auto-encoders could be used as a blueprint to detect subtle anomalies for medical purposes with a moderate size dataset.

While our results do not highlight a particular diffusion biomarker for early PD, they give an indication on the presence of diffuse anomalies mainly located in the white matter. The addition of more training data and other parameter maps could not be but beneficial to our classification performances.

Chapter 6

Characterization of MR multivariate data using mixture models

6.1 A word on mixture models

Clustering is an unsupervised learning problem intended to group unlabeled examples sharing common characteristics within a dataset. Clustering has a myriad of uses in image analysis going from segmentation to anomaly detection. When presented with simple problems where clusters are unlikely to overlap, hard clustering methods like the k-means algorithm may suffice. However, it is often useful to obtain a measure of probability or likelihood for every data point to be associated with a specific cluster. This is achieved by soft clustering methods, mainly mixture models.

Mixture models are statistical models that combine several probability distributions, each representing a cluster in the dataset. Formally put, let $\mathbf{Y} \in \mathbb{R}^{\mathrm{M}}$ be a real random variable in dimension M, $\mathrm{M} \in \mathbb{N}^*$. The corresponding mixture model is comprised by $\mathrm{K} \in \mathbb{N}^*$ probability distributions $\{f_1, ..., f_K\}$ described by their respective parameters $\theta = \{\theta_1, ..., \theta_{\mathrm{K}}\}$.

The probability density function p of the mixture is defined by:

$$\forall \mathbf{y} \in \mathbb{R}^{\mathrm{M}}, \quad p(\mathbf{y}; \theta, \pi) = \sum_{k=1}^{\mathrm{K}} \pi_k f_k(\mathbf{y}; \theta_k) \tag{6.1}$$

where $\pi = \{\pi_1, ..., \pi_{\mathrm{K}}\}$ are the mixture proportions: $\forall k \in [1; \mathrm{K}]$, $0 < \pi_k < 1$ and $\sum_{k=1}^{\mathrm{K}} \pi_k = 1$.

Given a sample of independent realizations $\{\mathbf{y}_1, ..., \mathbf{y}_{\mathrm{N}}\}$ from the mixture model, the likelihood of the sample is defined by:

$$p(\mathbf{y_1}, ..., \mathbf{y}_{\mathrm{N}}; \theta, \pi) = \prod_{n=1}^{\mathrm{N}} \sum_{k=1}^{\mathrm{K}} \pi_k f_k(\mathbf{y_n}; \theta_k) \tag{6.2}$$

The estimation of a mixture model from a sample is generally done with an Expectation-Maximization algorithm (EM). The EM algorithm jointly estimates the mixture proportions π and parameters θ in an iterative two step process :

- **E-step.** Estimate for each data point the probability of belonging to each possible cluster given the current values of the parameters.
- **M-step.** Update the mixture parameters using the previous probabilistic assignments to increase the likelihood of the observed data set.

Before the first E-step, the algorithm is initialized, for example using the results of the k-means algorithm and after each M-step the log-likelihood of the model is calculated until the increase of the computed log-likelihood is below a given threshold. The algorithm is shown to increase the likelihood at each iteration and is hoped to reach a (local) maximum:

$$\left(\widehat{\theta}, \widehat{\pi}\right) = \arg\max_{(\theta,\pi)} \log p(\mathbf{y}_1, ..., \mathbf{y}_N; \theta, \pi) \tag{6.3}$$

6.2 Mixtures of Multiple Scaled t-distributions (MMST)

It is fair to say that Gaussian distributions are the most popular probability distributions employed in mixture models, yet, like all other elliptical distributions (multivariate Student, Laplace, etc.) They are limited by the type of elliptical shapes they allow, making them unsuitable to model data presenting elongated and strongly non-elliptical subgroups which is often the case for physiological parameters.

As an alternative, Forbes & Wraith [14] introduced a richer family of multiple scaled t-distributions (MST) which is a generalization of the multivariate t-distributions. These heavy-tailed distributions are endowed with variable marginal amounts of tail-weight and their mixtures have been shown to provide an efficient alternative to Gaussian mixtures. In particular, their ability to model over-dispersed values is illustrated in Figure 6.1.

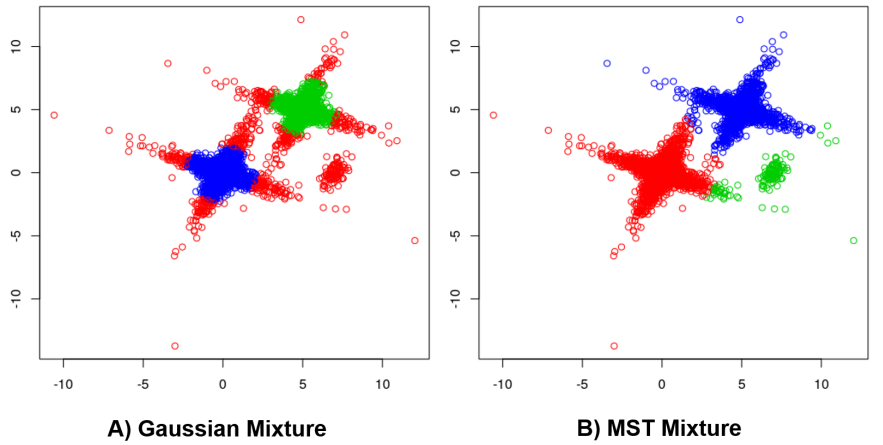


FIGURE 6.1: Gaussian versus MST mixture models on synthetic bivariate data.

Parting from a M-dimensional Gaussian distribution denoted $\mathcal{N}_M(\mathbf{y};\theta)$ of parameters $\theta = (\mu, \Sigma)$, where μ correspond to the mean and Σ to the covariance; a t-distribution can be represented as an infinite mixture of scaled Gaussians of the form:

$$p(\mathbf{y};\mu,\mathbf{\Sigma},\theta) \;=\; \int_0^\infty \mathcal{N}_M(\mathbf{y};\mu,\mathbf{\Sigma}/w)\; f_W(w;\theta)\; \mathrm{d}w \tag{6.4}$$

Most commonly, f_W takes the form of a Gamma distribution $\mathcal{G}(\nu/2,\nu/2)$ where ν denotes the degrees of freedom. The resulting density distribution denoted by $t_M((y;\mu,\mathbf{\Sigma},\nu)$ is expressed:

$$\begin{aligned} t_M(\mathbf{y};\mu,\mathbf{\Sigma},\nu) &= \int_0^\infty \mathcal{N}_M(\mathbf{y};\mu,\mathbf{\Sigma}/w)\; \mathcal{G}(w;\nu/2,\nu/2)\; \mathrm{d}w \\ &= \frac{\Gamma((\nu+M)/2)}{|\mathbf{\Sigma}|^{1/2}\,\Gamma(\nu/2)\,(\pi\nu)^{M/2}}\,[1+\delta(\mathbf{y},\mu,\mathbf{\Sigma})/\nu]^{-(\nu+M)/2} \end{aligned} \tag{6.5}$$

where $\delta(\mathbf{y},\mu,\mathbf{\Sigma}) = (\mathbf{y}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{y}-\mu)$ is the Mahalanobis distance between $\mathbf{y}$ and μ (^{T} means transpose) and Γ is the Gamma function: $\Gamma(x) = \int_0^{+\infty} t^{x-1}\exp(-t)dt\ ,\ x \in \mathbb{R}_+^*$. Note that μ is the mean when $\nu > 1$ but $\mathbf{\Sigma}$ is not strictly speaking the covariance matrix of the t-distribution which is $\nu/(\nu-2)\mathbf{\Sigma}$ when $\nu > 2$.

The weight variable W is commonly univariate which results in tails with the same heaviness in all dimensions. Forbes and Wraight propose to parametrize the scale matrix parameter into $\mathbf{\Sigma} = \mathbf{DAD}^T$, where $\mathbf{D}$ is the matrix of eigenvectors of $\mathbf{\Sigma}$ and $\mathbf{A}$ is a diagonal matrix with the corresponding eigenvalues of $\mathbf{\Sigma}$. The matrix $\mathbf{D}$ determines the orientation of the Gaussian and $\mathbf{A}$ its shape. The scaled Gaussian part is then set to $\mathcal{N}_M(\mathbf{y};\mu,\mathbf{D\Delta_w AD}^T)$, where $\mathbf{\Delta_w} = \mathrm{diag}(w_1^{-1},\ldots,w_M^{-1})$ is the $M\times M$ diagonal matrix whose diagonal components are the inverse weights $\{w_1^{-1},\ldots,w_M^{-1}\}$. The full expression then becomes:

$$p(\mathbf{y};\mu,\mathbf{D},\mathbf{A},\theta) \;=\; \int_0^\infty ... \int_0^\infty \mathcal{N}_M(\mathbf{y};\mu,\mathbf{D\Delta_w AD}^T)\; f_{\mathbf{w}}(w_1...w_M;\theta)\; \mathrm{d}w_1...dw_M \tag{6.6}$$

where $f_{\mathbf{w}}$ is now a M-variate density function. When the weights are all one, a standard multivariate Gaussian case is recovered. If the weights are independent so that, *i.e.* with $\theta = \{\theta_1,...,\theta_M\}$, $f_{\mathbf{w}}(w_1...w_M;\theta) = f_{W_1}(w_1;\theta_1)...f_{W_M}(w_M;\theta_M)$, we can then use the expression below:

$$\mathcal{N}_M(\mathbf{y};\mu,\mathbf{D\Delta_w AD}^T) \;=\; \prod_{m=1}^{M} \mathcal{N}_1([\mathbf{D}^T(\mathbf{y}-\mu)]_m;0,A_m w_m^{-1}) \tag{6.7}$$

where $[\mathbf{D}^T(\mathbf{y}-\mu)]_m$ denotes the mth component of vector $\mathbf{D}^T(\mathbf{y}-\mu)$ and A_m the mth diagonal element of the diagonal matrix $\mathbf{A}$ (or equivalently the mth eigenvalue of $\mathbf{\Sigma}$). Using

(6.7), it follows that

$$p(\mathbf{y};\mu,\mathbf{D},\mathbf{A},\theta) \;=\; \prod_{m=1}^{M}\int_0^{\infty}\mathcal{N}_1([\mathbf{D}^T(\mathbf{y}-\mu)]_m;0,A_m w_m^{-1})\; f_{W_m}(w_m;\theta_m)\;\mathrm{d}w_m\;. \quad (6.8)$$

The terms in the product reduce then to standard univariate scale mixtures and in particular, when setting $f_{W_m}(w_m;\theta_m)$ to a Gamma distribution $\mathcal{G}(w_m;\nu_m/2,\nu_m/2)$, it follows a generalization of the multivariate t-distribution. We can use (6.8) to express easily the density denoted by $p_{\text{MST}}(\mathbf{y};\mu,\mathbf{\Sigma},\nu)$ with $\nu=\{\nu_1,...,\nu_M\}$:

$$p_{\text{MST}}(\mathbf{y};\mu,\mathbf{D},\mathbf{A},\nu)=\prod_{m=1}^{M}\frac{\Gamma((\nu_m+1)/2)}{\Gamma(\nu_m/2)(A_m\nu_m\pi)^{1/2}}\left(1+\frac{[\mathbf{D}^T(\mathbf{y}-\mu)]_m^2}{A_m\nu_m}\right)^{-(\nu_m+1)/2} \quad (6.9)$$

We can then rewrite equation (6.1) as follows to have the density of a mixture of MST (MMST model) $\forall\mathbf{y}\in\mathbb{R}^{\mathrm{M}}$:

$$\begin{aligned} p_{\text{MMST}}(\mathbf{y};\psi) &= \sum_{k=1}^{\mathrm{K}}\pi_k p_{\text{MST}}(\mathbf{y};\mu_k,\mathbf{D}_k,\mathbf{A}_k,\nu_k) \quad (6.10)\\ &= \sum_{k=1}^{\mathrm{K}}\pi_k\prod_{m=1}^{\mathrm{M}}\frac{\Gamma\left(\frac{\nu_{k,m}+1}{2}\right)}{\Gamma\left(\frac{\nu_{k,m}}{2}\right)\left([\mathbf{A}_k]_{m,m}\nu_{k,m}\pi\right)^{\frac{1}{2}}}\left(1+\frac{[\mathbf{D}_k(\mathbf{y}-\mu_k)]_m^2}{[\mathbf{A}_k]_{m,m}\nu_{k,m}}\right)^{-\frac{\nu_{k,m}+1}{2}} \end{aligned}$$

with: $\psi=\{\psi_1,...,\psi_{\mathrm{K}}\}$ where $\psi_k=(\pi_k,\mu_k,\mathbf{D}_k,\mathbf{A}_k,\nu_k)$ for $k\in[1;\mathrm{K}]$.

6.3 Pipeline: applying MMST for anomaly detection

With all the mathematical tools available to fit a MMST onto a multivariate dataset, it is also possible to use the generative properties of mixture models to simulate data and to appraise the probability of new data to belong to an existing model.

Arnaud & colleagues [4] produced an original pipeline employing the MMST R library "ompmmsd_0.7" (available at: `https://www-ljk.imag.fr/membres/Stephane.Despreaux/MMSD/Download/0.7/`) to characterize tumors in a rodent model using MRI data. The procedure is data-driven and requires the availability of two data sets: one from healthy subjects and another one from subjects where we expect to find abnormalities, in this case, traces of a pathology. First, a mixture model is fitted to the healthy subjects voxels assuming that healthy subjects present a set of distinct healthy tissues characterized by different quantitative characteristics. This reference model is later used to detect voxels which exhibit abnormal MR features with respect to the reference model, in the healthy and pathological subjects. The newfound abnormal voxels are in turn used to build another mixture model that describes the classes of atypical characteristics found in the data. The proportions of

theses classes in each subject are used as a signature of the pathology and finally, these signatures are refined in an spatial post-processing step.

To ensure a fully automatic procedure Arnaud & colleagues opted for a data-driven way to choose the appropriate number of clusters K for the healthy and pathological mixture models. Considering a too small K may not allow a good fit of the data while a too large K may lead to overfitting issues. The choice of K was treated as a model selection problem by minimizing a penalized log-likelihood criterion to regulate the complexity of the model. The data-driven slope-heuristic method, developed by Baudry, Mugis & Michel [123] was employed to tune the penalization term.

In practice, for a collection of mixture models identified by their cluster numbers K between 1 and $K_{\max}$, a subset of $f_K = \{f_1, \ldots, f_{K_{\max}-2}\}$ models is extracted. Each model f_K is described by its log-likelihood γ_K and its number of free parameters P_K. On this subset, a robust linear regression on the negative log-likelihoods $\{\gamma_1, \ldots, \gamma_{K_{\max}-2}\}$ is computed with $\{P_1, \ldots, P_{K_{\max}-2}\}$ as regressors and the regression slope C_K is estimated.

The optimal number of clusters K^* corresponds to the model f_K^* which minimizes the associated penalized log-likelihood:

$$f_K^* = argmin_{(f_k \in f_K)} \{2C_K P_K - \gamma_K\}$$

This corresponds to the first model from which the log-likelihood is in a linear regime with respect to the number of free parameters. It is also the last model before overfitting.

Equipped with this automatic model selection functionality, the full pipeline yielded very satisfactory results and even allowed brain tumor classification [4]. Before applying it to the detection of anomalies in the quantitative MR brain data of 'de novo' Parkinsonian patients, some modifications were needed. One of the main differences between the two problems is that while tumors express strong differences with respect to healthy tissues, Parkinsonian anomalies are not clearly defined; what is more, they are expected to be subtle, especially in 'de novo' cases. For instance, it is well documented that gliome vascularity is altered and the diffusion of water within the tumor is different [124], on the other side, studies regarding brain perfusion in PD patients are limited and research in diffusion data report divided results as already exposed in Section 3. This comes into play when attempting to separate abnormal from normal voxels. In the work of [4], a two component mixture model was fitted to the log-likelihood scores of all voxels with respect to the reference model. This resulted in two distinct components with small tail, which is not the case in the PD application. For this reason, we decided to rather define a threshold of abnormality, as presented later in this section.

Another important difference, is the access to a ground truth segmentation of abnormalities. In effect, in the gliome application Arnaud & colleagues could calculate different performance scores (i.e. DICE, ARI) to compare their results to the manual segmentation and

characterization furnished by physicians whilst the application of the pipeline is completely exploratory. Our goal is not to improve or automatize an existing method for anomaly detection but instead to model a combined set of measures from multiple MRI modalities in the hope to provide a complete picture of the initial functional changes caused by PD.

Additionally, some optimizations were added to enable the application of the pipeline to bigger datasets. As mentioned before, the models are fit onto the parameter values at voxel-level and not surprisingly humans brains are bigger than rodent brains and their MR images often have a better resolution. As a matter of fact a human brain measures 1 200 cm^3 in average whilst a rat brain measures merely 2.59 cm^3. In a diffusion parameter map, this translates to 175 000 voxels for a human brain and approximately 8 000 for a rat brain.

Hereafter the modified pipeline of five steps is presented in more detail.

6.3.1 Reference Model

Each subject is associated to M co-localized MR parameter maps that provide for each voxel v a M-dimensional vector of parameters denoted $\mathbf{y}_v$. The set of healthy voxels is denoted $\mathcal{V}_H$ and that of patients voxels $\mathcal{V}_H$, while their corresponding datasets are defined as $Y_H = \{y_v, v \in \mathcal{V}_H\}$ and $Y_P = \{y_v, v \in \mathcal{V}_P\}$ for the healthy and pathological individuals correspondingly.

In this first step of the procedure an MST mixture is fitted to Y_H building a reference model of density f_H that describes healthy tissues.

$$f_H(y|\pi, \theta) = \sum_{k=1}^{K_H} \pi_k \mathcal{MST}(y; \psi_k), \tag{6.11}$$

where K_H is the number of mixture components, and each component is characterized by a proportion π_k and an MSD parameter ψ_k. The EM algorithm is used to obtain the best fit of the MST mixture on Y_H. Iteratively, the MST mixture model parameters $\psi = \{\mu, \mathbf{D}, \mathbf{A}, \nu\}$ and the π proportions of every cluster are optimized to maximize the global log-likelihood of the model.

The K_H clusters in f_H account for the potential heterogeneity in the parameter values belonging to different brain tissues or structures within the human brain. f_H is referred to as the reference model and its generative properties permit the detection of abnormalities.

6.3.2 Outlier Detection

Now the goal is to identify any voxels manifesting abnormal MR parameter values. For this anomaly detection task the log-likelihood score is considered as a measure of proximity for every v (associated to value y_v) to the reference healthy model (represented by f_H).

The rationale is that voxels from healthy subjects are more likely to have a high log-score while pathological voxels may not be well explained by the reference model and consequently have a lower log-score.

This is not restricted to the pathological dataset. Surely, it is interesting to recognize any voxels that are not well statistically explained by f_H within the healthy dataset, in order to differentiate healthy heterogeneity from abnormalities related to PD.

Therefore, the log-likelihood score with respect to f_H of all voxels included in the healthy and pathological datasets $\{\log f_H(y_v), v \in V_H \cup V_P\}$ is computed and the ensemble of values is used to establish an abnormality threshold noted τ_α.

Every possible threshold τ_α is associated to a certain false positive error rate (FPR) α. In other words, the probability a log-score is smaller than τ_α, although the log-score is that of an healthy voxel, is α.

Formally τ_α is the value such that:

$$P(\log(f_H(Y)) < \tau_\alpha) = \alpha,$$

when Y is a random variable following the f_h reference model distribution. In practice, while f_H is known explicitly, the probability distribution of $\log(f_H(Y))$ is not. However, it is easy to simulate this distribution so that τ_α can be computed using empirical quantiles.

α is fixed to an acceptable value (i.e. 5%) and then all the voxels whose log-score is below τ_α are labeled as abnormal and the corresponding measures provide a set of parameters that are referred to as the abnormal data set:

$$Y_A = \{y_v, v \in V_H \cup V_P,\ s.t.\ \log f_H(y_v) < \tau_\alpha\}.$$

We note that several thresholds of abnormality can be established to study the proportions and localization of the resulting nested anomaly segmentations.

6.3.3 Atypical Model

The following step is dedicated to characterize the newly found abnormal voxels. A new model denoted by f_A is then fitted to Y_A in the same manner as the f_H model was fitted to the Y_H dataset.

$$f_A(y|\eta, \phi) = \sum_{k=1}^{K_A} \eta_k \mathcal{MST}(y; \phi_k). \tag{6.12}$$

This abnormality model is used to account for the fact that voxels detected as abnormal may belong to different abnormality classes, with different physiological characteristics. Typically, the above formula indicates that among the Y_A set there are K_A different clusters.

6.3.4 Signature Extraction

The different abnormality classes can in turn be used to build a signature ρ^S of each subject S by determining the proportion of voxels that are assigned to each of the K_A group. The probability that a voxel $v \in \mathcal{V}_S$ belongs to cluster k among K_A clusters is expressed as :

$$p_k^v = \frac{\eta_k \mathcal{MST}(\mathbf{y}_v; \phi_k)}{\sum_{l=1}^{K_A} \eta_l \mathcal{MST}(\mathbf{y}_v; \phi_l)}$$

and thus, the mean probability over all voxels in $\mathcal{V}_S$ is

$$p_k^S = \frac{\sum_{v \in \mathcal{V}_S} p_k^v}{n_S} \tag{6.13}$$

The resulting signature is then :

$$\rho^S = (p_1^S, \ldots, p_{K_A}^S) \tag{6.14}$$

This signature aims to characterize and classify subjects. Notably, discriminant analysis models can be used to learn the differences between healthy and pathological signatures to later classify unlabeled subjects. In addition, the presence of distinct signature patterns among the patients may indicate different phenotypes of the disease.

6.3.5 Spatial post-processing

As explained in [4], we can make use of this classification information and of additional spatial information to refine abnormality detection. For instance, clusters of less than 4 atypical voxels are discarded. Figure 6.2 shows the effect of such a post-processing.

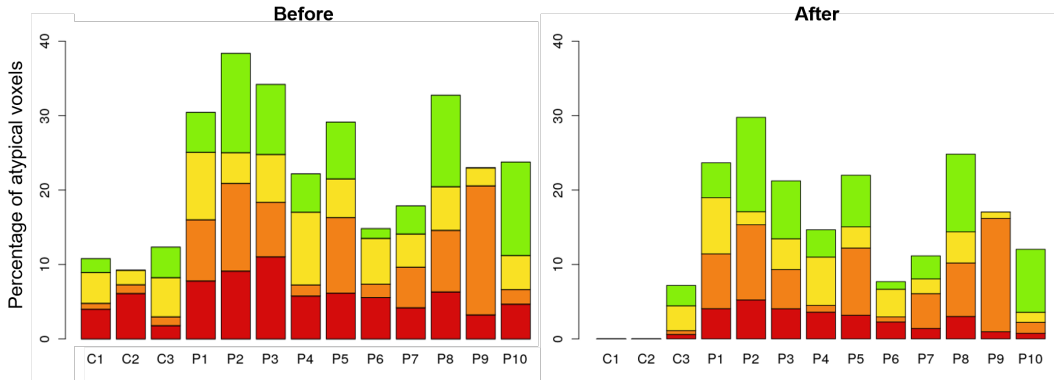


FIGURE 6.2: Effect of post-treatment on the subject's signatures

6.4 Application 1: AGIR-Park data study

6.4.1 Data presentation and pre-processing

The first application of the previously presented pipeline to characterize PD abnormalities was carried out on a small population consisting on three healthy subjects (C1-C3) (Age: $\{28, 40, 50\}$; Sex: 2M,1F) and ten *de novo* PD patients (P1-P10) (Age $\in [36, 66]$; H&Y score ≤ 2; Sex: 7M, 3F).

For each individual three MR sequences were acquired on a Philips 3T scanner at the IR-MaGe platform in Grenoble.

- **Anatomical image:** T1 3D MP-RAGE with a spatial resolution of 1 x 1 x 1 mm3, 180 sagittal slices, an acquisition matrix of 256×240 and a TR/TE/TI = 4.8/2.3/616 ms and a flip angle of $9°$.
- **Diffusion scan:** Two DTI iso SENSE scans of 26 slices centered on the visual tract of dimensions 112×112 were acquired with 64 gradient directions. One of the volumes was acquired on the anterior/posterior orientation and the other one on the posterior/anterior orientation. Both with a TE/TR = 90/3300 ms.
- **Blood perfusion scan:** pCASL scan covering the whole brain with a labelling at 1800 ms and a post-labelling delay of 1643 ms and an EPI single-shot ($3.5 \times 3.5 \times 3.5$ mm3, 20 slices, TE/TR = 4230/12ms).

We extracted FA and MD parameter maps from the DTI images following a classic pipeline of MRtrix3 which includes denoising, Gibbs artifacts correction, merging of the two DTI orientations, eddy current correction, tensor calculation and metric inference. The CBF maps were derived from the pCASL images with an in-house method following the indications in [125].

The computed maps were coregistered and resliced (1mm isotropic) with respect to the T1-weighted image using SPM12. Having the images in the same space permits us to have for each voxel v, three corresponding feature values $y_v = (\text{FA}_v, \text{MD}_v, \text{CBF}_v)$ to which we could fit the reference and atypical MMST models.

We opted to fit the model on the voxels belonging to the subcortical structures in the MNI PD25 atlas [120] (introduced in Chapter 5). Indeed, as the diffusion scan only covered a section of the brain, we decided to concentrate in the subcortical structures that were, for their majority, entirely comprised in the 26 axial slices available.

6.4.2 Results

Several MMST models of $K \in \{1; 12\}$ were fitted to the control subjects data, the slope heuristic retained a reference model (f_H) of $K_H = 6$ clusters that obtained the best compromise between log-likelihood and complexity of the model. The f_H clusters seemingly respect spatial cues such as symmetry and structural composition. A projection of these reference classes onto the control subjects subcortical structures is showcased in Figure 6.3-D).

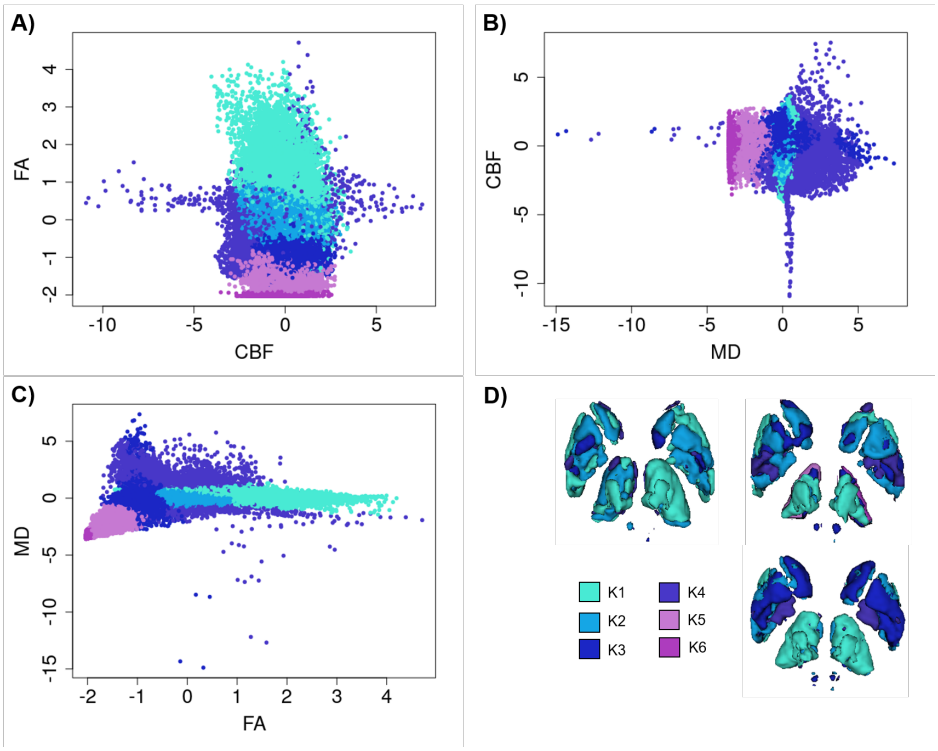


FIGURE 6.3: Two-by-two scatter plots of the multivariate physiological values to which the reference model was fitted colored by cluster: A)FA as a function of CBF. B)CBF as a function of MD. C) MD as a function of FA. D) Spatial localization of the classes in the subject's subcortical structures.

In Figure 6.3 it is visually apparent that the distributions of the physiological parameters here studied (FA,MD and CBF) are not Gaussian (i.e. elliptical), thus renforcing the utility of MMST models. To simplify their comparison, we characterize every f_H cluster through their center μ_K, which is the vector that indicates the mean value of the cluster in every parameter dimension, and their size in the mixture π_K given as a percentage, as reported in Table 6.1.

Outliers were detected based on their low likelihood of belonging to this reference model. The accepted false positive error rate (FPR) α was empirically fixed at 10%, which corresponded to a t_α pathological threshold of -5.558722, where $\alpha = 100\%$ equates to a negative log-likelihood threshold of -1.276832.

A total of 51533 voxels (21.6%) with a negative log-likelihood score above t_α were deemed as atypical and thus the MMST atypical model (f_A) was fitted to their properties (Y_A). The

	K_H1	K_H2	K_H3	K_H4	K_H5	K_H6
CBF($ml/100g/min$)	38.63	39.37	41.98	29.89	37.81	36.91
FA (scalar)	0.386	0.267	0.158	0.251	0.058	0.003
MD ($\mu m^2/s$)	696.2	677.3	660.3	805.8	261.7	55.8
Size (%)	32.21	31.81	18.38	9.13	5.79	2.70

TABLE 6.1: Characteristics of the reference model clusters: mean value for every parameter and overall size in the model.

resulting model counted with $K_A = 4$ clusters, according to the slope heuristic. Only 32719 voxels survived post-treatment, that is 13.7% of all voxels.

Each cluster in f_A corresponds to a MST distribution. Table 6.2 provides a summary of these distributions by their centers and sizes.

	K_A1	K_A2	K_A3	K_A4
CBF ($ml/100g/min$)	30.92	74.79	33.51	45.69
FA (scalar)	0.427	0.244	0.145	0.110
MD ($\mu m^2/s$)	729.4	673.6	1032.3	218.5
Size (%)	27.95	27.61	23.42	21.02

TABLE 6.2: Characteristics of the atypical model clusters: mean value for every parameter and overall size in the model.

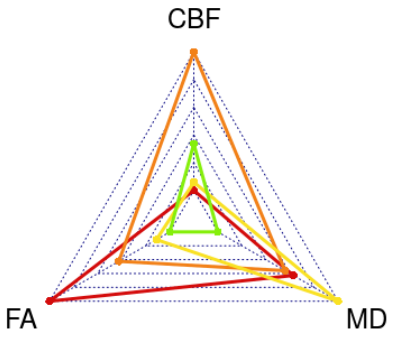


FIGURE 6.4: Clusters comparison by parameter means.

The atypical clusters are present in a different proportion for every subject and by observing their signatures it is possible to verify that there is an evident difference between controls and patients but also to identify subgroups of signatures within the pathological group. The signatures of the study after post-treatment are presented in Figure 6.5 (right).

Hierarchical clustering on these signatures identifies two subtypes of patients, the first composed by P4, P6, P7, P10 and the second one by P1, P2, P3, P5, P8 and P9.

The localization of the atypical voxels is also of great importance. Just by looking at the projection of the atypical voxels with their classes onto the subcortical structures of our subjects (like in Figure 6.5) we are able to understand that the anomalies encompassed by the K_A4 cluster (green) are probably due to partial volume. As a matter of fact, these voxels are generally in the rear part of the putamen and the caudate nucleus, near the point were last diffusion slices were found (see 6.4.1).

Furthermore, we can notice that K_A1 clusters (red) are predominantly located on the red nucleus, the substantia nigra and the subthalamic nucleus, on average between the ten patients they make up for 21.7, 12.9 and 25.9 % of these structures respectively. In the reference model, the RN, SN and STN of the control subjects is characterised by the K_H1 cluster.

Comparing the two clusters parameter averages there seems to be a decrease in CBF and an increase in both FA and MD.

The K_A2 cluster has the biggest CBF average (74.79 $ml/100g/min$) out of both healthy and atypical clusters. It is mainly present on the biggest subcortical structures, the thalamus (8.7%) the putamen (6.6%), the external segment of the globus pallidus (4.9%) and the caudate nucleus (4.1%).

K_A3 abnormalities have the greatest MD values on average (1023.3 $\mu m^2/s$). They were also found in the caudate nucleus (6.4%) and in smaller quantities in the internal segment of the globus pallidus (5.0%), the putamen (3.0%) and the thalamus (2.2%). Half of the patients present important K_A3 clusters in the colliculus (superior and/or inferior) going up to 76.2% of the structures for P8.

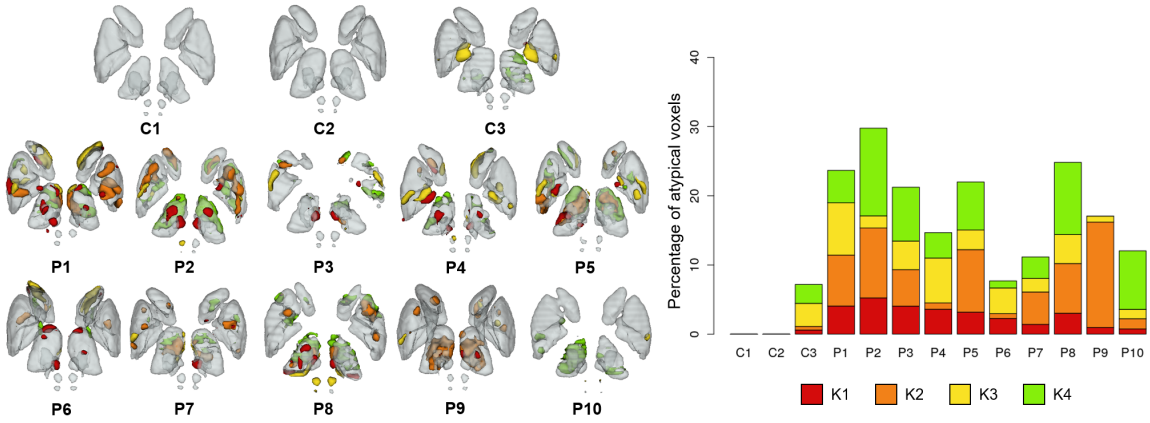


FIGURE 6.5: Spatial localization of the identified atypical clusters for all study subjects (left) and their atypical signatures built from the proportion of each cluster on the patients.

The persistent abnormalities in the third control subject are majoritarily from the K_A3 cluster and can be found in the globus pallidus, where they represent 55.8% of the internal segment on and 15.2% of the external segment.

We calculated the percentage of abnormal voxels belonging to the first three atypical clusters ($K_A = 1, 2, 3$) in every structure (Table 6.3). In smaller regions (RN, SN, STN, SC, IC), atypical clusters often take up a bigger part of the structure, while for bigger structures (putamen, thalamus, caudate nucleus), they constitute a smaller percentage even though the size of the clusters are considerable.

6.4.3 Discussion

The presented method for the detection of abnormalities in *de novo* PD patients using MST mixture models produced encouraging results on a small cohort of three controls and ten patients. The MMST models were able to capture the complexity of FA, MD and CBF parameter distributions that are evidently non-elliptical. Despite the fact that the reference

%	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
RN	2.5	57.8	36.4	25.0	13.2	25.7	30.6	25.7	0	0
SN	46.4	29.4	5.2	16.0	3.0	0	0	18.4	11.3	0
STN	43.6	77.3	0	59.7	16.1	15.1	0	17.0	30.8	0
Cau	32.1	12.9	16.8	18.4	7.1	16.6	3.9	3.6	5.4	0
Put	15.9	26.8	22.4	3.3	12.0	2.0	5.3	11.7	3.1	2.5
GPe	21.6	21.1	7.1	10.5	14.2	2.5	1.9	8.3	1.0	1.9
GPi	31.2	10.2	9.5	50.2	29.1	4.6	1.7	1.2	0.2	5.4
Th	13.5	3.7	9.3	4.7	19.8	4.7	13.1	18.8	39.9	5.8
SC	0	30.5	0	33.3	23.4	0	0	76.2	0	52.4
IC	0	31.9	0	0	0	0	0	33.3	0	0

TABLE 6.3: Percentages of abnormal voxels after post-processing in each patient subcortical structure.

model was estimated without any spatial information, the retained model is coherent from an anatomical standpoint and respects symmetry.

The estimation of the atypical model was proven useful in allowing us to eliminate abnormalities linked to partial volume interference, what is more, we were able to study the location of the abnormal clusters and found that the K_A1 clusters, mainly located on the substantia nigra, the red nucleus and the subthalamic nucleus, are characterized by low CBF and high FA values, which is in accordance with the results of Fernandez-Seara & al. [15] (Table 3.5) and Couisineau & al. [60] (Table 3.4).

Also, the K_A3 abnormalities in the striatum coincide with the higher MD values reported by Peran & al. [16] (Table 3.6). Nevertheless, these conclusions are not unanimous across the literature and our cohort is quite small.

Clearly, more subjects are needed to bring out robust biomarkers of PD. However, these preliminary results show that the application of mixture models of relevant distributions is informative and promising to correctly discriminate the pathology.

6.5 Application 2: PPMI data study

6.5.1 Data presentation and pre-processing

For the second application we employed diffusion data from the PPMI [6] database, like in the previous chapter. From PPMI, we pooled the DTI scans of 56 healthy controls (Age: 61.1 $\pm$ 9.8, Sex: 34M, 23F) and 129 'de novo' PD patients (Age: 61.74 $\pm$ 8.96, Sex: 80M, 49F), all measured with a 3T Siemens Trio Tim MRI scanner and configured with the same acquisition parameters (90° flip angle, 64 gradient directions, EP Pulse sequence, TE=88ms) to limit bias in our study.

As this method is designed to model quantitative data, we computed the FA (fractional anisotrophy) and MD (mean diffusivity) parameter maps for all of the subjects using MRtrix3.

This being a bigger dataset with approximately 7.2 million voxels for the control subjects and 35.9 million voxels for the patients, a considerable amount of RAM memory was required. The complete pipeline was implemented using R 3.6.3 on a Intel Xeon Gold6130 machine with 32 cores of 2.1 GHz of processor speed and 384 Go of system memory.

6.5.2 Results

Twelve MMST models of $K \in \{1; 12\}$ were fitted to the diffusion features of all voxels belonging to the control subjects ($Y_H = y_v; v \in V_H\}$ where $y_v = \{\text{FA}_v, \text{MD}_v\}$). The model chosen with the slope heuristic to be the reference model (f_H) counted with $K_H = 6$ clusters. In Figure 6.6 the six clusters are displayed directly on the distribution of FA and MD values of the control subjects.

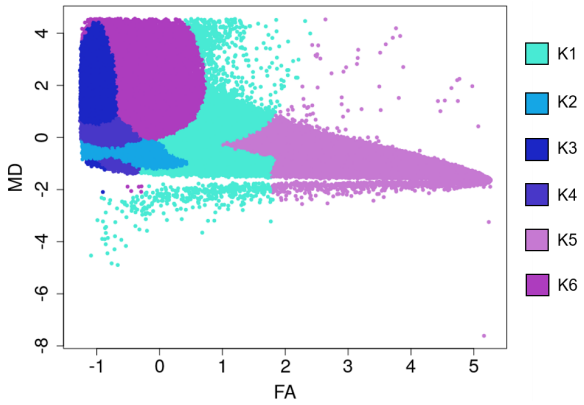


FIGURE 6.6: Scatter plot of the FA and MD features belonging to the healthy subjects. The colors represent the f_H clusters.

Each cluster is summarized on Table 6.4 by their cluster mean values and their proportion on the mixture.

	K_H1	K_H2	K_H3	K_H4	K_H5	K_H6
FA (scalar)	0.376	0.187	0.062	0.097	0.540	0.151
MD ($\mu m^2/s$)	730.0	760.7	2281.8	1172.7	754.6	2237.5
Size (%)	23.49	23.38	17.78	15.53	13.06	6.76

TABLE 6.4: Characteristics of the reference model clusters: mean value for every parameter and overall size in the model.

Outlier detection was performed on the ensemble of voxels from controls and patients. The atypical model (f_A) was constructed from the voxels with a log-likelihood FPR $\alpha \geq 10\%$ of belonging to the reference model ($Y_A = \{y_v, v \in V_H \cup V_P,\ s.t.\ \log f_H(y_v) < \tau_{10\%}\}$). According to the slope heuristic, the best f_A is composed by $K_A = 4$ clusters that can be described

by their center (μ_K), that is the intersection between their mean FA and MD values, and their proportion within the model (η_K). This characteristics are available on Table 6.5.

	K_A1	K_A2	K_A3	K_A4
FA (scalar)	0.274	0.153	0.771	0.582
MD ($\mu m^2/s$	1460.3	3088.6	834.2	362.5
Size (%)	28.08	27.32	25.14	19.46

TABLE 6.5: Characteristics of the atypical model clusters: mean value of FA and MD as well as size in the model.

From the 1 810 395 voxels detected as atypical, that is 5.5% of all voxels, only 620 950 (1.9%) remained after post-treatment. The resulting signatures of the study control and patient subjects are illustrated on Figure 6.7.

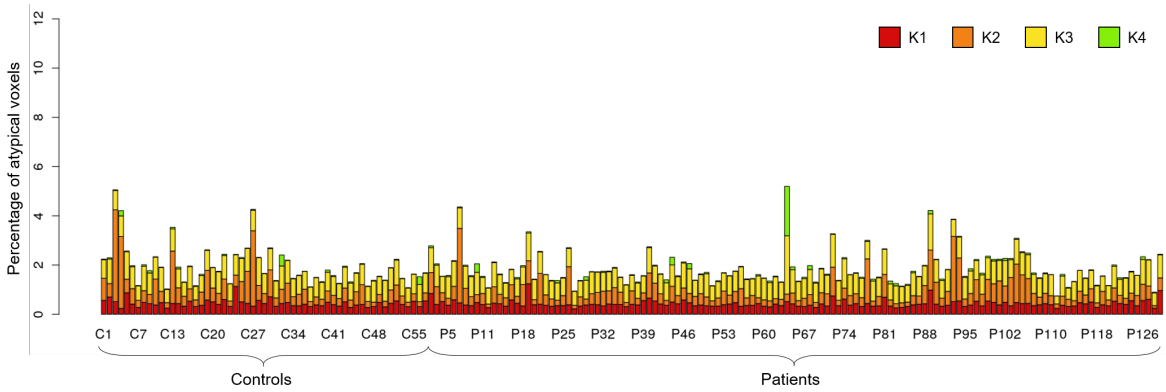


FIGURE 6.7: Subjects' signatures describing the proportion of each atypical subject in their brain.

There is no apparent difference between neither on the amount nor the class of abnormalities found in PD patients compared to healthy controls. When studying the localization of the abnormalities, we discovered that most abnormalities were found on the corpus callosum and around the ventricules in general. An additional experiment concentrating on the subcortical structures exclusively did not produce conclusive results either.

6.5.3 Discussion

The application of MMST models for the detection of diffusion anomalies on a considerably large dataset produced inconclusive results.

Once again the non-gaussianity of the data indicated that MST distributions are better suited to describe physiological parameters, however, diffusion anomalies were observed in equal quantity for control subjects and PD patients.

Most of the abnormalities were located on the corpus callosum. This is the largest white matter structure of the brain, consisting on a flat bundle of commisural fibers that connect

the left and right hemispheres. The implication of this structure in Parkinson's disease and has been related to cognitive decline [126], yet, a recent case study reporting a PD patient without corpus callosum (complete agenesis) [127] suggests that the bilateral degenerative changes in PD may occur independent of the state of the corpus callosum.

It is quite probable that diffusion data alone is not enough to discriminate PD patients from healthy controls, at least on the early stages of the disease, notwithstanding, to continue the study it would be interesting to employ our method on a carefully chosen subset of white matter structures related to specific symptoms assessed by a clinical score from which we can position our results.

We note that the estimation of the models on a dataset of this size is both time consuming and computationally expensive. Computational developments are on their way to optimize the model estimation and adapt the pipeline for large datasets.

6.6 Application 3: SignaPark data study

The INNOBIOPARK study (PI: E. Moro) is a group project of several laboratories and hospital services within the Université Grenoble Alpes, designed to identify potential new and innovative biomarkers in de novo Parkinson's disease patients.

The study was designed to include 60 healthy controls and 60 *de novo* PD patients matched in age and gender. Five different modalities are assessed at baseline, 1 year and 2 years follow-up:

1. Detailed clinical evaluation using the current available validated clinical scales (MoCA, UPDRS, SCOPA, etc.)
2. Dopamine transporter (DAT) SPECT imaging.
3. Comprehensive brain evaluation using MRI.
4. Cortical brain mapping and transcranial magnetic stimulation assessment using a robotized approach.
5. Emotional responses study using a novel paradigm.

The third modality, called SignaPark, is embedded in this thesis project. Healthy volunteers and *de novo* PD patients have been included in the protocol since July 2019. Each individual was imaged during one hour in a Philips Achieva dStream 3T imager at the IRMaGe platform (CHU Grenoble Alpes). Anatomical, perfusion, diffusion and relaxometry sequences parameterized at IRMaGe were performed.

6.6.1 Data presentation and pre-processing

In this preliminary study we compare the MRI data of 2 healthy volunteers (Ages: 56,58, Sex: 2F) and 12 early PD patients (Age: 62.9 $\pm$ 7.9, Sex: 6M, 6F, H&Y=2, MoCA>26). The protocol's scans relevant to this study are the following:

- **Anatomical image:** 3D MPRAGE T1-weighted scan of 220 sagittal images of 288×288 with a spatial resolution of 0.89×0.89mm, slice-thickness = 1mm, flip angle = 8° and TE/TR=3.75/8.1ms.
- **Diffusion scan:** Two DTI scans of 90 gradient directions each acquired in the posterior-anterior and anterior-posterior directions with a b=800. The volume consisted 60 images of 96×96 with a spatial resolution of 2.5×2.5mm, slice thickness= 2.5mm and TE/TR=85/3000ms.
- **Blood perfusion scan:**
  - 3D pCASL scan of 25 axial 64×64 images with a spatial resolution of 3.5×3.5mm, slice thickness=5mm, flip angle = 90°, TE/TR=12.96/4003ms. The labelling distance was set at 105mm, labelling was effectuated at 1800 ms and post-labelling at 1800ms.
  - DSC routine perfusion scan of 25 (112×112) axial images of 2×2mm spatial resolution, slice thickness = 4mm, TE/TR = 40/1559.4ms. A gadolinium contrast agent was administered with a dosage of 2mMoL/ml/kg at 66/s 30 seconds after the start of the sequence and is followed by a physiological serum administered at the same speed.
- **T1 relaxometry scan:** DCE scans at flip angle = 5, 15, 20 and 35. Each volume is composed of 110 (240×240) sagittal images with a spatial resolution of 1×1mm, slice thickness=4mm and TE/TR=3.85/8.1ms.
- **T2* relaxometry scan:** Four SWI scans at TE = 7.2, 13.4, 19.6 and 25.8ms. Each volume is made up of 130 (240×240) axial images with a spatial resolution of 0.95×0.95, slice thickness=1mm, flip angle=17°, TR = 31ms.

The MP3 (Medical software for Processing multi-Parametric images Pipelines) software [128], developed at the Grenoble Institute of Neurosciences, was employed to extract the quantitative parameter maps of all sequences, except for the pCASL sequence, for which CBF (cerebral blood flow) maps are calculated directly at acquisition. In order to differentiate these maps from the DSC CBF maps we refer to pCASL CBF maps as pCBF.

From the DTI sequences, FA and MD images were computed; from the DSC perfusion sequence, relative CBF and CBV (cerebral blood volume) maps were extracted as well as MTT (mean transit time) maps; from the four flip angle T1 scans a T1 relaxometry map was estimated and finally from the SWI scans at multiple echos, a T2 star relaxometry map was constructed.

The eight resulting parameter maps were coregistered and resliced (1mm isotropic) with respect to the T1-weighted anatomical scan using SPM12. Additionally, ventricules were supressed.

The rCBF, rCBV and MD features were eliminated based on their high correlation (>0.5) to other features. In consequence, every voxel v was associated to the feature vector $y_v = \{\text{FA}_v, \text{pCBF}_v, \text{MTT}_v, \text{T1}_v, \text{T2star}_v\}$.

In the first part of this application we chose to fit the MMST models to all brain voxels and on the second part we only studied the voxels belonging to the subcortical structures in the MNI PD25 atlas.

6.6.2 Results for the brain

The controls brain data was fitted with several MMST models from which the model presenting $K_H = 6$ clusters was retained as the reference model f_H by the slope heuristic. Pair plots of the physiological parameters describing the controls' voxels are featured on Figure 6.8 along with projections of the clusters on 2D slices of the two control subjects.

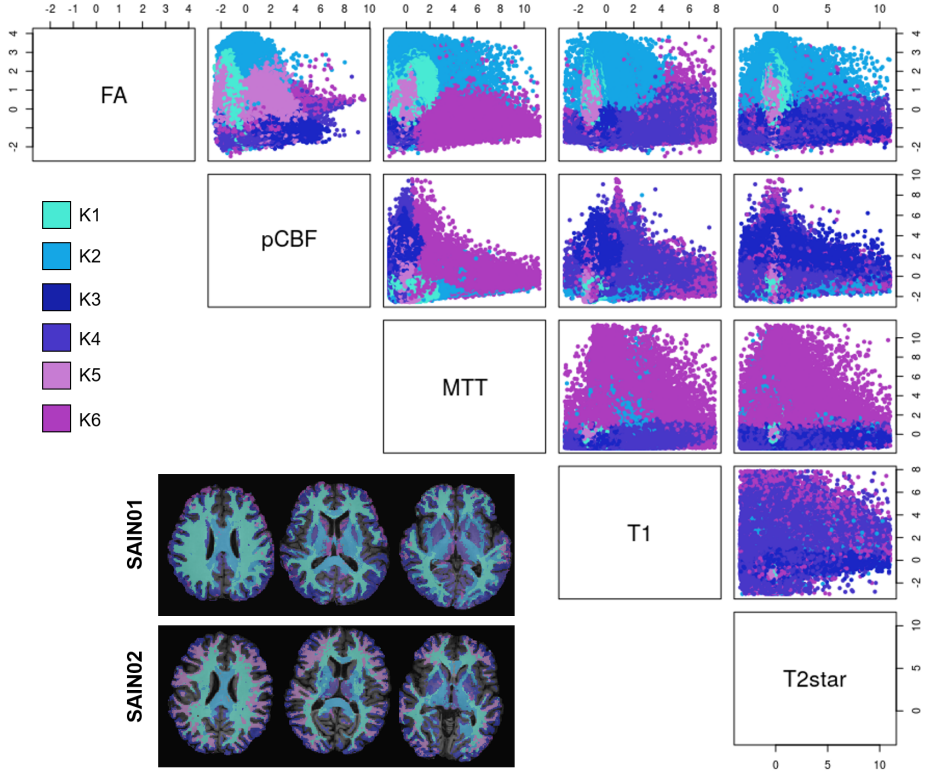


FIGURE 6.8: Scatter plots of the parameter features belonging to the healthy subjects. The colors represent the f_H clusters.

The repartition of the classes on the anatomical images is visually coherent. Most of the white matter is described by the K_H1 and K_H2 clusters, while the gray matter is mainly

described by the K_H3-5 clusters. For simplicity we summarize the clusters by their mean values (μ_K) for every parameter and their size in the mixture (π_K) on Table 6.6.

	K_H1	K_H2	K_H3	K_H4	K_H5	K_H6
FA (%)	0.391	0.430	0.155	0.138	0.308	0.145
pCBF ($ml/100g/min$)	14.49	20.93	35.30	25.47	32.58	24.32
MTT (s)	5.27	4.83	4.26	4.52	4.85	12.31
T1 (ms)	1105	1560	1313	1837	950	1427
T2star (ms)	47.33	44.48	53.65	56.16	48.52	55.79
Size (%)	25.2	19.52	19.11	15.74	12.59	7.84

TABLE 6.6: Characteristics of the reference brain model clusters: mean value of FA and MD as well as size in the model.

Outlying voxels from the controls and patients voxels were detected in function of their negative log-likelihood of belonging to the reference model. An allowed FPR was fixed at 5% and so any voxel with a negative log-likelihood above $t_{5\%}$ was declared as atypical.

An atypical MMST model of $K_A = 5$ clusters was retained, their mean values and sizes are reported on Table 6.7.

	K_A1	K_A2	K_A3	K_A4	K_A5
FA (%)	0.300	0.314	0.078	0.408	0.472
pCBF ($ml/100g/min$)	23.03	13.41	34.94	40.15	7.21
MTT (s)	8.34	8.67	7.15	10.52	13.98
T1 (ms)	1991	635	1053	1101	1192
T2star (ms)	63.78	48.84	69.69	48.66	49.08
Size (%)	22.59	20.57	20.45	18.82	17.57

TABLE 6.7: Characteristics of the atypical brain model clusters: mean value of FA and MD as well as size in the model.

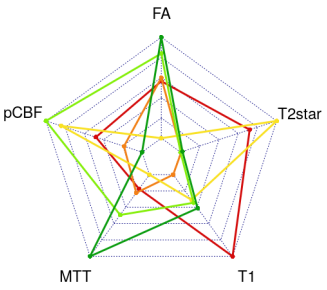


FIGURE 6.9: Clusters comparison by parameter means.

From an initial number of 1 987 053 (20.2%) voxels, 1 284 815 (13.1%) survived post-treatment. The atypical signatures of the control subjects and patients after post-treatment are displayed on Figure 6.10. For additional information, Table 6.8 presents the percentage of anomalies in several brain structures defined from the Neuromorphometrics atlas, presented on Chapter 5.

The signatures of patients P5, P7 and P8 show that a large percentage of these subjects' brains is considered anomalous, 28%, 26.5% and 35.6% respectively. The data on Table 6.8 indicates that a large portion of these patients' anomalies is found on the white matter, and more specifically on the corpus callosum. Indeed, 24.5, 31 and 43% of the white matter and 58.5, 51.6 and 43.9% of the corpus callosum is described by an atypical cluster.

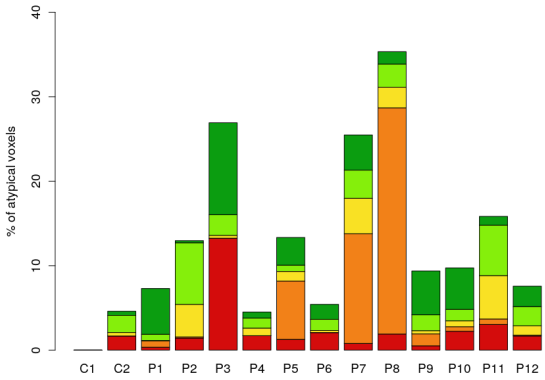


FIGURE 6.10: Signatures of the subjects (brain) after post-treatment.

Additional analysis shows that the K_A1 cluster, exhibiting the largest T1 relaxation times, is predominantly present on the subcortical structures. This K_A1 cluster takes up 52% of the subcortical structures in patient P3 who exhibits the largest quantities of this cluster.

Cluster K_A2, exhibited in large percentages by patients P5, P7 and P8 describes abnormalities both in white matter (outside of the corpus callosum) and cortical gray matter lobes. This cluster has the lowest T1 relaxation times. Cluster K_A3, identified by dramatically reduced FA, is present on the same structures but on smaller proportion.

Finally, the K_A4 and K_A5 clusters primarily describe anomalies in the corpus callosum and, in a smaller percentage, anomalies in the subcortical structures. The two clusters are on positive extremes regarding perfusion, K_A4 has the greatest pCBF average and K_A5 the smallest. Both clusters, however, share high MTT averages.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Subcortical	12.4	23.3	56.0	15.8	18.2	18.4	17.1	17.3	12.6	20.6	30.6	21.3
Frontal	1.0	9.8	4.3	1.4	6.3	1.2	20.4	38.5	8.0	2.4	13.5	5.3
Parietal	0.0	9.3	13.0	1.1	3.9	1.2	13.3	36.6	5.9	2.9	10.6	3.3
Temporal	0.1	9.3	11.7	1.0	20.1	1.3	21.4	23.3	1.9	3.7	14.9	2.6
Occipital	0.2	30.5	26.6	2.2	23.0	1.8	20.7	18.4	3.1	2.0	23.8	4.7
White Matter	6.9	5.5	24.5	2.5	13.5	3.6	31.0	43.0	9.8	11.1	11.2	5.9
Cing/Ins.	1.3	16.1	36.9	13.9	4.1	11.7	16.1	10.0	8.9	9.8	20.5	13.7
C. Callosum	13.8	25.3	58.5	25.2	35.9	36.3	51.6	43.9	30.2	33.5	41.6	32.3

TABLE 6.8: Percentages of abnormal voxels after post-processing in each patient in the subcortical structures, the gray matter lobes, the white matter and the corpus callosum.

6.6.3 Results for the subcortical structures

A new reference model was fitted to the healthy voxels included on the subcortical structures exclusively. According to the slope heuristic, the model with $K_H = 7$ clusters offered

the best compromise between maximizing the log-likelihood of the model and limiting complexity.

The clusters are presented on different parameter distributions on Figure 6.11. This time the projection of the clusters on the subjects anatomy shows quite different results, despite the fact that both control subjects are females with similar ages.

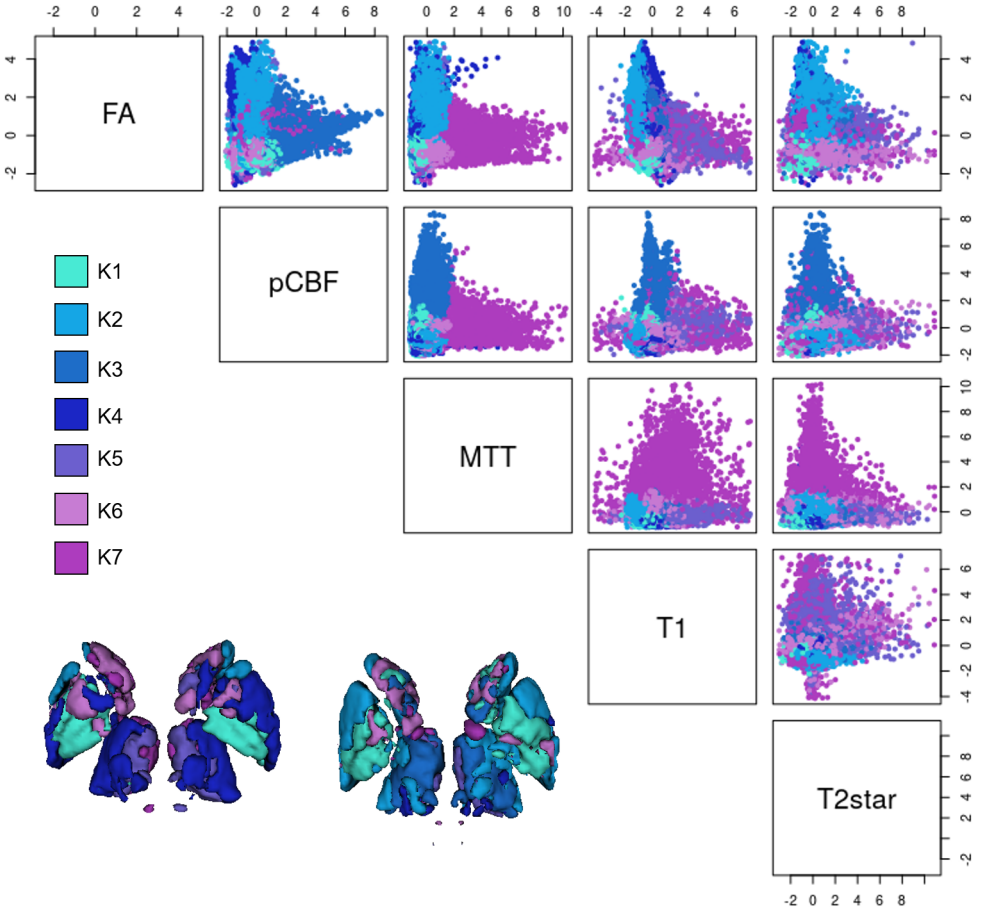


FIGURE 6.11: Scatter plots of the parameter features belonging to the healthy subjects. The colors represent the f_H clusters.

The structures predominantly described by cluster K_H4 on C1 are represented K_H2 and K_H3 on C2. By observing the mean values of the clusters in every parameter (Table 6.9) we notice that in average C1 has reduced perfusion (CBF), MTT and T2* relaxation times on the RN, SN, STN and thalamus compared to C2.

Once again the FPR was fixed at 5% and all voxels from the controls and patients above the threshold associated to this FPR, that is $t_{5\%}$, were deemed as atypical. An atypical model of $K_A = 4$ clusters was chosen by the slope heuristic to represent the outlying voxels.

The cluster centers in every dimension and their size in the mixture are reported on Table 6.10.

	K_H1	K_H2	K_H3	K_H4	K_H5	K_H6	K_H7
FA (%)	0.132	0.317	0.267	0.325	0.272	0.137	0.222
pCBF ($ml/100g/min$)	25.58	24.01	44.24	17.00	27.98	22.51	25.02
MTT (s)	2.82	5.21	4.40	3.21	4.63	5.87	12.85
T1 (ms)	1641	1422	1870	1870	2304	1868	2285
T2star (ms)	32.83	44.60	47.31	40.09	53.43	52.92	47.61
Size (%)	18.85	17.82	14.44	14.31	12.63	12.37	9.6

TABLE 6.9: Characteristics of the reference subcortical structures model clusters: mean value for every parameter and overall size in the model.

	K_A1	K_A2	K_A3	K_A4
FA(scalar)	0.229	0.188	0.355	0.243
pCBF($ml/100g/min$)	25.17	23.30	24.09	43.48
MTT(s)	5.72	8.83	9.80	13.21
T1(ms)	923	2480	1323	1520
T2star(ms)	37.53	81.65	59.94	45.65
Size(%)	31.98	22.78	22.63	22.61

TABLE 6.10: Characteristics of the atypical model clusters: mean value for every parameter and overall size in the model.

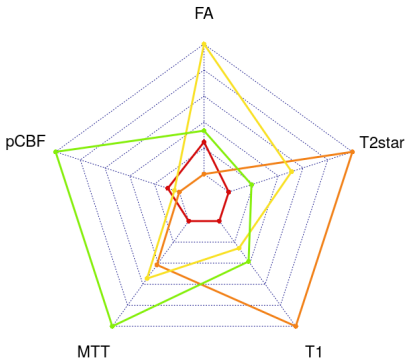


FIGURE 6.12: Clusters comparison by parameter means.

From 109 572 (27.5%) initial atypical samples, 74 773 (18.8%) survived post-treatment. The atypical signatures of all subjects after post-treatment are shown on Figure 6.10, in addition to the projection of the atypical clusters on the patients anatomical structures.

Patients P5, P7 and P8 present a very important percentage of K_A1 abnormalities characterized by the smallest T1 relaxation times in average in comparison with all other healthy and atypical clusters. These clusters are mainly found on the putamen, the caudate nucleus and the globus pallidus (external and internal segments) but also in the thalamus, SN and STN for P8. The three patients are males, which could explain this type ob abnormalities, however, patients P1, P3 and P10, who are males as well, do not present K_A1 clusters of comparable size.

The most generalized abnormalities are characterized by the K_A4 cluster. They are present on the thalamus, the caudate nucleus and the internal segment of the globus pallidus and suggest high CBF and MTT values.

On Table 6.11 is reported the percentage of voxels in every structure considered as atypical. We can observe generalized abnormalities in the caudate nucleus and the thalamus. Patients P7, P8 and P5 present huge amounts of abnormalities in almost all of the subcortical structures studied, but the colliculus.

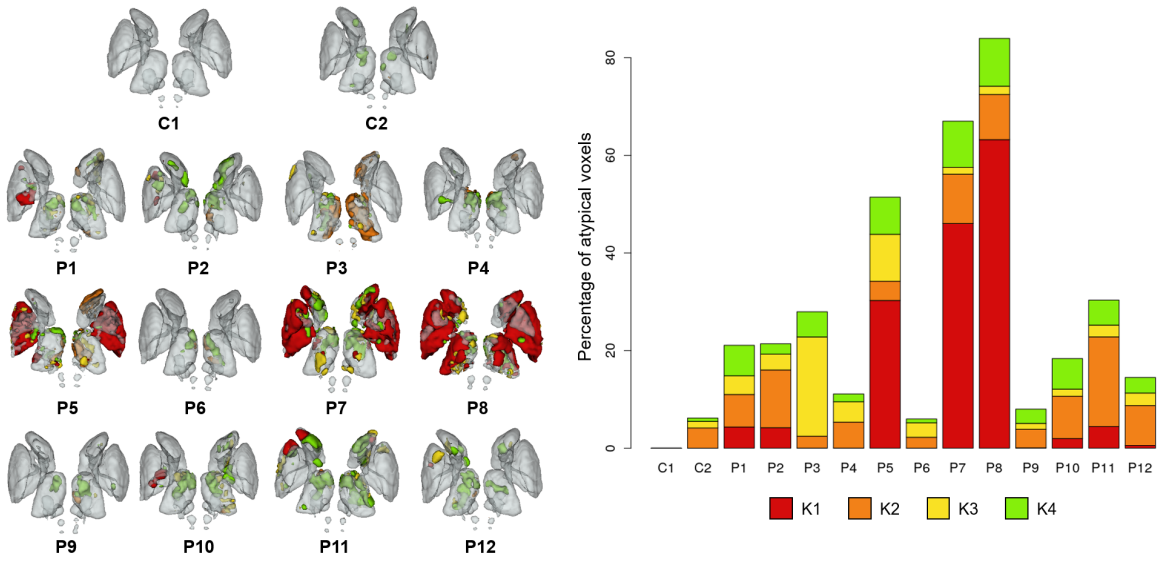


FIGURE 6.13: Signatures of the subjects (subcortical structures) after post-treatment.

%	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
RN	4.7	0	20.6	0	50.7	0.0	18.3	30.8	0	5.3	0	0
SN	3.8	0	24.7	0	38.7	0.0	59.8	57.8	0	3.7	2.2	3.2
STN	0	0	0	0	1.1	0	73.3	85.5	0	0	7.8	2.0
Cau	10.2	11.5	12.1	3.9	18.0	0.3	61.5	67.0	3.4	9.0	33.8	9.1
Put	1.5	6.5	1.5	0	63.3	0.2	74.4	83.3	0	0.1	12.7	2.4
GPe	20.4	2.1	0	1.8	81.7	0.0	79.8	96.0	1.6	19.7	0	0
GPi	15.8	3.8	0	9.8	82.5	0.0	90.3	91.7	0	19.3	0	0
Th	18.4	12.4	39.7	17.7	25.4	8.3	18.5	62.1	7.4	18.4	25.4	12.0
CS	0	0	21.2	13.5	9.5	0	0	0	16.3	0	17.4	11.9
CI	0	0	31.4	0	0	0	0	0	0	0	0	0

TABLE 6.11: Percentages of abnormal voxels after post-processing in each patient subcortical structure.

6.6.4 Discussion

On this application we employed our anomaly detection pipeline based on MMST models to detect abnormalities on a small dataset with multiple physiological parameters. These results should be interpreted delicately as the reference model was not learnt from a representative sample of the population, the available data appertain to two female controls of 58 and 56 years of age.

Nevertheless, the preliminary results here reported are encouraging and plausible with respect to the literature. In the whole brain analysis, the anomalies characterized by increased T1 relaxation times were found on the subcortical structures, which could indicate microstructural changes in the tissues [17].

Mean transit times on both analysis were clearly higher for PD patients, one explication

could be a decrease in cerebral perfusion pressure which is compensated by vasodilatation and in turn produces a reduction of CBF and prolonged MTT (Monro-Kellie theory). On a side note, the predilection for pCBF over relative CBF maps extracted from DSC perfusion sequences should be further studied.

Patients P5, P7 and P8 presented the biggest percentage of anomalies with respect to both the brain and the subcortical structures healthy models. While this could be influenced by their sex (they are three males of 62; 58 and 63 years), the other males presented similar signatures to the females for both the brain and subcortical structures study.

All in all, the combination of diffusion, perfusion and tissue relaxation times seems to be all the more informative than the separate measures.

The addition of more healthy and pathological subjects to the study cohort is of utmost importance to produce robust results that convey more information on the pathophysiology of the early stages of Parkinson's Disease.

6.7 Conclusions

MMST models have proven to be well suited to characterize healthy and pathological tissues based on physiological parameters. The developed pipeline for anomaly detection produced coherent results with respect to the literature in the cases where more than one MR modality was employed. Indeed, on the early stages of Parkinson's Disease, anomalies are likely to be subtle and diffusion properties alone appear to be insufficient to characterize PD patients.

Our method, in contrast to supervised machine learning techniques, does not rely on a large set of annotated data, which are difficult to obtain in medical contexts. Moreover, it relies on interpretable statistical tools that can be tuned and compared, providing a model of physical properties alterations.

In addition, it does not require ground-truth comparison, making it a valuable tool for the exploration of physiological changes in a wide variety of pathologies.

Computational developments are on their way to optimize the estimation of MMST models on large datasets. Methodological changes have been evoked as well. Coresets could be employed to reduce the size of the datasets before model estimation [129], that is, the dataset can be summarized by a fraction of its instances with corresponding weights. Also, fractionation and refractionation [130] could be used to constitute a general model by merging several partial models fitted to subsets of the dataset.

GENERAL DISCUSSION AND PERSPECTIVES

Three very different methods to study different datasets were explored in this thesis. In this section we attempt to discuss our main findings within the scope of the project.

Regarding our methods

Volume, deformation and surface-based morphometry

The study of brain morphometry through MR imaging is an undoubtedly valuable technique. It has allowed researchers to quantify anatomical features of the brain in terms of shape, mass, volume; to study local changes in gray matter, atrophic changes, cortical thinning and other alterations of the cortical ribbon. Numerous studies both on healthy (ageing, learning) and pathological populations have used VBM, DBM or SBM and achieved remarkable results. Most employ widely recognized pipelines like CAT12 while others develop their own pipelines.

Methods heterogeneity is one of the principal challenges we encountered while studying early changes in Parkinson's disease patients. On one side the variability of results may be a direct product of the differences in the algorithms employed. This leads us to believe that investigations employing state-of-the-art established pipelines like CAT12 to study large size populations should be privileged. On the other side, novel algorithms to analyze brain morphometry are on constant development and the community should be open to test them and promote improvements.

The availability of large open datasets like PPMI and the current tendency for data sharing are of great aid to study morphometry and to develop new methods. Learning-based segmentation methods such as volBrain's largely benefit from these databases to produce a highly accurate segmentation of multiple subcortical structures.

Volume, deformation and surface based morphometry are excellent tools for the investigation of structural changes in a population. The absence of significant morphological differences in this study does not indicate the contrary. It is merely an indication that the spatial resolution achieved by the current MR settings is not sufficient to detect subtle alterations

or even that morphological changes probably occur in the later stages of the disease. A longitudinal investigation could be implemented to further study this claims.

Auto-encoders

Deep-learning methods have gained great popularity in the medical imaging field. While most networks have been applied to structure segmentation and supervised learning of lesions, the development of unsupervised deep-learning architectures, such as auto-encoders, holds great promise for the democratization of these techniques. Indeed, lesion labelling is both time and resource-consuming, plus it is not always possible.

In our application, we trained auto-encoder architectures to reconstruct healthy diffusion MR slices and then compared the reconstruction error profiles of early PD patients compared to a population of test controls. The hypothesis being that the amount of abnormalities in PD reconstruction maps would be significantly higher that that of the healthy test population.

There are two main issues with the investigation. The first one is the amount of healthy data available for training and specially for testing. The reconstructions achieved by the sAE were quite satisfactory and those of the sVAE were quite good as well, but perhaps more training data (41 controls) would have allowed the dVAE to yield better results. This being said we considerably increased the learning and testing data available by considering 2D images (40 per patients) and separating the two hemispheres.

During testing, we compared the reconstructions associated to 15 controls (120 hemisphere images) to those of 129 PD patients (10 320 hemisphere images). All the necessary precautions were taken to correctly carry out this very unbalanced classification task, notably we followed a cross-validation procedure and chose an appropriate metric. Nevertheless, more test controls would give us more statistical confidence.

The second issue is the bias introduced by our thresholds. We defined a threshold to identify abnormalities at the voxel level. Any voxel with a reconstruction error above the value corresponding to the 98% quantile of the distribution of errors in the healthy test population was considered as abnormal. Then we defined another threshold to discriminate healthy from pathological subjects using a ROC curve. The threshold corresponded to the optimal cut-point of the curve which in this case was the point minimizing the Euclidean distance between the curve and perfect sensitivity and specificity. Both thresholds are tailored to the data but probably over-summarize the information included on the reconstruction error maps.

Future research should devise a strategy to use the reconstruction error maps in raw form, perhaps through voxel-level analysis, or even to train a classifier directly from the latent vectors produced by the encoding block of the architecture.

Mixture models

Mixtures models are statistical models commonly used in clustering tasks where the goal is identify sub-populations within an overall data population, every component in the mixture corresponds to an univariate or multivariate distribution that describes one of the sub-populations under consideration. These models have been widely used in economics, botany, medicine and many other fields.

Among the various possible mixtures of distributions, the ones mixing multiple scaled t-distributions (MST) [14] have been shown to be better suited to characterize physiological data than other elliptical distributions. The variable tail-weight in these distributions gives them a flexible nature specially useful when studying outliers.

In three different applications we applied an in-house developed pipeline that utilizes MST mixtures for anomaly detection and characterization. A MST mixture model is fitted to healthy data to constitute a reference model to which new data can be compared. The log-likelihood of the patients data to belong to the reference model is computed and any voxels below a determined threshold are labeled as atypical. The atypical voxels are then characterized by another MST mixture model.

In one application we modeled the FA and MD diffusion parameters of the same population that was studied with auto-encoder reconstructions. No relevant differences were found in the percentage of atypical voxels found in PD patients versus controls nor on classes to which the atypical voxels belonged.

In the two remaining applications the multivariate data fitted by the model came from more than one modality. The results were interesting; atypical clusters of the same class emerged in similar locations for all patients despite the fact that no spatial information is entered in the model. However, the studies included very few participants (3 controls and 10 patients in one and 2 controls and 12 patients in the other) so the results are merely preliminary.

In future research a larger cohort of multivariate data should be studied. Notably new acquisitions of the SignaPark project could be added to the study.

This clustering-based anomaly detection framework appears to be a great means to study the physiopathological changes affecting the brains of newly diagnosed Parkinson's disease patients. The interaction of different features is taken into account and the models are easily interpretable. Furthermore, the method can be easily applied to other pathologies. Finally, if the reference and atypical models are built from a representative population, they can be used in a direct manner to classify new subjects.

Regarding the physiopathology of Parkinson's Disease

Structural changes

No structural changes were found on the morphological assessment presented on Chapter 4. Two different methods were employed for tissue segmentation, yet neither of the two produced significant differences between the local gray matter concentrations (VBM) of PD patients versus controls in the brain or in the subcortical structures studied. One of this methods is included in a pipeline (CAT12) that also allows for deformation (DBM) and cortical analysis (SBM). No significant atrophy was found using DBM and neither did any cortical ribbon differences in 'de novo' PD patients using SBM.

It is our belief that quantitative data is a more appropriate alternative to study the Parkinson's disease, specially on 'de novo' patients where any changes are expected to be subtle.

As a matter of fact, the preliminary results presented on Section 6.6 give us a glimpse of the potential discriminant information contained in T1 and T2* relaxometry maps that could indicate microstructural changes in the tissues as proposed by Deoni and colleagues [17].

Diffusion changes

DTI is the only MR modality that was explored through two different techniques in this thesis. First through a reconstruction-based anomaly detection pipeline employing autoencoder architectures and secondly by the clustering of MR diffusion properties with MMST models. In both cases the DTI scans were pooled from the PPMI database.

In the first study we found a larger percentage of abnormal instances in the brains of PD patients compared to controls, this resulted on an average g-mean classification performance of 64.6% in a ten-fold cross-validation. Most of the abnormalities were found in the white matter, which came as no surprise due to the nature of DTI. In addition, we found that the patients exhibiting mild cognitive impairment presented more abnormalities and in different locations than those with normal cognition.

In the second study, no relevant differences emerged between the 'de novo' PD patients and controls but most atypical instances were found in the corps callosum, a white matter structure responsible for the communication between the two hemispheres.

In summary, we found evidence of white matter abnormalities in the early stages of PD, however further research should concentrate in characterizing these differences in patients without cognitive impairment. Moreover, the combination of diffusion parameters along with other physiological properties should improve classification.

Perfusion changes

While no independent studies of perfusion were performed, the integration of perfusion parameters in our clustering-based anomaly detection approach gave a clear indication its potential to classify PD patients from controls, even at the early stages of the disease.

Furthermore, its interaction with other physiological properties should be of benefit to classification performances.

In further work, we shall investigate perfusion changes on larger cohorts. Also, the standardization of DSC relative parameter maps (rCBV, rCBF) should be explored.

...

Ton conclude, the integration of quantitative MR measures reflecting multiple properties of the brain is certainly key to detect abnormalities in the brains of PD patients from the earliest stages possible. Diffusion, perfusion and relaxometry measures can paint an informative picture of the normal physiology of the brain. Moreover, mixture models and auto-encoders are great candidates to detect abnormalities linked to the pathophysiology of PD from this data.

Bibliography

[1] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(September):1–58, 2009.

[2] Raghavendra Chalapathy and Sanjay Chawla. Deep Learning for Anomaly Detection: A Survey. arXiv, 2019.

[3] Marco A.F. Pimentel, David A. Clifton, Lei Clifton, and Lionel Tarassenko. A review of novelty detection. *Signal Processing*, 99:215–249, 2014.

[4] Alexis Arnaud, Florence Forbes, Nicolas Coquery, and Nora Collomb. Fully Automatic Lesion Localization and Characterization: Application to Brain Tumors Using Multiparametric MRI Data. *IEEE Transactions on Medical Imaging*, 2017.

[5] José V. Manjón and Pierrick Coupé. volBrain: An Online MRI Brain Volumetry System. *Frontiers in Neuroinformatics*, 10:30, jul 2016.

[6] Parkinson's Foundation. Parkinson's Prevalence Project, 2018.

[7] Yashar Zeighami, Miguel Ulla, Yasser Iturria-Medina, Mahsa Dadar, Yu Zhang, Kevin Michel Herve Larcher, Vladimir Fonov, Alan C. Evans, D. Louis Collins, and Alain Dagher. Network structure of brain atrophy in de novo parkinson's disease. *eLife*, 4, 2015.

[8] Rene Seiger, Sebastian Ganger, Georg S. Kranz, Andreas Hahn, and Rupert Lanzenberger. Cortical Thickness Estimations of FreeSurfer and the CAT12 Toolbox in Patients with Alzheimer's Disease and Healthy Controls. *Journal of Neuroimaging*, 28(5):515–523, sep 2018.

[9] Anders Eklund, Thomas E Nichols, and Hans Knutsson. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, 113(28):7900–5, jul 2016.

[10] J. Donald Tournier, Robert Smith, David Raffelt, Rami Tabbara, Thijs Dhollander, Maximilian Pietsch, Daan Christiaens, Ben Jeurissen, Chun Hung Yeh, and Alan Connelly. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation, nov 2019.

[11] Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Deep Autoencoding Models for Unsupervised Anomaly Segmentation in Brain MR Images. arXiv, 2018.

[12] Norbert Schuff, I. Wei Wu, Shannon Buckley, Eric D. Foster, Christopher S. Coffey, Darren R. Gitelman, Susan Mendick, John Seibyl, Tanya Simuni, Yu Zhang, Joseph Jankovic, Christine Hunter, Caroline M. Tanner, Linda Rees, Stewart Factor, Daniela Berg, Isabel Wurster, Katharina Gauss, Fabienne Sprenger, Klaus Seppi, Werner Poewe, Brit Mollenhauer, Susanne Knake, Zoltan Mari, Arita Mccoy, Madelaine Ranola, and Kenneth Marek. Diffusion imaging of nigral alterations in early Parkinson's disease with dopaminergic deficits. *Movement Disorders*, 30(14):1885–1892, 2015.

[13] Marta M Correia, Timothy Rittman, Christopher L Barnes, Ian T Coyle-Gilchrist, Boyd Ghosh, Laura E Hughes, and James B Rowe. Towards accurate and unbiased imaging-based differentiation of Parkinson's disease, progressive supranuclear palsy and corticobasal syndrome. *Brain Communications*, 2020.

[14] Florence Forbes and Darren Wraith. A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: application to robust clustering. *Statistics and Computing*, 24(6):971–984, 2014.

[15] María A Fernández-Seara, Elisa Mengual, Marta Vidorreta, Maite Aznárez-Sanado, Francis R Loayza, Federico Villagra, Jaione Irigoyen, and María A Pastor. Cortical hypoperfusion in Parkinson's disease assessed using arterial spin labeled perfusion MRI. *NeuroImage*, 59:2743–2750, 2012.

[16] Patrice Péran, Andrea Cherubini, Francesca Assogna, Fabrizio Piras, Carlo Quattrocchi, Antonella Peppe, Pierre Celsis, Olivier Rascol, Jean Franois Démonet, Alessandro Stefani, Mariangela Pierantozzi, Francesco Ernesto Pontieri, Carlo Caltagirone, Gianfranco Spalletta, and Umberto Sabatini. Magnetic resonance imaging markers of Parkinson's disease nigrostriatal signature. *Brain*, 133(11):3423–3433, 2010.

[17] Sean C L Deoni. Quantitative Relaxometry of the Brain. *Topics in Magnetic Resonance Imaging*, 21:101–113, 2010.

[18] Christopher G. Goetz. The history of Parkinson's disease: Early clinical descriptions and neurological therapies. *Cold Spring Harbor Perspectives in Medicine*, 1(1), 2011.

[19] Werner Poewe. The natural history of Parkinson's disease. *Journal of Neurology*, 253(SUPPL. 7):2–6, 2006.

[20] World Health Organization. Neurological disorders: a public health approach. *Neurological disorders: public health challenges.*, pages 41–176, 2006.

[21] Werner Poewe, Klaus Seppi, Caroline M. Tanner, Glenda M. Halliday, Patrik Brundin, Jens Volkmann, Anette Eleonore Schrag, and Anthony E. Lang. Parkinson disease. *Nature Reviews Disease Primers*, 3(1):1–21, mar 2017.

[22] A Elbaz, L Carcaillon, S Kab, and F Moisan. Epidemiology of Parkinson's disease. *Revue Neurologique*, 172:14–26, 2016.

[23] Sara Meoni, Antonella Macerollo, and Elena Moro. Sex differences in movement disorders Nature reviews | Neurology. *Nature Reviews Neurology*, 16:84–96, 2020.

[24] Naveed Malek, Diane M.A. Swallow, Katherine A. Grosset, Michael A. Lawton, Sarah L. Marrinan, Alexander C. Lehn, Catherine Bresner, Nin Bajaj, Roger A. Barker, Yoav Ben-Shlomo, David J. Burn, Thomas Foltynie, John Hardy, Huw R. Morris, Nigel M. Williams, Nicholas Wood, and Donald G. Grosset. Tracking Parkinson's: study design and baseline patient data. *Journal of Parkinson's Disease*, 5(4):947–959, nov 2015.

[25] Ronald B. Postuma. *Nonmotor Aspects of Parkinson's Disease—How Do They Help Diagnosis?*, volume 133. Elsevier Inc., 1 edition, 2017.

[26] Emmanuelle Bellot. *Le colliculus supérieur dans la maladie de Parkinson : un biomarqueur possible ?* PhD thesis, Université Grenoble Alpes, 2017.

[27] Ying Jiao Zhao, Hwee Lin Wee, Yiong-Huak Chan, Soo Hoon Seah, Wing Lok Au, Puay Ngoh Lau, Emmanuel Camara Pica, Shu Chuen Li, Nan Luo, and Louis C.S. Tan. Progression of Parkinson's disease as evaluated by Hoehn and Yahr stage transition times. *Movement Disorders*, 25(6):710–716, apr 2010.

[28] Margaret M Hoehn and Melvin D Yahr. Parkinsonism : onset, progression, and mortality Parkinsonism: onset, progression, and mortality. *Neurology*, 17(5):427–442, 1967.

[29] Thomas R. Barber, Johannes C. Klein, Clare E. Mackay, and Michele T.M. Hu. Neuroimaging in pre-motor Parkinson's disease. *NeuroImage: Clinical*, 15:215–227, jan 2017.

[30] R. B. Postuma, A. E. Lang, J. F. Gagnon, A. Pelletier, and J. Y. Montplaisir. How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder. *Brain*, 135(6):1860–1870, jun 2012.

[31] Glenda Halliday, Andrew Lees, and Matthew Stern. Milestones in Parkinson's disease-Clinical and pathologic features. *Movement Disorders*, 26(6):1015–1021, may 2011.

[32] Jeff M. Bronstein, Michele Tagliati, Ron L. Alterman, Andres M. Lozano, Jens Volkmann, Alessandro Stefani, Fay B. Horak, Michael S. Okun, Kelly D. Foote, Paul Krack, Rajesh Pahwa, Jaimie M. Henderson, Marwan I. Hariz, Roy A. Bakay, Ali Rezai, William J. Marks, Elena Moro, Jerrold L. Vitek, Frances M. Weaver, Robert E. Gross, and Mahlon R. DeLong. Deep brain stimulation for Parkinson disease an expert consensus and review of key issues. *Archives of Neurology*, 68(2):165–171, feb 2011.

[33] Elena Moro, Andres M Lozano, Pierre Pollak, Yves Agid, Stig Rehncrona, Jens Volkmann, Jaime Kulisevsky, Jose A Obeso, Alberto Albanese, Marwan I Hariz,

Niall P Quinn, Jans D Speelman, Alim L Benabid, Valerie Fraix, Alexandre Mendes, Marie-Laure Welter, Jean-Luc Houeto, Philippe Cornu, Didier Dormont, Annalena L Tornqvist, Ron Ekberg, Alfons Schnitzler, Lars Timmermann, Lars Wojtecki, Andres Gironell, Maria C Rodriguez-Oroz, Jorge Guridi, Anna R Bentivoglio, Maria F Contarino, Luigi Romito, Massimo Scerrati, Marc Janssens, Anthony E Lang, and Toronto Western Hospital. Long-Term Results of a Multicenter Study on Subthalamic and Pallidal Stimulation in Parkinson's Disease Potential conflict of interest: None reported. *Movement Disorders*, 25(5):578–586, 2010.

[34] Usman Saeed, Jordana Compagnone, Richard I Aviv, Antonio P Strafella, Sandra E Black, Anthony E Lang, and Mario Masellis. Imaging biomarkers in Parkinson's disease and Parkinsonian syndromes: current and emerging concepts. *Translational neurodegeneration*, 6:8, 2017.

[35] Marios Politis. Neuroimaging in Parkinson disease: from research setting to clinical practice. *Nature Publishing Group*, 10:708–722, 2014.

[36] Paul Tofts. *Quantitative MRI of the brain*. Wiley, 1 edition, aug 2005.

[37] A. Jon Stoessl, Stephane Lehericy, and Antonio P. Strafella. Imaging insights into basal ganglia function, Parkinson's disease, and dystonia. *PubMed Central CANADA*, 49(2):743–750, 2014.

[38] Nadya Pyatigorskaya, Cécile Gallea, Daniel Garcia-Lorenzo, Marie Vidailhet, and Stéphane Lehericy. A review of the use of magnetic resonance imaging in Parkinson's disease. *Therapeutic Advances in Neurological Disorders*, 7(4):206–220, 2014.

[39] Salil Soman, Jose A Bregni, Francisco Marroquin, Guatemala Berkin Bilgic, Ursula Nemec, Audrey Fan, Zhe Liu, Robert L Barry, Jiang Du, Keith Main, Jerome Yesavage, Berkin Bilgic, Maheen M Adamson, Michael Moseley, and Yi Wang. Susceptibility-Based Neuroimaging: Standard Methods, Clinical Applications, and Future Directions. *Current Radiology Reports*, 5(3), 2017.

[40] Denis Le Bihan and Heidi Johansen-Berg. Diffusion MRI at 25: Exploring brain tissue structure and function, jun 2012.

[41] Christopher Summerfield, Carme Junqué, Eduardo Tolosa, Pilar Salgado-Pineda, Beatriz Gómez-Ansón, Maria José Martí, Pau Pastor, Blanca Ramírez-Ruíz, and José Mercader. Structural Brain Changes in Parkinson Disease With Dementia. *Archives of Neurology*, 62(2):281, feb 2005.

[42] Eric M Nyberg, Jody Tanabe, Justin M Honce, Theodore Krmpotich, Erika Shelton, Jessica Hedeman, and Brian D Berman. Morphologic changes in the mesolimbic pathway in Parkinson's disease motor subtypes. *Parkinsonism and Related Disorders*, 21:536–540, 2015.

[43] Andrius Radziunas, Vytenis Pranas Deltuva, Arimantas Tamasauskas, Rymante Gleizniene, Aiste Pranckeviciene, Kestutis Petrikonis, and Adomas Bunevicius. Brain MRI morphometric analysis in Parkinson's disease patients with sleep disturbances. *BMC Neurology*, 18(1):88, dec 2018.

[44] X Jia, P Liang, Y Li, L Shi, D Wang, and K Li. Longitudinal Study of Gray Matter Changes in Parkinson Disease. *American journal of neuroradiology*, 36(12):2219–26, dec 2015.

[45] Xingfeng Li, Yue Xing, Stefan T. Schwarz, and Dorothee P. Auer. Limbic grey matter changes in early Parkinson's disease. *Human Brain Mapping*, 00, 2017.

[46] P. Borghammer, K. Østergaard, P. Cumming, A. Gjedde, A. Rodell, N. Hall, and M. M. Chakravarty. A deformation-based morphometry study of patients with early-stage Parkinson's disease. *European Journal of Neurology*, 17(2):314–320, 2010.

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

[48] Yashar Zeighami, Seyed-Mohammad Fereshtehnejad, Mahsa Dadar, D Louis Collins, Ronald B Postuma, Mi Si C, and Alain Dagher. A clinical-anatomical signature of Parkinson's disease identified with partial least squares and magnetic resonance imaging. *NeurImage*, 190:69–78, 2017.

[49] Yashar Zeighami, Seyed Mohammad Fereshtehnejad, Mahsa Dadar, D. Louis Collins, Ronald B. Postuma, and Alain Dagher. Assessment of a prognostic MRI biomarker in early de novo Parkinson's disease. *NeuroImage: Clinical*, 24:101986, jan 2019.

[50] Sule Tinaz, Maureen G. Courtney, and Chantal E. Stern. Focal cortical and subcortical atrophy in early Parkinson's disease. *Movement Disorders*, 26(3):436–441, feb 2011.

[51] Joana Braga Pereira, Naroa Ibarretxe-Bilbao, Maria Jose Marti, Yaroslau Compta, Carme Junqué, Nuria Bargallo, and Eduardo Tolosa. Assessment of cortical degeneration in patients with Parkinson's disease by voxel-based morphometry, cortical folding, and cortical thickness. *Human Brain Mapping*, 33(11):2521–2534, 2012.

[52] Naroa Ibarretxe-Bilbao, Carme Junque, Barbara Segura, Hugo C. Baggio, Maria J. Marti, Francesc Valldeoriola, Nuria Bargallo, and Eduardo Tolosa. Progression of cortical thinning in early Parkinson's disease. *Movement Disorders*, 27(14):1746–1753, 2012.

[53] Frederic Sampedro, Juan Marín-Lahoz, Saul Martínez-Horta, Javier Pagonabarraga, and Jaime Kulisevsky. Dopaminergic degeneration induces early posterior cortical thinning in Parkinson's disease. *Neurobiology of Disease*, 124:29–35, 2018.

[54] Yuanchao Zhang, Jiuquan Zhang, Jinping Xu, Xiu Wu, Yanling Zhang, Hua Feng, Jian Wang, and Tianzi Jiang. Cortical Gyrification Reductions and Subcortical Atrophy in Parkinson's Disease. *Movement Disorders*, 2013.

[55] Nicholas W. Sterling, Ming Wang, Lijun Zhang, Eun Young Lee, Guangwei Du, Mechelle M. Lewis, Martin Styner, and Xuemei Huang. Stage-dependent loss of cortical gyrification as Parkinson disease "unfolds". *Neurology*, 86(12):1143–1151, mar 2016.

[56] Stefan T. Schwarz, Maryam Abaei, Vamsi Gontu, Paul S. Morgan, Nin Bajaj, and Dorothee P. Auer. Diffusion tensor imaging of nigral degeneration in Parkinson's disease: A region-of-interest and voxel-based study at 3 T and systematic review with meta-analysis. *NeuroImage: Clinical*, 3:481–488, 2013.

[57] Guangwei Du, Mechelle M. Lewis, Martin Styner, Michele L. Shaffer, Suman Sen, Qing X. Yang, and Xuemei Huang. Combined R2* and Diffusion Tensor Imaging Changes in the Substantia Nigra in Parkinson's Disease. *Movement Disorders*, 26(9):1627–1632, 2011.

[58] Claire J. Cochrane and Klaus P. Ebmeier. Diffusion tensor imaging in parkinsonian syndromes: A systematic review and meta-analysis, feb 2013.

[59] Cyril Atkinson-Clement, Serge Pinto, Alexandre Eusebio, and Olivier Coulon. Diffusion tensor imaging in Parkinson's disease: Review and meta-analysis. *NeuroImage: Clinical*, 16(July):98–110, 2017.

[60] Martin Cousineau, Pierre Marc Jodoin, Félix C. Morency, Verena Rozanski, Marilyn Grand'Maison, Barry J. Bedell, and Maxime Descoteaux. A test-retest study on Parkinson's PPMI dataset yields statistically significant white matter fascicles. *NeuroImage: Clinical*, 16(July):222–233, 2017.

[61] Aron S. Talai, Jan Sedlacik, Kai Boelmans, and Nils D. Forkert. Widespread diffusion changes differentiate Parkinson's disease and progressive supranuclear palsy. *NeuroImage: Clinical*, 20(July):1037–1043, 2018.

[62] Rachel P Guimarães, Brunno M Campos, Thiago J De Rezende, Luiza Piovesana, Paula C Azevedo, Augusto C Amato-Filho, and Anelyssa D'abreu. Is Diffusion Tensor Imaging a Good Biomarker for Early Parkinson's Disease? *Frontiers in Neurology*, 9:1–8, 2018.

[63] Jingqiang Wang, Fan Zhang, Changchen Zhao, Qingrun Zeng, Jianzhong He, Lauren J O'donnell, and Yuanjing Feng. Investigation of local white matter abnormality in Parkinson's disease by using an automatic fiber tract parcellation. *Behavioural Brain Research*, 394:1–10, 2020.

[64] William J. Jagust, Bruce R. Reed, Eileen M. Martin, Jamie L. Eberling, and Ruth A. Nelson-abbott. Cognitive function and regional cerebral blood flow in parkinson's disease. *Brain*, 115(2):521–537, apr 1992.

[65] M J Firbank, S J Colloby, D J Burn, I G Mckeith, and J T O'brien. Regional cerebral blood flow in Parkinson's disease with and without dementia. 2003.

[66] Young-Do Kim, Hyeonseok S. Jeong, In-Uk Song, Yong-An Chung, Eun Namgung, and Yong-Duk Kim. Brain perfusion alterations in depressed patients with Parkinson's disease. *Annals of Nuclear Medicine*, 30(10):731–737, 2016.

[67] Tracy R Melzer, Richard Watts, Michael R Macaskill, John F Pearson, Sina Rü, Toni L Pitcher, Leslie Livingston, Charlotte Graham, Ross Keenan, Ajit Shankaranarayanan, David C Alsop, John C Dalrymple-Alford, and Tim J Anderson. Arterial spin labelling reveals an abnormal cerebral perfusion pattern in Parkinson's disease. *Brain*, 134:845–855, 2011.

[68] Laura K Teune, Remco J Renken, Bauke M de Jong, Antoon T Willemsen, Matthias J van Osch, Jos B T M Roerdink, Rudi A Dierckx, and Klaus L Leenders. Parkinson's disease-related perfusion and glucose metabolic brain patterns identified with PCASL-MRI and FDG-PET imaging. *NeuroImage. Clinical*, 5:240–4, 2014.

[69] Sarah Al-Bachari, Rishma Vidyasagar, Hedley Ca Emsley, and Laura M Parkes. Structural and physiological neurovascular changes in idiopathic Parkinson's disease and its clinical phenotypes. *Journal of Cerebral Blood Flow & Metabolism*, 37(10):3409–3421, 2017.

[70] Roberto Erro, Sara Ponticorvo, Renzo Manara, Paolo Barone, Marina Picillo, Sara Scannapieco, Giulio Cicarelli, Massimo Squillante, Giampiero Volpe, Fabrizio Esposito, and Maria Teresa Pellecchia. Subcortical atrophy and perfusion patterns in Parkinson disease and multiple system atrophy. *Parkinsonism and Related Disorders*, 72:49–55, 2020.

[71] Alessandro Arrigo, Alessandro Calamuneri, Demetrio Milardi, Enricomaria Mormina, Laura Rania, Elisa Postorino, Silvia Marino, Giuseppe Di Lorenzo, Giuseppe Pio Anastasi, Maria Felice Ghilardi, Pasquale Aragona, Angelo Quartarone, and Michele Gaeta. Visual System Involvement in Patients with Newly Diagnosed Parkinson Disease. *Radiology*, 000(0):161732, 2017.

[72] Nico J. Diederich, Vannina Pieri, Géraldine Hipp, Olivier Rufra, Sara Blyth, and Michel Vaillant. Discriminative power of different nonmotor signs in early Parkinson's disease. A case-control study. *Movement Disorders*, 25(7):882–887, 2010.

[73] Konrad Szewczyk-Krolikowski, Ricarda A.L. Menke, Michal Rolinski, Eugene Duff, Gholamreza Salimi-Khorshidi, Nicola Filippini, Giovanna Zamboni, Michele T.M. Hu, and Clare E. Mackay. Functional connectivity in the basal ganglia network differentiates PD patients from controls. *Neurology*, 83(3):208–214, 2014.

[74] Ludovica Griffanti, Philipp Stratmann, Michal Rolinski, Nicola Filippini, Enikő Zsoldos, Abda Mahmood, Giovanna Zamboni, Gwenaëlle Douaud, Johannes C. Klein,

Mika Kivimäki, Archana Singh-Manoux, Michele T. Hu, Klaus P. Ebmeier, and Clare E. Mackay. Exploring variability in basal ganglia connectivity with functional MRI in healthy aging. *Brain Imaging and Behavior*, 0(0):1–6, 2018.

[75] F. Du Bois Bowman, Daniel F. Drake, and Daniel E. Huddleston. Multimodal imaging signatures of Parkinson's disease. *Frontiers in Neuroscience*, 10(APR):1–11, 2016.

[76] Nicola Amoroso, Marianna La Rocca, Alfonso Monaco, Roberto Bellotti, and Sabina Tangaro. Complex networks reveal early MRI markers of Parkinson's disease. *Medical Image Analysis*, 48:12–24, 2018.

[77] Maria M Laganà, Alice Pirastru, Laura Pelizzari, Federica Rossetto, Sonia Di Tella, Niels Bergsland, Raffaello Nemni, Mario Meloni, and Francesca Baglio. Multimodal Evaluation of Neurovascular Functionality in Early Parkinson's Disease. *Frontiers in Neurology*, 11:831, 2020.

[78] Bjoern M. Eskofier, Sunghoon I. Lee, Jean Francois Daneault, Fatemeh N. Golabchi, Gabriela Ferreira-Carvalho, Gloria Vergara-Diaz, Stefano Sapienza, Gianluca Costante, Jochen Klucken, Thomas Kautz, and Paolo Bonato. Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson's disease assessment. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, volume 2016-October, pages 655–658. Institute of Electrical and Electronics Engineers Inc., oct 2016.

[79] Julià Camps, Albert Samà, Mario Martín, Daniel Rodríguez-Martín, Carlos Pérez-López, Joan M. Moreno Arostegui, Joan Cabestany, Andreu Català, Sheila Alcaine, Berta Mestre, Anna Prats, Maria C. Crespo-Maraver, Timothy J. Counihan, Patrick Browne, Leo R. Quinlan, Gearóid Laighin, Dean Sweeney, Hadas Lewy, Gabriel Vainstein, Alberto Costa, Roberta Annicchiarico, Àngels Bayés, and Alejandro Rodríguez-Molinero. Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. *Knowledge-Based Systems*, 139:119–131, jan 2018.

[80] Clayton R. Pereira, Silke A.T. Weber, Christian Hook, Gustavo H. Rosa, and Joao P. Papa. Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics. In *Proceedings - 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2016*, pages 340–346. Institute of Electrical and Electronics Engineers Inc., jan 2017.

[81] Hakan Gunduz. Deep Learning-Based Parkinson's Disease Classification Using Vocal Feature Sets. *IEEE Access*, 7:115540–115551, aug 2019.

[82] Juan Camilo Vásquez-Correa, Tomas Arias-Vergara, J. R. Orozco-Arroyave, Björn Eskofier, Jochen Klucken, and Elmar Nöth. Multimodal Assessment of Parkinson's Disease: A Deep Learning Approach. *IEEE Journal of Biomedical and Health Informatics*, 23(4):1618–1630, jul 2019.

[83] C. Stamate, G. D. Magoulas, S. Kueppers, E. Nomikou, I. Daskalopoulos, M. U. Luchini, T. Moussouri, and G. Roussos. Deep learning Parkinson's from smartphone data. In *IEEE International Conference on Pervasive Computing and Communications*, pages 31–40. Institute of Electrical and Electronics Engineers Inc., may 2017.

[84] Andong Zhan, Srihari Mohan, Christopher Tarolli, Ruth B Schneider, Jamie L Adams, Saloni Sharma, Molly J Elson, Kelsey L Spear, Alistair M Glidden, Max A Little, Andreas Terzis, E Ray Dorsey, Suchi Saria, Corresponding Author, and Malone Hall. Using Smartphones and Machine Learning to Quantify Parkinson Disease Severity The Mobile Parkinson Disease Score Supplemental content. *JAMA Neurol*, 75(7):876–880, 2018.

[85] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *Information Processing in Medical Imaging (IPMI)*, volume 10265 LNCS, pages 146–147, 2017.

[86] Varghese Alex, K. P. Mohammed Safwan, Sai Saketh Chennamsetty, and Ganapathy Krishnamurthi. Generative adversarial networks for brain lesion detection. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 10133 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 101330G, February 2017.

[87] Vu Ha Son, Ueta Daisuke, Hashimoto Kiyoshi, Maeno Kazuki, Sugiri Pranata, and Shen Sheng Mei. Anomaly Detection with Adversarial Dual Autoencoders. arXiv, 2019.

[88] Shiqi Li, Haijun Lei, Feng Zhou, Jamal Gardezi, and Baiying Lei. Longitudinal and multi-modal data learning for Parkinson's disease diagnosis via stacked sparse autoencoder. In *International Symposium on Biomedical Imaging (ISBI)*, volume 16, pages 384–387. IEEE, 2019.

[89] Sumeet Shinde, Shweta Prasad, Yash Saboo, Rishabh Kaushick, Jitender Saini, Pramod Kumar Pal, and Madhura Ingalhalikar. Predictive markers for Parkinson's disease using deep neural nets on neuromelanin sensitive MRI. *NeuroImage: Clinical*, 22(January):101748, 2019.

[90] S. Sivaranjini and C. M. Sujatha. Deep learning based diagnosis of Parkinson's disease using convolutional neural network. *Multimedia Tools and Applications*, 79(21-22):15467–15479, 2020.

[91] Ravi Bansal, Andrew J. Gerber, and Bradley S. Peterson. Brain morphometry using anatomical magnetic resonance imaging. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47(6):619–621, 2008.

[92] John Ashburner and Karl J Friston. Voxel-Based Morphometry-The Methods. *NeuroImage*, 11:805–821, 2000.

[93] Christian Gaser, Hans-Peter Volz, Stefan Kiebel, Stefan Riehemann, and Heinrich Sauer. Detecting Structural Changes in Whole Brain Based on Nonlinear Deformations-Application to Schizophrenia Research. *Neuroimage*, 10:107–113, 1999.

[94] Christian Gaser, Igor Nenadic, Bradley R Buchsbaum, Erin A Hazlett, and Monte S Buchsbaum. Deformation-Based Morphometry and Its Relation to Conventional Volumetry of Brain Lateral Ventricles in MRI. *NeuroImage*, 13:1140–1145, 2001.

[95] Robert Dahnke, Rachel Aine Yotter, and Christian Gaser. Cortical thickness and central surface estimation. *NeuroImage*, 65:336–348, 2013.

[96] José V. Manjón, Pierrick Coupé, Luis Martí-Bonmatí, D. Louis Collins, and Montserrat Robles. Adaptive non-local means denoising of MR images with spatially varying noise levels. *Journal of Magnetic Resonance Imaging*, 31(1):192–203, jan 2010.

[97] Nicholas J. Tustison, Brian B. Avants, Philip A. Cook, Yuanjie Zheng, Alexander Egan, Paul A. Yushkevich, and James C. Gee. N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, jun 2010.

[98] John Ashburner and Karl J Friston. Unified segmentation. *NeuroImage*, 26:839–851, 2005.

[99] Bruce Fischl and Anders M. Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, 97(20):11050–11055, sep 2000.

[100] Rachel Aine Yotter, Robert Dahnke, Paul M Thompson, and Christian Gaser. Topological Correction of Brain Surface Meshes Using Spherical Harmonics. *Human Brain Mapping*, 00:0–000, 2010.

[101] John Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95–113, oct 2007.

[102] Rachel A Yotter, Igor Nenadic, Gabriel Ziegler, Paul M Thompson, and Christian Gaser. Local cortical surface complexity maps from spherical harmonic reconstructions. *NeuroImage*, 56:961–973, 2011.

[103] E Luders, P M Thompson, K L Narr, A W Toga, L Jancke, and C Gaser. A curvature-based approach to estimate local gyrification on the cortical surface. *NeuroImage*, 29:1224–1230, 2006.

[104] Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, Marilyn S Albert, and Ronald J Killiany. An automated labeling system for

subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31:968–980, 2006.

[105] Cristina Scarpazza, Stefania Tognin, Silvia Frisciata, Giuseppe Sartori, and Andrea Mechelli. False positive rates in Voxel-based Morphometry studies of the human brain: Should we be worried? *Neuroscience & Biobehavioral Reviews*, 52:49–55, may 2015.

[106] Veronica Popescu, Menno M. Schoonheim, Adriaan Versteeg, Nimisha Chaturvedi, Marianne Jonker, Renee Xavier de Menezes, Francisca Gallindo Garre, Bernard M. J. Uitdehaag, Frederik Barkhof, and Hugo Vrenken. Grey Matter Atrophy in Multiple Sclerosis: Clinical Interpretation Depends on Choice of Analysis Method. *Plos One*, 11(1):1–17, 2016.

[107] Farnaz Farokhian, Iman Beheshti, Daichi Sone, and Hiroshi Matsuda. Comparing CAT12 and VBM8 for Detecting Brain Morphological Abnormalities in Temporal Lobe Epilepsy. *Frontiers in neurology*, 8:428, 2017.

[108] John V. Hindle. Ageing, neurodegeneration and Parkinson's disease, jan 2010.

[109] Bart Post, Dino Muslimovic, Nan van Geloven, Johannes D. Speelman, Ben Schmand, and Rob J. de Haan. Progression and prognostic factors of motor impairment, disability and quality of life in newly diagnosed Parkinson's disease. *Movement Disorders*, 26(3):449–456, feb 2011.

[110] Pierre Baldi. Autoencoders, Unsupervised Learning, and Deep Architectures. *ICML Unsupervised and Transfer Learning*, pages 37–50, 2012.

[111] Geoffrey E Hinton and Richard S Zemel. Autoencoders, Minimum Description Length and Helmholtz Free Energy. In *NIPS'93: Proceedings of the 6th International Conference on Neural Information Processing Systems*, pages 3–10, 1993.

[112] Derek K. Jones and Mara Cercignani. Twenty-five pitfalls in the analysis of diffusion MRI data, aug 2010.

[113] Konstantinos Kamnitsas, Christian Ledig, Virginia F.J. Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, 36:61–78, 2017.

[114] Tanya Nair, Doina Precup, Douglas L Arnold, and Tal Arbel. Exploring uncertainty measures in deep networks for Multiple sclerosis lesion detection and segmentation. *Medical Image Analysis*, 59:101557, 2020.

[115] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. volume 37 of *Proceedings of Machine Learning Research*, pages 448–456. PMLR, 2015.

[116] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR 2015*, 2014.

[117] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations, ICLR 2014*. International Conference on Learning Representations, ICLR, 2014.

[118] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[119] Rembrandt Bakker, Paul Tiesinga, and Rolf Kötter. The Scalable Brain Atlas: Instant Web-Based Access to Public Brain Atlases and Related Content. *Neuroinformatics*, 13(3):353–366, jul 2015.

[120] Yiming Xiao, Vladimir Fonov, Silvain Bériault, Fahd Al Subaie, M. Mallar Chakravarty, Abbas F. Sadikot, G. Bruce Pike, and D. Louis Collins. Multi-contrast unbiased MRI atlas of a Parkinson's disease population. *International Journal of Computer Assisted Radiology and Surgery*, 10(3):329–341, 2015.

[121] Susumu Mori, Kenichi Oishi, Hangyi Jiang, Li Jiang, Xin Li, Kazi Akhter, Kegang Hua, Andreia V. Faria, Asif Mahmood, Roger Woods, Arthur W. Toga, G. Bruce Pike, Pedro Rosa Neto, Alan Evans, Jiangyang Zhang, Hao Huang, Michael I. Miller, Peter van Zijl, and John Mazziotta. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *NeuroImage*, 40(2):570–582, apr 2008.

[122] Veronica Munoz Ramírez, Virgilio Kmetzsch, Florence Forbes, and Michel Dojat. Deep learning models to study the early stages of parkinson's disease. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1534–1537, 2020.

[123] Jean Patrick Baudry, Cathy Maugis, and Bertrand Michel. Slope heuristics: Overview and implementation. *Statistics and Computing*, 22(2):455–470, apr 2012.

[124] Meng Law, Stanley Yang, James S. Babb, Edmond A. Knopp, John G. Golfinos, David Zagzag, and Glyn Johnson. Comparison of cerebral blood volume and vascular permeability from dynamic susceptibility contrast-enhanced perfusion MR imaging with glioma grade. *American Journal of Neuroradiology*, 25(5):746–755, may 2004.

[125] David C Alsop, John A Detre, Xavier Golay, Matthias Günther, Jeroen Hendrikse, Luis Hernandez-Garcia, Hanzhang Lu, Bradley J Macintosh, Laura M Parkes, Marion Smits, Matthias J P Van Osch, Danny J J Wang, Eric C Wong, and Greg Zaharchuk. Recommended Implementation of Arterial Spin-Labeled Perfusion MRI for Clinical Applications: A Consensus of the ISMRM Perfusion Study Group and the European Consortium for ASL in Dementia. *Magnetic Resonance in Medicine*, 73:102–116, 2015.

[126] Ian O Bledsoe, Glenn T Stebbins, Doug Merkitch, and Jennifer G Goldman. White matter abnormalities in the corpus callosum with cognitive impairment in Parkinson disease. *Neurology*, 91(24):2244–2255, 2018.

[127] Kuan H Kho, Quinten H Leijten, and Lucille D A Dorresteijn. A Parkinson's Disease Patient without Corpus Callosum. *Journal of Parkinson's Disease*, 9:441–442, 2019.

[128] Clément Brossard, Olivier Montigon, Fabien Boux, Aurélien Delphin, Thomas Christen, Emmanuel L Barbier, and Benjamin Lemasson. MP3: Medical software for Processing multi-Parametric images Pipelines. working paper or preprint, September 2020.

[129] Mario Lucic, Matthew Faulkner, Andreas Krause, and Dan Feldman. Training Gaussian Mixture Models at Scale via Coresets. *Journal of Machine Learning Research*, 18:1–25, 2017.

[130] Jeremy Tantrum, Alejandro Murua, and Werner Stuetzle. Hierarchical model-based clustering of large datasets through fractionation and refractionation. *Information Systems*, 29(4):315–326, 2004.

Appendix A

Publication: « No structural differences are revealed using VBM in 'de novo' Parkinsonian patients »

This conference paper was accepted at MEDINFO 2019: Health and Wellbeing e-Networks for All. This international conference took place in August,2019 in Lyon, France. The proceedings were published in the Journal of Studies of Health Technology and Informatics.

The scope of this paper includes the VBM study presented in Chapter 4.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190225

No Structural Differences Are Revealed by VBM in 'De Novo' Parkinsonian Patients

Verónica Muñoz Ramírez[a,b], Florence Forbes[b], Pierrick Coupé[c], Michel Dojat[a]

[a] *Univ. Grenoble Alpes, Inserm U1216, CHU Grenoble Alpes, Grenoble Institut Neurosciences, 38000 Grenoble, France*
[b] *Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK Laboratoire Jean Kuntzmann, 38000 Grenoble, France*
[c] *Laboratoire Bordelais de Recherche en Informatique, UMR 5800, Univ. Bordeaux, Talence, France*

Abstract

The identification of brain morphological alterations in newly diagnosed PD patients (i.e. 'de novo') could potentially serve as a biomarker and accelerate diagnosis. However, presently no consensus exists in the literature possibly due to several factors: small size cohorts, differences in segmentation techniques or bad control of false positive rates. In this study, we use the CAT12 pipeline, to seek for morphological brain differences in gray and white matter of 66 controls and 144 de novo PD patients from the PPMI database. Moreover, we search for subcortical structure differences using the VolBrain pipeline. We found no structural brain differences in this de novo Parkinsonian population, neither in tissues using a whole brain analysis nor in any of nine subcortical structures analyzed separately. We conclude that some results published in the literature may appear as false positives and we contest their reproductibility.

Keywords:

Biomarkers, Magnetic Resonance Imaging, Brain

Introduction

Parkinson's Disease (PD) is a complex neurodegenerative disorder that affects more than 10 million people worldwide [1]. It is mainly characterized by the depletion of dopaminergic neurons situated in the substantia nigra that consequentially disturbs the functions of subcortical nuclei and triggers cortical neuropathological changes causing a plethora of heavily disabling motor and non-motor symptoms [2].

In general, the diagnosis of PD takes place after the manifestation of motor symptoms, which have been found to occur once 50 % of nigrostriatal neurons are lost and dopamine levels are dropped by 80 % [2], [3], creating an urgent need to detect PD biomarkers at the earliest pre-clinical stages of illness possible [4].

The study of morphological brain differences between pathological and healthy groups could potentially identify key regions affected during the PD prodromal phase to better understand PD pathophysiology and its treatment. Magnetic Resonance Imaging (MRI) has positioned itself as a valuable tool for the non-invasive study of the brain's structure. Many automated non operator-dependent techniques have been developed for the analysis of structural MRI data. Voxel-based morphometry (VBM) is the most popular, it allows the detection of subtle morphometric group differences at voxel level [5].

In order to elucidate the nature of morphological differences in de novo PD patients, we investigated 210 subjects from the PPMI (Parkinson Progressive Markers Initiative) through both 1) the well-established Computational Anatomy Toolbox (CAT12) (University of Jena) via the current version of the Statistical Parametric Mapping (SPM12) software and 2) via a new online platform: volBrain [6]. Both pipelines have complementary strengths that are exploited in this study: volBrain performs state of the art quality segmentation of subcortical nuclei [7] and CAT12 facilitates group analysis. Furthermore, we looked for quantitative differences between the tissue classification performed by the two approaches, both including partial volume estimation.

Methods

It is well-known that gathering large cohorts of subjects is a time and resource-consuming task. This is why several efforts have been made by the community to generate databases that benefit for more than one research group. The PPMI (Parkinson Progression Markers Initiative) project is a longitudinal study that gathers data from 35 centers that follows PD patients for five years. The database is openly available for researchers and contains, among other clinical test results, structural MRI images at baseline for 412 patients and 182 healthy subjects.

The scans being heterogeneous, we chose to pool data acquired with the same acquisition parameters, notably magnetic field and scanner manufacturer, to eliminate any additional sources of bias. As a result, our study included 144 de novo PD patients (age: 61.30 ± 9.06; sex: 53 F, 91 M) and 66 healthy controls (age: 60.12 ± 11.39; sex: 23 F, 43 M) from the PPMI database. The structural T1-weighted MRI images extracted were acquired with a 3T Siemens Trio Tim scanner with repetition time (TR) = 2300 ms; echo time (TE) = 2.98 ms; flip angle = 9 degrees; field of view (FOV) = 240 x 256 mm; matrix size : 240 x 256; thickness = 1mm. We note that although T2-weighted images are generally prefered to the delineation of brain structures in neurodegenerative diseases, the available scans on PPMI are provided with low-resolution and thus barely suitable for VBM studies.

Using the CAT12 pipeline

Imaging data were first analyzed using the CAT12 toolbox included in SPM12. All 3D T1-w MRI scans follow a pre-processing protocol including intensity normalization, bias and noise-correction with the Spatially Adaptive Non-Local Means (SANLM) filter introduced in [8] that removes spatially varying noise while maintaining edges. Then the images were spatially normalized using an affine and non-linear (DARTEL and Geodesic Shooting) registration to a reference template brain. Tissue segmentation served to classify the MRI scans into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) components. CAT12 integrates a classical Markov Random Field and the Adaptive Maximum Posterior (AMAP) technique that reduces the dependency on Tissue Probability Maps (TPM). In addition, the segmentation approach uses a Partial Volume Estimation (PVE), taking the three pure tissue classes

as input and estimating two additional mixed classes: GM-WM and GM-CSF. This allows for more precise segmentation as single voxels are likely to contain more than one tissue type. Next, the total intracranial volumes (TIV) were estimated for each subject and the segmented images were modulated by scaling with the amount of volume changes due to non-linear spatial registration, so that the total amount of grey matter in the modulated image remains the same as it would be in the original image.

The resulting images, appearing in Figure 1, were smoothed with an isotropic Gaussian kernel (8mm), and ready for statistical analysis.

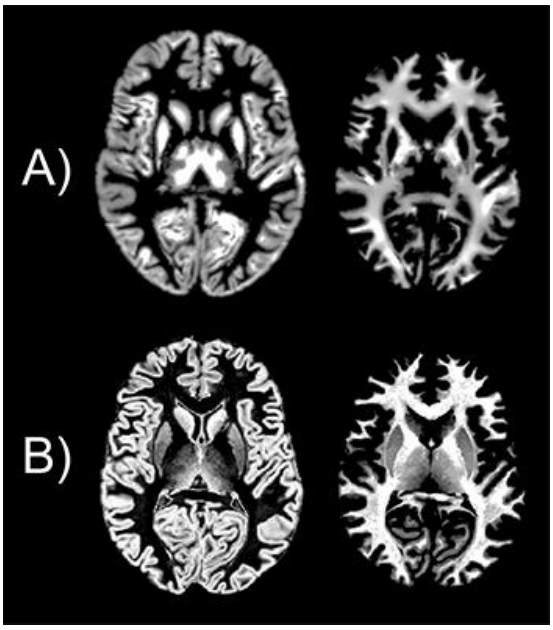


Figure 1– A) CAT12 GM and WM segmentations (modulated) B) volBrain GM and WM segmentations (raw).

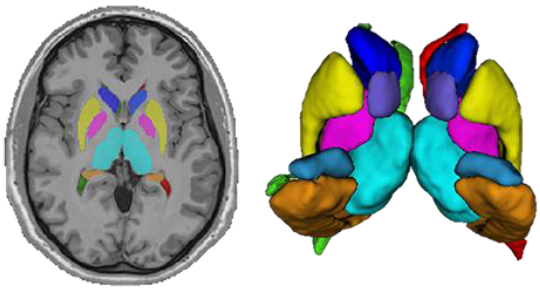


Figure 2– Segmented structures using volBrain.

Using the volBrain pipeline

In parallel, imaging data were analyzed via the volBrain online platform. This system not only provides a state-of-the-art segmentation of the brain tissues (WM, GM, CSF and TIV) (Figure 1), but also segments brain regions like the cerebrum, cerebellum and brainstem; the ventricules; and GM structures such as the putamen, the caudate, the globus pallidus, the thalamus, the hipocampus, the amygdala, and the accumbens [9], as shown in Figure 2.

The multi-template method employed to segment the above mentioned structures considers non-local label fusion schemes using a library built from the manual segmentation of 50 subjects.

The segmentation performed by volBrain provides results in the native and MNI space along with a report containing normality bounds corresponding to the age and sex of the considered subject. These bounds were estimated from the IXI dataset containing 600 normal subjects covering most of adult lifespan.

The pipeline starts by some pre-processing steps. The image is denoised using a SANLM filter, goes through a rough inhomogeneity correction using the N4 method, is registered to the MNI space with a linear affine transformation, goes though a fine SPM based inhomogeneity correction and intensity normalization. Then, segmentation takes place. Tissue classification is obtained by the TMS method that robustly estimates the mean values of the different tissues by excluding partial volume voxels from the estimation jointly with the use of an unbiased robust mean estimator. Partial Volume Coefficients (PVC) are computed from the mean values and completely leave aside tissue probability maps. Next, GM and WM are divided into cerebrum, cerebellum and brainsteam, discriminating between the two hemispheres; and last, subcortical structure segmentation is performed.

VolBrain results analysis by CAT12

Since some subcortical structures of the brain are impacted by PD, we decided to do VBM analysis for the regions provided by volBrain. To do this, we brought volBrain output images to the template space of CAT12 by applying the forward deformation DARTEL field. Once in the same space, the segmented images were used as input for the subsequent statistical analysis. For tissue segmentation analysis (GM & WM), corresponding volBrain's PVC maps were, similarly to CAT12's PVE maps, spatially smoothed with a 8mm kernel.

Statistical analysis

We chose to employ a two-sample T-test to compare the CAT12 modulated tissue maps (GM and WM) of patients versus controls with a general linear model (GLM) where age, sex, and TIV were entered as covariates. The same test was effectuated on volBrain's PVC maps.

A recent study investigating the high rate of false positive present in VBM studies recommends the use of the same group size to detect morphological differences between two groups [10]. Following this recommendation, we repeated our analysis five times to compare the tissue maps of 66 controls versus 66 randomly selected patients using sampling with replacement technique. Their age and sex characteristics are summarized on Table 1.

Table 1– Characteristics of the original study population and the 5 sub-samples of patients equal in size to the control group

	Age	**Sex**
Controls	60.1 ± 11.4	43 M, 23F
Patients	61.3 ± 9.1	91 M, 53F
PD sample 1	61.0 ± 8.7	40 M, 26F
PD sample 2	60.6 ± 9.7	41 M, 25F
PD sample 3	61.7 ± 9.5	44 M, 22F
PD sample 4	61.9 ± 8.7	38 M, 28F
PD sample 5	59.6 ± 8.6	38 M, 28F

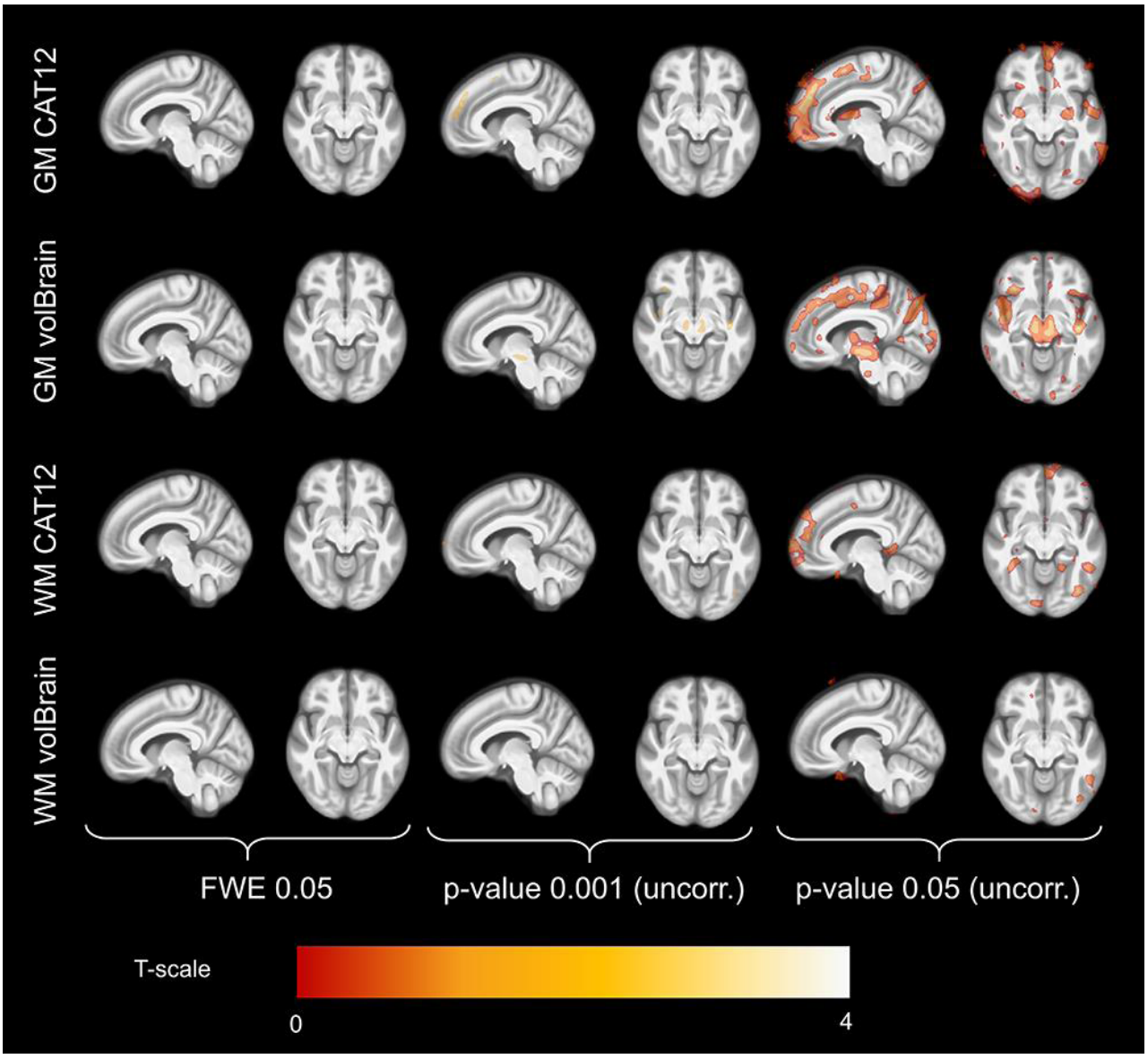


Figure 3– Comparison of PD patients vs controls. Clusters detected for GM an WM diminution in patients using CAT12 and volBrain for different statistical thresholds. The selected slices in the template's MNI space are x=99 and z=64.

Results

When choosing a p-value of 0.05 with Family Wise Error (FWE) correction for multiple comparisons, no voxels survive the difference analysis between PD patient and control groups with tissue map computed with CAT12 or volBrain. In order to replicate some literature results (exploratory study), we decreased the statistical threshold to $p<0.001$ and $p<0.05$ and refrained from any type of correction. Several clusters were then found in PD patients showing volume decrease both in GM and WM as seen in Figure 3.

Also, two-sample T-test comparisons of each independent subcortical structure (computed by volBrain) failed to detect any differences in GM and WM contents $p<0.05$ while FWE corrected. Differences were found in the caudate nucleus, the hipocampus and the putamen for an uncorrected p-value of 0.001.

"Small volume" analysis in SPM12 was used as well to study possible morphometric changes in the substantia nigra, key structure in PD research, using volBrain maps. We observed that differences were only present in gray matter for an uncorrected p-value of 0.001 and did not survived multiple comparison correction.

For all of the 5 new equal size sub-populations (see Table 1), no differences were found in GM or WM for $p< 0.05$ FWE corrected, whilst several significant clusters appeared for an uncorrected p-value of 0.001, esspecially in the frontal cortex for gray matter.

Discussion

Using two recent approaches for accurate segmentation of tissues (CAT12 and volBrain) and subcortical structures (volBrain) we failed to detect robust structural differences in de novo PD patients and healthy controls. We took special care to consider a relatively large cohort of subjects, consider the

effects of an unbalanced number of patients and controls and correct for multiple comparisons. We controlled for multiple comparison using FWE approach, which is known nevertheless to produce some false positives [11]. Following these precautions, no morphological differences were found in PD patients, neither on whole brain GM and WM group analysis or on the analysis of several subcortical structures separately.

In the literature, several studies have reported structural brain differences in PD patients compared to controls. However, these findings tend to be contradictory. In studying a different PD population than in our study, Summerfield and colleagues detected gray matter loss on the right hipocampus, the left anterior cingulate region and the left superior temporal gyrus (p=0.001 uncorrected) in PD patients (n= 13) compared to controls (n=13) [12]. Nyberg and colleagues found an augmentation in the volume of the hipocampus (p=0.03 uncorrected) of PD patients (n=21) and shape deformations of the right accumbens nucleus (p=0.005 uncorrected) compared to controls (n=20) [13]. Radziunas and colleagues observed that PD patients (n=28) with sleep disturbances had bigger ventricules and smaller hipocampus (p-FDR<0.05) than healthy controls (n=28) [14].

Similarly to our study, some VBM studies used the PPMI database and reported structural differences in PD patients. Jia and colleagues noted gray matter losses (p-FWE<0.001) in the fronto-parietal areas and the caudate nucleus, as well as an increase in the size of the limbic and paralimbic areas, the globus pallidus and the putamen of PD patients (n=89) [15] versus controls (n=55) using SPM8.

This lack of consensus on the morphological differences present in de novo PD patients may be due to a variety of factors.

Some studies were carried out on small cohorts, no more than 60 subjects in total, so one may argue that the inconsistencies could be resolved with a larger cohort more representative of the population.

Although, in [10], it was brought to light that sample size does not appear to influence false positive rate, a small sample may incorrectly represent a pathological population, hindering the reproductibility of results.

We note that there is a wide variety of softwares for pre-processing MRI images (i.e. SPM, Freesurfer, FSL), all using different techniques that will inevitably influence the final statistical results as proven by [16] on the study of Multiple Sclerosis. By combining the latest improvements on VBM analysis present in CAT12 (notably denoising and partial volume estimation) with the state of the art segmentations of volBrain [7] we sought to reduce estimation bias considerably.

Finally, correction for multiple comparison is vital to reduce the false positive rate, even if it is not perfect to hope providing robust and reproducible results [11]. Exploratory studies, which use lenient statistical thresholds, could be interesting to indicate some trends in the observed population, that should be confirmed by more robust studies. Then, in our exploratory action (p<0.001 uncorrected) we were able to reproduce some GM results reported in [17]. In the case of [15], the tests were FWE corrected, but the the study was effectuated on VBM8 while, according to [18], the CAT12 toolbox can contribute to more robust detection compared to VBM8.

Regarding the differences we observed between the tissue classification with CAT12 versus volBrain, raw volBrain's PVC maps seem to better distinguish the presence of gray and white matter in the subcortical nuclei. However, as in this study, no morphological robust differences were found between PD patients and controls, a more in depth investigation would be necessary to pertinently test the performances of both methods of partial volume estimation.

In order to further this research, other morphometric methods should be explored, notably Surface Based Morphometry (SBM) and Deformation Based Morphometry (DBM) [17].

Conclusions

In sight of the lack of morphological differences, we suspect that early PD biomarkers may lie on the physiological properties of the Parkinsonian brain and could be investigated through quantitative MRI techniques.

Finally, we reinforce the message that VBM is a delicate technique involving many parameters that should be handled with care to avoid false positive influencing the final results.

Acknowledgements

VMR is supported by a grant from NeuroCoG IDEX UGA in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02).

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org.

PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Allergan, Avid, Biogen, BioLegend, Bristol Meyers Squibb, Denali, GE Healthcare, Genentech, GSK, Lilly, Merck, MSD, Pfizer, Piramal, Roche, Sanofi, Servier, Takeda, Teva, UCB.

References

[1] Parkinson's Foundation, "Parkinson's Prevalence Project," 2018. [Online]. Available: http://parkinson.org/Understanding-Parkinsons/Causes-and-Statistics/Statistics.

[2] M. Politis, "Neuroimaging in Parkinson disease: from research setting to clinical practice," *Nat. Publ. Gr.*, vol. 10, pp. 708–722, 2014.

[3] T. R. Barber, J. C. Klein, C. E. Mackay, and M. T. M. Hu, "Neuroimaging in pre-motor Parkinson's disease," *NeuroImage Clin.*, vol. 15, pp. 215–227, Jan. 2017.

[4] W. Poewe, "The natural history of Parkinson's disease," *J. Neurol.*, vol. 253, no. SUPPL. 7, pp. 2–6, 2006.

[5] J. Ashburner and K. J. Friston, "Unified segmentation," *Neuroimage*, vol. 26, no. 3, pp. 839–851, Jul. 2005.

[6] J. V. Manjón and P. Coupé, "volBrain: An Online MRI Brain Volumetry System," *Front. Neuroinform.*, vol. 10, p. 30, Jul. 2016.

[7] J. V. Manjón *et al.*, "Automatic thalamus and hippocampus segmentation from MP2RAGE: comparison of publicly available methods and implications for DTI quantification," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 11, pp. 1979–1991, 2016.

[8] J. V. Manjón, P. Coupé, L. Martí-Bonmatí, D. L. Collins, and M. Robles, "Adaptive non-local means denoising of MR images with spatially varying noise

levels," *J. Magn. Reson. Imaging*, vol. 31, no. 1, pp. 192–203, Jan. 2010.

[9] P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *Neuroimage*, vol. 54, no. 2, pp. 940–954, Jan. 2011.

[10] C. Scarpazza, S. Tognin, S. Frisciata, G. Sartori, and A. Mechelli, "False positive rates in Voxel-based Morphometry studies of the human brain: Should we be worried?," *Neurosci. Biobehav. Rev.*, vol. 52, pp. 49–55, May 2015.

[11] A. Eklund, T. E. Nichols, and H. Knutsson, "Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 28, pp. 7900–5, Jul. 2016.

[12] C. Summerfield *et al.*, "Structural brain changes in parkinson disease with dementia: A voxel-based morphometry study," *Arch. Neurol.*, vol. 62, no. 2, pp. 281–285, Feb. 2005.

[13] E. M. Nyberg *et al.*, "Morphologic changes in the mesolimbic pathway in Parkinson's disease motor subtypes.," *Parkinsonism Relat. Disord.*, vol. 21, no. 5, pp. 536–40, May 2015.

[14] A. Radziunas *et al.*, "Brain MRI morphometric analysis in Parkinson's disease patients with sleep disturbances," *BMC Neurol.*, vol. 18, no. 1, p. 88, Dec. 2018.

[15] X. Jia, P. Liang, Y. Li, L. Shi, D. Wang, and K. Li, "Longitudinal Study of Gray Matter Changes in Parkinson Disease.," *AJNR. Am. J. Neuroradiol.*, vol. 36, no. 12, pp. 2219–26, Dec. 2015.

[16] A. Versteeg *et al.*, "Grey Matter Atrophy in Multiple Sclerosis: Clinical Interpretation Depends on Choice of Analysis Method," *PLoS One*, vol. 11, no. 1, p. e0143942, 2016.

[17] Y. Zeighami *et al.*, "Network structure of brain atrophy in de novo parkinson's disease," *Elife*, vol. 4, no. September 2015, 2015.

[18] F. Farokhian, I. Beheshti, D. Sone, and H. Matsuda, "Comparing CAT12 and VBM8 for Detecting Brain Morphological Abnormalities in Temporal Lobe Epilepsy," *Front. Neurol.*, vol. 8, p. 428, Aug. 2017.

Address for correspondence

Grenoble Institut des Neurosciences
Bâtiment Edmond J. Safra
Chemin Fortuné Ferrini, Site Santé
38706 La Tronche Cedex
Ph. 04 56 52 05 92
e-mail: veronica.munoz-ramirez@univ-grenoble-alpes.fr

Appendix B

Publication: « Deep learning models to study the early stages of Parkinson Disease »

This conference paper was accepted at ISBI 2020 - IEEE International Symposium on Biomedical Imaging. This international conference took place in April 2020 in Iowa City, USA. The proceedings were published by IEEE.

This paper includes part of the study presented in 5. Notably, the scope of this paper does not include cross-validation or prediction assessment.

DEEP LEARNING MODELS TO STUDY THE EARLY STAGES OF PARKINSON'S DISEASE

Verónica Muñoz Ramírez[⋆†] *Virgilio Kmetzsch*[†⋆] *Florence Forbes*[†] *Michel Dojat*[⋆]

[⋆] Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, GIN, 38000 Grenoble, France
[†] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

ABSTRACT

Current physio-pathological data suggest that Parkinson's Disease (PD) symptoms are related to important alterations in subcortical brain structures. However, structural changes in these small regions remain difficult to detect for neuroradiologists, in particular, at the early stages of the disease (*de novo* PD patients). The absence of a reliable ground truth at the voxel level prevents the application of traditional supervised deep learning techniques. In this work, we consider instead an anomaly detection approach and show that autoencoders (AE) could provide an efficient anomaly scoring to discriminate *de novo* PD patients using quantitative Magnetic Resonance Imaging (MRI) data.

***Index Terms*—** Brain, Anomaly detection, Autoencoder, Diffusion Imaging, MRI

1. INTRODUCTION

Today, there is a pressing need for objective and reliable biomarkers that allow the detection of Parkinson's Disease (PD) from its early stages. MRI has played a vital role in the characterization of multiple neurological diseases like Multiple Sclerosis and brain cancer [1]. However, structural MR images appear to be insufficient to detect the subtle changes caused by PD, especially in the sub-cortical structures of the brain [2]. This motivates the study of PD through quantitative MRI techniques such as DTI (Diffusion Tensor Imaging) that measures the displacement of water molecules in the brain. Indeed, DTI has been useful in the study of Alzheimer's disease [1], making it very attractive for the study of PD.

In this work, we propose to implement an anomaly detection framework to uncover alterations in the diffusion MR images of newly diagnosed (i.e. de novo) patients.

While there exists numerous techniques for anomaly detection [3], deep learning models have achieved remarkable results in recent brain lesion classification challenges such as BRATS, BrainLes and ISLES at MICCAI conferences (see for instance [4]). Nonetheless, these deep learning approaches for anomaly detection are often supervised, meaning that the networks are trained with representatives of both healthy and diseased voxels so as to learn their characteristics. Seeing that the only information available for PD MR scans is an indication of the global presence or absence of the disease, we designed a semi-supervised framework employing three different types of fully convolutional auto-encoders.

More specifically, we trained the auto-encoders to reconstruct healthy diffusion MR scans with a healthy training dataset. Since the network is learned only from healthy subjects, there is no guarantee that it provides good reconstructions outside this population. We therefore compared the scan reconstruction errors of a healthy test data set with those of a pathological data set to identify unusual patterns in the sub-cortical structures of *de novo* PD patients.

We follow the work of [5], which demonstrated the advantages of employing full MR slices as input for the network architectures instead of down sampling and patch division that result in important loss of spatial information.

2. INPUT DATA

We pooled our data from the PPMI (Parkinson Progression Markers Initiative) database. The PPMI is a longitudinal study that follows *de novo* PD patients of 35 centers for five years. The database is openly available for researchers and contains, among other clinical test results, structural and diffusion MR images. To eliminate any additional sources of bias, we only selected scans acquired with the same acquisition parameters, notably magnetic field and scanner manufacturer. As a result, we pooled 129 *de novo* patients (age: 62 ± 9; sex: 48 F, 80 M) and 57 healthy controls (age: 61 ± 10; sex: 23 F, 34 M).

From this data, two features were extracted: mean diffusivity (MD), accounting for the overall water displacement, and fractional anisotropy (FA), an indication of diffusion orientation. Values of FA and MD were normalized into the range $[0, 1]$. Each volume was composed of $116 \times 116 \times 72$ voxels. Forty coronal slices (116×72) were extracted from the center of the brain encompassing the sub-cortical structures under study.

To avoid data leakage, the control dataset was divided into

VMR is supported by a grant from NeuroCoG IDEX UGA in the framework of the "Investissements dávenir" program (ANR-15-IDEX-02).

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database. For up-to-date information on the study, visit www.ppmi-info.org.

a training dataset (42 MR volumes) and a testing dataset (15 MR volumes). Special care was taken to maintain similar age and sex profiles in both datasets. In order to choose the best network architecture and tune the corresponding hyper-parameters, all models were trained and assessed with 7-fold cross-validation, with 36 MR volumes used for training and 6 MR volumes for validation in each fold. Once the final models were defined, they were retrained with the full control training set (1680 slices) and evaluated with the PD dataset (5160 slices) and the healthy control test set (600 slices).

3. ARCHITECTURE DESIGN

Three autoencoder-based models were developed and evaluated: a spatial autoencoder (sAE), a spatial variational autoencoder (sVAE) and a dense variational autoencoder (dVAE).

As seen in Figure 1-A, all autoencoders come with two parts, an encoder f_ϕ and a decoder g_θ. The encoder maps the input images x to a lower dimensional latent representation z, then the decoder maps the latent vector z to the reconstructed output images $\hat{x} \in \mathbb{R}^{H\times W\times C}$:

$$z = f_\phi(x) \qquad\qquad \hat{x} = g_\theta(z) \tag{1}$$

We extended the work of [5] by using multiple quantitative MRI measures simultaneously as input. Thus, every input can be expressed as $x \in \mathbb{R}^{H\times W\times C}$, where H is the height, W the width and C the number of channels (FA and MD).

Depending on the network architecture, the latent code may be a simple vector ($z \in \mathbb{R}^d$) or a third-order tensor ($z \in \mathbb{R}^{h\times w\times c}$). The former is referred as a dense bottleneck and the latter as a spatial bottleneck.

All models were implemented using Python 3.6.8, PyTorch 1.0.1, CUDA 10.0.130 and trained on a NVIDIA GeForce RTX 2080 Ti GPU with batches of 32 images. After each convolutional layer, batch normalization [6] was applied for its regularization properties. The nonlinear activation function in each layer was the rectified linear unit (ReLU), except for the last layer where a sigmoid was employed in order to have output pixels normalized between $[0, 1]$. The loss functions were optimized using Adam [7], a popular optimization algorithm for training deep neural networks.

3.1. Spatial autoencoder (sAE)

The spatial autoencoder model is fully convolutional, 4 convolutional layers go from input to bottleneck and 4 transposed convolutional layers from bottleneck to output. As depicted in Figure 1-B, the output of the encoder network is directly the latent vector z and loss function is simply the reconstruction error:

$$\mathcal{L} = \|x - \hat{x}\|_1 \tag{2}$$

This model was trained for 160 epochs, with a learning rate of 10^{-3}. 5×5 kernels were convolved using padding of 1 pixel and a stride of (2, 2), and there were no pooling layers.

3.2. Spatial variational autoencoder (sVAE)

Our spatial variational autoencoder model is shown in Figure 1-C. Similar to sAE, the model is fully convolutional, however, the encoder generated the parameters of the approximate posterior of the latent variable given the input, constrained to follow a multivariate normal distribution. A sampling operation was needed to obtain an actual value for z.

Training lasted for 200 epochs using a learning rate of 0.3×10^{-3}. A 5×5 kernel was chosen as filter, along with a padding of 1 and stride of (2, 2). No pooling layers were used. The loss function was computed as follows:

$$\mathcal{L} = \lambda\|x - \hat{x}\|_1 + (1-\lambda)\Big[-\tfrac{1}{2}\sum_{j=1}^{J}(1 + \log((\sigma_j)^2) - (\mu_j)^2 - (\sigma_j)^2)\Big] \tag{3}$$

where μ and σ denote the mean and the variance of the approximate posterior, J is the number of dimensions of the latent space and λ controls the proportions between the two terms. The first term is the reconstruction error and the second term is the Kullback-Leibler (KL) divergence between the approximate posterior and the prior of the latent variable, for the Gaussian case [8]. To favor good reconstructions over a Gaussian-like distribution of the latent variables, we put more weight (90%) in the reconstruction term and less weight (10%) in the KL divergence term.

3.3. Dense variational autoencoder (dVAE)

The main difference of the dense variational autoencoder when compared to the sVAE is its dense bottleneck. Encoder and decoder also have fully connected layers in addition to the convolutional layers, as shown in Figure 1-D.

For regularization purposes, the dropout [9] technique was used to turn off 30% of the units in fully-connected layers during training. This model was trained for 100 epochs with a learning rate of 0.3×10^{-3}. There were no pooling layers. Kernels for all convolutional layers were 5×5, and convolutions were performed with a padding of 1 and stride of (2, 2). The dVAE shared the same loss function as the sVAE and we kept the same 90/10 proportion between the reconstruction term and the KL divergence term.

4. PD ANOMALY DETECTION

A reference model is learned from healthy MR images using the autoencoder-based models presented in Section 3. During the training process the model parameters are tuned so as to minimize a loss function that favors good reconstructions. Since the network was solely trained on healthy subjects, we hypothesized that the MR scans of the PD population would

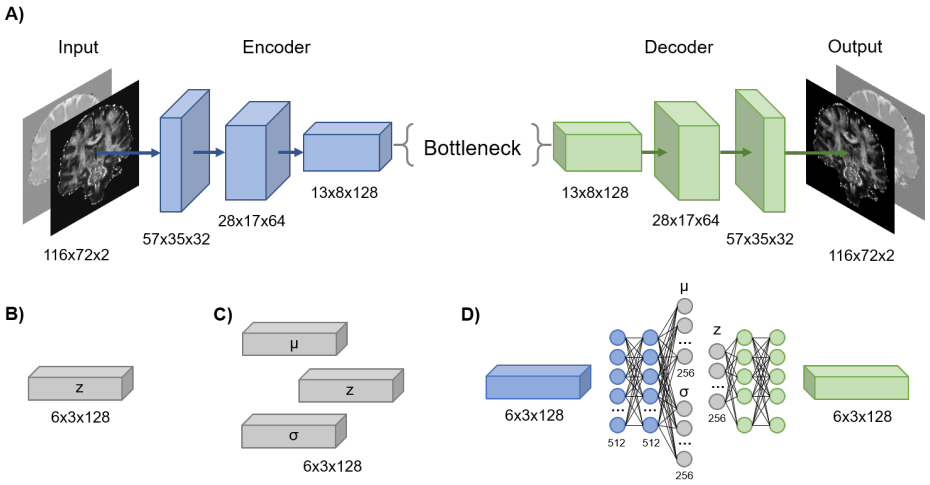


Fig. 1. A) The general architecture of the implemented autoencoders with an unspecified bottleneck; B) sAE spatial bottleneck; C) sVAE spatial bottleneck and D) fourth convolutional layer of the dVAE along with its fully connected layers and dense bottleneck. μ and σ describe the approximate posterior of the latent variable, z is obtained by a sampling operation.

have greater reconstruction errors in some regions. The idea is therefore to use the reconstruction error as an anomaly score.

The voxel-wise reconstruction errors in one image can be computed as $|x_i - \hat{x}_i|$. Seeing that our decoders output two images ($\hat{\text{FA}}, \hat{\text{MD}}$), we defined the joint reconstruction error of every voxel as:

$$\sqrt{(\text{FA}_i - \hat{\text{FA}}_i)^2 + (\text{MD}_i - \hat{\text{MD}}_i)^2} \tag{4}$$

We identified four sources of reconstruction errors : 1) noise in the input data, 2) loss of information due to dimension reduction in the latent space, 3) variability of healthy controls not captured by the model and 4) finally real anomalies caused by PD. Because we were only interested in the latter, the best way to evaluate and compare the models is by measuring their ability to discriminate between controls and PD patients, based on the intensity and localization of the reconstruction errors.

To help evaluate the localization of anomalies, we employed the MNI PD25 atlas [10], specifically designed for PD patient exploration. It contains 8 regions: substantia nigra (SN), red nucleus (RN), subthalamic nucleus (STN), globus pallidus interna and externa (GPi, GPe), thalamus, putamen and caudate nucleus. In addition, we considered the superior colliculus (SC) and the inferior colliculus (IC), where we recently found functional deficits [11].

Regarding comparisons of reconstruction error intensities, we investigated extreme reconstruction errors with the idea that PD patients should exhibit very abnormal voxels in larger quantities. Accordingly, we considered an extreme quantile (eg. the 95% quantile) of the distribution of errors in the control population as possible threshold value to decide whether or not a given voxel was considered as abnormal. For each control or PD subject, we counted the number of extreme abnormalities detected in every structure. The idea being to classify a subject as PD or healthy when this number was above a certain value. The critical choice of this value was investigated using a ROC curve of sensitivity and specificity to account for class imbalance. We chose the AUC as our principal indicator of discrimination performance.

5. RESULTS: RECONSTRUCTION ASSESSMENT

We used the reconstruction errors of every structure in every subject to classify patients from controls as explained in Section 4. Amongst the three models tested, we noticed that dVAE detected the biggest reconstruction errors, as an example the mean reconstruction error of the control's subcortical structures is of 0.075, 0.086 and 0.106 for the sAE, sVAE and dVAE respectively. However, the absence of ground truth at a voxel level prevented us from determining which model was the most accurate relative to PD abnormalities.

The number of voxels over the 95% quantile abnormality threshold were generally superior in a patient than in a healthy control. We employed the ROC curve as a performance measurement for our classification problem, we were able to measure the AUC to have an indication on the ability of the model to distinguish between patients and healthy controls. The structures with the highest AUC were the Substantia Nigra, the Red Nucleus, the Thalamus and the combination of all subcortical structures as seen in Table 1. This was in accordance with the literature. The total white matter present in the reconstructed slices obtained even better results with an AUC of 0.83, 0.80 and 0.74 for the sAE, sVAE and dVAE respectively.

	sAE	sVAE	dVAE
Red Nucleus	0.75	0.76	0.65
Substantia Nigra	0.74	0.73	0.72
Sub-thalamic Nucleus	0.59	0.64	0.53
Caudates	0.70	0.64	0.61
Putamen	0.72	0.74	0.63
Globus Pallidus ext.	0.65	0.69	0.69
Globus Pallidus int.	0.69	0.69	0.71
Thalamus	0.71	0.73	0.72
Superior Colliculus	0.51	0.59	0.54
Inferior Colliculus	0.54	0.56	0.49
All subcortical structures	0.76	0.77	0.73
White Matter	0.83	0.80	0.74

Table 1. AUC values obtained in different structures when counting the number of voxels above the 95% percentile to discriminate between patients and controls

When assessing the reconstructions of FA and MD separately, FA obtained the highest ROC AUC for the caudates and the putamen, whereas for the white matter and the red nucleus MD was better. The substantia nigra and the ensemble of all subcortical structures benefited from the joint measures of FA and MD.

6. DISCUSSION AND CONCLUSION

Although preliminary, these results offer compelling evidence that deep learning-based models are useful to identify subtle anomalies in *de novo* PD patients, even when trained with a moderate number of images and only two parametric maps as input. The good discriminative performances of the sub-cortical structures are in accordance to our physiopathological knowledge of PD. The dopaminergic neuron deficit in Substantia Nigra is known as critical in the development of PD. What is more, the absence of motor symptoms in the pre-clinical stages of PD may be the result of compensation mechanisms involving the structures in the motor coordination pathways of the brain such as the Thalamus and the Red Nucleus [12].

We have shown that no structural changes, including in the White Matter (WM), can be robustly observed from T1-weighted images to automatically distinguish between controls and *de novo* PD patients [2]. To explain the good performances of autoencoders based on white matter voxels, we may hypothesize that 1) diffusion parameters (FA an MD) are more informative than grey levels from T1-weighted imaging, or that 2) they dispose of more voxels, compared to subcortical structures, to build a model that captures the variability of healthy controls.

Undeniably, experiments on a larger cohort are necessary to confirm our results. The control group available for our experiments contained only 57 MRI volumes, with gender and age imbalance. Although we were able to discriminate between healthy controls and individuals affected by PD with good performances, we cannot rule out that other possible causes of variability in brain properties, such as age and gender, and other hidden parameters, might have influenced our classification performance.

Similar anomaly detection frameworks could be of interest for studying other neurological disorders where small lesions are suspected and difficult to localize for a human observer. In future work, it could be beneficial to include other quantitative MR measures, such as perfusion and relaxation times, to encode a more complete picture of the pathophysiology of the disease.

7. REFERENCES

[1] P. Tofts, *Quantitative MRI of the brain*, Wiley, 1st edition, 2005.

[2] V. Munoz Ramirez & al., "No structural differences are revealed by vbm in de novo parkinsonian patients," in *MEDINFO*, 2019.

[3] M. Pimentel & al., "Review: A review of novelty detection," *Signal Process.*, vol. 99, pp. 215–249, June 2014.

[4] "MICCAI 2018 Proceedings," in *MICCAI*, Granada, 2018, Springer, Cham.

[5] C. Baur & al., "Deep autoencoding models for unsupervised anomaly segmentation in brain mr images," in *Brainlesion*. 2019, pp. 161–169, Springer Intl.

[6] S. Ioffe & C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *ICML*, 2015, pp. 448–456.

[7] D. Kingma & J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv e-prints*, Dec 2014.

[8] D. Kingma & M. Welling, "Auto-encoding variational bayes," *CoRR*, 2013.

[9] N. Srivastava & al., "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014.

[10] Y. Xiao & al., "Multi-contrast unbiased MRI atlas of a Parkinson's disease population," *Int. J. Comput. Assist. Radiol. Surg*, vol. 10, no. 3, pp. 329–341, 2015.

[11] E. Bellot & al., "First steps of visual processing are altered in de novo Parkinson's disease patients," in *IAPRD Conf, Lyon, FR*, 2018.

[12] I. Philippens & al., "Involvement of the Red Nucleus in the Compensation of Parkinsonism may Explain why Primates can develop Stable Parkinson 's Disease," *Sci. Rep.*, pp. 1–9, 2019.

Appendix C

Publication: « Quantitative MRI characterization of brain abnormalities in de novo Parkinsonian patients »

This conference paper was accepted at ISBI 2019 - IEEE International Symposium on Biomedical Imaging. This international conference took place in April 2019 in Venice, Italy. The proceedings were published by IEEE.

This paper includes the application of the MMST model pipeline presented in Chapter 6 on the AGIR-Park dataset. Important results differences arise due to the use of different preprocessing pipelines to extract FA and MD diffusion parameters. FSL was employed in this paper and MRtrix in Chapter 6. What is more, an additional patient was included in this manuscript.

QUANTITATIVE MRI CHARACTERIZATION OF BRAIN ABNORMALITIES IN *DE NOVO* PARKINSONIAN PATIENTS

Verónica Muñoz Ramírez[⋆†] *Florence Forbes*[†] *Julyan Arbel*[†] *Alexis Arnaud*[†⋆] *Michel Dojat*[⋆]

[⋆] Grenoble Institut des Neurosciences, Inserm U1216, Univ. Grenoble Alpes, 38000 Grenoble, France
[†] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

ABSTRACT

Currently there is an important delay between the onset of Parkinson's disease and its diagnosis. The detection of changes in physical properties of brain structures may help to detect the disease earlier. In this work, we propose to take advantage of the informative features provided by quantitative MRI to construct statistical models representing healthy brain tissues. This allows us to detect atypical values for these features in the brain of Parkinsonian patients. We introduce mixture models to capture the non-standard shape of the data multivariate distribution. Promising preliminary results demonstrate the potential of our approach in discriminating patients from controls and revealing the subcortical structures the most impacted by the disease.

***Index Terms*—** Brain, Biomarker, Statistical mixture models

1. INTRODUCTION

Parkinson's Disease (PD) is a complex neurodegenerative disease that can be divided into two stages. First, a prodromal phase lasting from a few years to several decades where dopaminergic neurons are lost progressively, affecting mainly subcortical brain structures. Then, a clinical phase involving cortical networks and marked by PD motor symptoms, such as gait impairment, rigidity and resting tremor. It would be of great importance, with respect to research and treatment, to identify biomarkers that may yield early PD diagnosis. Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique useful for the detection, characterization and monitoring of neurological changes in *de novo*, i.e. just diagnosed, PD patients. Several innovative *quantitative* MRI techniques have been developed to measure meaningful physical or chemical brain properties [1]. The combination of measures from multiple MRI modalities may provide a complete picture of functional and structural changes caused by PD [2].

In the quest to find biomarkers that accurately differentiate PD patients from healthy subjects, we propose an abnormality detection approach that identifies unusual patterns in multi-parametric MRI data. For this purpose, we fit a mixture model to data from healthy volunteers (reference model), then compare to *de novo* PD data to search for outliers, i.e. possible abnormalities, in different brain regions, to which we fit another mixture model (atypical model) to represent the potentially complex distribution of abnormal tissues.

Mixture models are probabilistic models that represent subpopulations within an overall population. They do not require prior knowledge of which subpopulation a data point belongs to, allowing the model to learn automatically the subpopulations in an unsupervised manner.

In our study, the reference model is built assuming that healthy subjects present a set of distinct healthy tissues characterized by different quantitative characteristics. These classes of characteristics are captured using a mixture model with an appropriate number of components, whose number is determined automatically from the data. Abnormal regions in PD patients are then detected as those with a too low log-density score in the reference model. The detected abnormal regions are in turn used to build another mixture that defines a new model for atypical characteristics.

2. QUANTITATIVE MRI DATA

Quantitative MR data are considered, in particular, to be as independent as possible of the MRI scanner or the study center [1]. The scanner is no longer considered only as a camera but as a means of measurement. In this paper, we consider Diffusion imaging (DI) and Perfusion Imaging (PI). Through DI, mean diffusivity [MD] describes the overall displacement of water molecules and fractional anisotropy [FA] indicates the orientation of diffusion. Although there are some inconsistent findings regarding diffusivity in PD patients, DI has been helpful to discriminate PD from other syndromes [3, 4]. Moreover, with the development of arterial spin labelling (ASL), brain perfusion can be assessed non-invasively. Cerebral blood flow (CBF) has been reported to be decreased in the cortex of PD patients and either preserved or decreased in their basal nuclei [5, 6].

To explore the feasibility of our method described in Section 3, we extracted FA and MD parameter maps from DI, and CBF maps from PI, for three healthy subjects (C1-C3)

VMR is supported by a grant from NeuroCoG IDEX UGA in the framework of the Investissements dávenir program (ANR-15-IDEX-02).

(Age: $\{28, 40, 50\}$; 2 Males) and nine *de novo* PD patients (P1-P9) (Age $\in [36, 66]$; H&Y score ≤ 2; 7 Males). All images were acquired on a Philips 3T scanner. For each individual, the computed maps were coregistered and resliced (1mm isotropic) with SPM12 to obtain, for each voxel v, three corresponding feature values $y_v = (\text{FA}_v, \text{MD}_v, \text{CBF}_v)$.

In this first attempt, we have searched for outliers in the subcortical brain structures where changes were most likely to take place in *de novo* PD patients, namely the substantia nigra (SN), the red nucleus (RN), the subthalamic nucleus (STN), the Globus Pallidus interna and externa (GPi, GPe), the thalamus, the putamen and the caudate nucleus. To obtain a mask of these Regions of Interest (ROIs), we performed a non-linear deformation of the MNI PD25 atlas [7], specifically designed for PD patient exploration. In addition, we considered the superior colliculus (SC) and the inferior colliculus (IC), where we recently found functional deficits [8]. The results of this experiment are summarized in Section 4.

3. ABNORMALITY DETECTION VIA MIXTURE MODELS

To eliminate possible redundant information, we perform a Principal Component Analysis (PCA) on the three maps (CBF, FA and MD) for our healthy subjects. Each parameter contributes equally to the information of explained variance $(38\%, 33\%, 29\%)$ on axes $(1, 2, 3)$ and all parameters are then considered in the statistical analysis. A following step is then to decide on an appropriate model that best accounts for the data distribution. Although Gaussian distributions are the most widely used statistical distributions for their tractability and representation power, they are constrained by elliptical shapes. Mixtures of Gaussians can help in modelling richer distributional shapes but they are still not appropriate when the data present elongated and strongly non-elliptical subgroups. As an alternative, we consider a richer family of distributions based on multiple scaled t-distribution (MSD) mixtures. These heavy-tailed distributions are endowed with variable marginal amounts of tail-weight and their mixtures have been showed to provide an efficient alternative to Gaussian mixtures [9, 10]. In particular, their ability to model over-dispersed values is illustrated in Figure 1.

The tridimensional measures (FA, MD, CBF) are separated into healthy and patient datasets,

$$Y_H = \{y_v, v \in V_H\} \quad \text{and} \quad Y_P = \{y_v, v \in V_P\}, \qquad (1)$$

where V_H and V_P respectively represent the voxels belonging to healthy volunteers and to patients. By fitting an MSD mixture to the healthy data Y_H, we build a reference model density f_H that describes healthy tissues as follows:

$$f_H(y|\boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{k=1}^{K_H} \pi_k \text{MSD}(y; \theta_k), \qquad (2)$$

where K_H is the number of mixture components, and each component is characterized by a proportion π_k and an MSD parameter θ_k. The optimal number of components K_H is chosen using the Bayesian Information Criterium (BIC). The Expectation-Maximization algorithm (EM) [9] is used to obtain the best fit of the MSD mixture on Y_H: it jointly estimates the weights $\boldsymbol{\pi}$ and the parameters $\boldsymbol{\theta}$ by reaching a (local) maximum of the model log-likelihood.

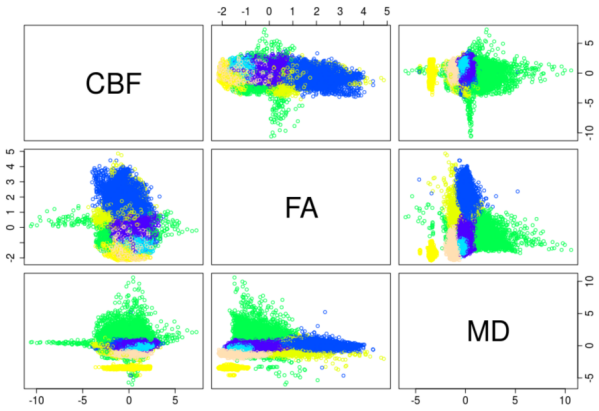


Fig. 1. Bivariate marginal distributions of our three parameter maps (CBF, FA, MD). The colors represent the different classes assigned by the MSD mixture reference model (2).

The log-density (or log-score) of each voxel (in V_H and V_P) with respect to f_H, $\{\log f_H(y_v), v \in V_H \cup V_P\}$, can then be used to compute a threshold to detect abnormal voxels. The log-score can be considered as a measure of proximity of one voxel v (associated to value y_v) to the reference healthy model (represented by f_H). The rational is that voxels from healthy subjects are more likely to have a high log-score while pathological voxels may not be well explained by the reference model and consequently have a lower log-score. A threshold can be computed in a data-driven way by fitting a 2-component mixture model on the log-scores, defining the threshold at the intersection of the two distributions. This separation into high and low score groups would correspond to a certain false positive error rate (FPR) that can be determined as proposed in [10]. We choose another option which consists of deciding on an acceptable FPR α and determining the corresponding threshold τ_α: the probability the log-score is smaller than τ_α, although the log-score is that of an healthy voxel, is α. In other words, τ_α is the value such that

$$P(\log(f_H(Y)) < \tau_\alpha) = \alpha,$$

when Y is a random variable following the f_h reference model distribution. In practice, while f_H is known explicitly, the probability distribution of $\log(f_H(Y))$ is not. However, it is easy to simulate this distribution so that τ_α can be computed using empirical quantiles (Figure 2).

All the voxels whose log-score is below τ_α are then labeled as abnormal and the corresponding measures provide a

set of parameters that are referred to as the abnormal data set

$$Y_A = \{y_v, v \in V_H \cup V_P,\ s.t.\ \log f_H(y_v) < \tau_\alpha\}\ .$$

An abnormality model denoted by f_A is then constructed following the same procedure as for the reference model:

$$f_A(y|\boldsymbol{\mu}, \boldsymbol{\phi}) = \sum_{k=1}^{K_A} \mu_k \text{MSD}(y; \phi_k). \quad (3)$$

This abnormality model is used to account for the fact that voxels detected as abnormal may belong to different abnormality classes, with different physiological characteristics. Typically, the above formula indicates that among the Y_A set there are K_A different groups. Theses groups can be used to build a signature ρ^s of each subject s by determining the proportion of voxels that are assigned to each of the K_A classes:

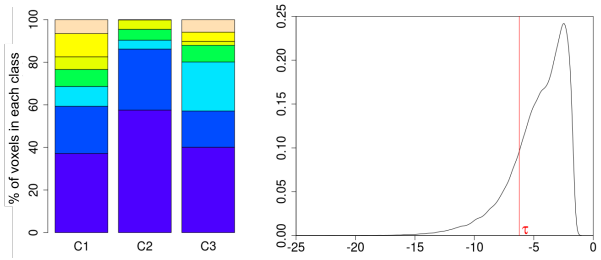
$$\rho^s = (p^s_1, \ldots, p^s_{K_A}). \quad (4)$$

Figure 3 provides an example of such signatures. The top plot illustrates that abnormal voxels can be detected in healthy subject although in a much lower proportion that in PD subjects. In addition the abnormality pattern in healthy subjects is usually different and may be distinguished from PD patterns. A model of this difference can be quantified using standard discriminant analysis models and the known status of each subject (control or patient). As explained in [10], we can make use of this classification information and of additional spatial information to refine abnormality detection. Notably, clusters of less than 4 atypical voxels are not selected. Figure 3-bottom shows the effect of such a post-processing.

4. RESULTS

For control data, the highest BIC score for the MSD mixture reference model (2) is obtained for $K_H = 7$ components. It is critical that the reference model correctly characterizes the physiological properties of brain structures while encompassing the individual variability. Interestingly, while individual differences emerge, a spatial symmetry is observed over the two brain hemispheres, when looking at individual brain slices annotated with the reference classification. The proportion of each class in the control subjects provides an insight into these differences, see Figure 2(A).

To separate atypical from healthy voxels we choose a FPR of $\alpha = 5\%$. Figure 2(B) shows the density of the log-scores computed for all voxels with respect to the reference model. Voxels considered as abnormal are located left to the vertical red line representing the threshold τ_α. 49 196 voxels are labelled as abnormal (i.e. 20.4% of the set of voxels). Note that the majority of these abnormal voxels belong to PD patients, however, but some are detected in the controls. This is expected for two reasons. First, because each healthy individual is unique and can bring some non-pathological extreme values, just like for a patient, the physiological measures are not



(A) (B)

Fig. 2. Panel (A): proportion of each of the $K_H = 7$ classes for controls C1, C2 and C3 given by the reference model f_H. Panel (B): density of the log-scores computed for all voxels (controls and patients) with respect to f_H. The red line represents the abnormality threshold τ_α.

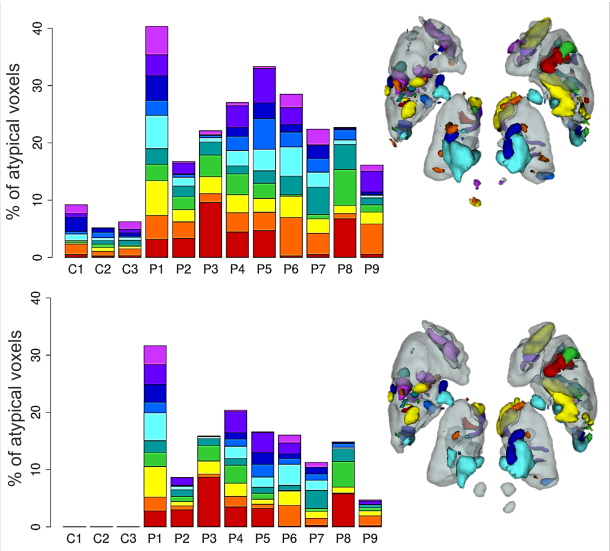


Fig. 3. Subjects signatures derived from f_A before (top) and after post-processing (bottom). Each color represents one the $K_A = 10$ abnormal classes. Their localization in the brain is illustrated for PD patient P1 (right).

all abnormal; and second, because we have admitted a non-zero FPR.

The abnormal model is best defined by $K_A = 10$ classes according to BIC score. Figure 3 displays the abnormal signatures of all subjects before and after post-processing. After post-processing only 28 377 voxels, i.e 11.8% of all voxels, remain labelled as abnormal. For Controls, no more voxels are detected as abnormal. The largest numbers of abnormal voxels are detected for patients P1 and P4. This is coherent with the corresponding additional functional MRI data obtained for these two subjects, which show alteration in visual information processing in their subcortical structures[8].

Table 1 indicates the localization for each PD patient of

abnormal voxels with regard to the corresponding subcortical structures. We computed the percentage of each ROI that was detected as abnormal. In accordance with the literature, with found that the substantia nigra (SN), the red nucleus (RN) and the globus pallidus (GP) were the most impacted by the pathology. Indeed,the degeneration of dopaminergic neurons in the SN pars compacta and the subsequent denervation of the dorsal striatum are at the origin of PD [11]. Abnormal signatures found in SN, GP and Cau are also in line with the reduced degree of myelination found in these regions for PD patients [12, 5]. Moreover, SN and RN show an augmented FA [3] and there is evidence of a decrease in structural connectivity between SN, ipsilateral putamen and thalamus [13];

%	P1	P2	P3	P4	P5	P6	P7	P8	P9
SN	71	19	18	64	10	49	60	29	7
RN	48	1	0	7	41	31	27	24	0
STN	22	2	0	45	0	11	12	0	0
GPe	49	4	9	24	29	21	19	5	10
GPi	63	4	1	21	11	39	24	0	22
Th	20	10	32	11	12	16	5	32	3
Put	30	11	7	25	20	5	7	4	3
Cau	27	1	4	14	28	13	15	5	0
SC	0	0	13	0	0	22	16	0	0
IC	33	0	0	0	0	18	58	13	0

Table 1. Percentages of abnormal voxels after post-processing in each patient subcortical structure.

5. CONCLUSION AND PERSPECTIVES

We presented a method for the detection of abnormalities in *de novo* PD patients. We demonstrated its potential on a small cohort of three controls and nine PD patients. Clearly, more subjects are needed to bring out robust biomarkers of PD. However, these preliminary results show that the application of mixture models of relevant distributions is informative and promising to correctly discriminate the pathology. In addition, our method, in contrast to supervised machine learning techniques, does not rely on a large set of annotated data, which are difficult to obtain in medical contexts. Moreover, it relies on interpretable statistical tools that can be tuned and compared, providing a model of physical properties alterations. Last but not least, it does not require ground-truth comparison, making it a valuable tool for the exploration of any physiological changes. Future work will extend our method to the study of the entire brain on large cohorts and on the inclusion of new physiological measures to fully exploit the potential of multi-parametric quantitative MRI.

6. REFERENCES

[1] P. Tofts, *Quantitative MRI of the brain*, Wiley, 1st edition, 2005.

[2] F. D. B. Bowman & al., "Multimodal imaging signatures of Parkinson's disease," *Frontiers in Neuroscience*, vol. 10, pp. 1–11, 2016.

[3] S. T. Schwarz & al., "Diffusion tensor imaging of nigral degeneration in Parkinson's disease: A region-of-interest and voxel-based study at 3 T and systematic review with meta-analysis," *NeuroImage: Clinical*, vol. 3, pp. 481–488, 2013.

[4] G. Du & al., "Combined R2* and Diffusion Tensor Imaging Changes in the Substantia Nigra in Parkinson's Disease," *Movement Disorders*, vol. 26, no. 9, pp. 1627–32, 2011.

[5] N. Pyatigorskaya & al., "A review of the use of magnetic resonance imaging in Parkinson's disease," *Therapeutic Advances in Neurological Disorders*, vol. 7, no. 4, pp. 206–220, 2014.

[6] T. R. Barber & al., "Neuroimaging in pre-motor Parkinson's disease," *NeuroImage: Clinical*, vol. 15, pp. 215–227, jan 2017.

[7] Y. Xiao & al., "Multi-contrast unbiased MRI atlas of a Parkinson's disease population," *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, no. 3, pp. 329–341, 2015.

[8] E. Bellot & al., "First steps of visual processing are altered in de novo Parkinson's disease patients," in *IAPRD Conf, Lyon, FR*, 2018.

[9] F. Forbes and D. Wraith, "A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: application to robust clustering," *Statistics and Computing*, vol. 24, no. 6, pp. 971–984, 2014.

[10] A. Arnaud & al., "Fully Automatic Lesion Localization and Characterization: Application to Brain Tumors Using Multiparametric Quantitative MRI Data," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1678–1689, 2018.

[11] M. Politis, "Neuroimaging in Parkinson disease: from research setting to clinical practice," *Nature*, vol. 10, pp. 708–722, 2014.

[12] P. Péran & al., "Magnetic resonance imaging markers of Parkinson's disease nigrostriatal signature," *Brain*, vol. 133, no. 11, pp. 3423–33, 2010.

[13] L. Griffanti & al., "Exploring variability in basal ganglia connectivity with functional MRI in healthy aging," *Brain Imaging and Behavior*, pp. 1–6, 2018.

Anomaly characterization in the MRI data of ‘de novo’ Parkinson’s patients

Parkinson's disease (PD) is a progressive disorder of the nervous system characterized by the degeneration of dopaminergic neurons found in the substantia nigra. In general, the diagnosis occurs once the patients start experiencing the well-known motor symptoms of this disease, namely stiffness, akinesia and resting tremor. By this moment, it is estimated that 60 to 80% of the dopamine-producing neurons have already been lost or impaired.
However, as the loss of these neurons disrupts the functioning of the subcortical structures, many non-motor symptoms such as visual and olfactory disturbances, mood disorders or changes in cerebral perfusion occur earlier in the pathology.
In this context, our project aims to find specific signatures in the newly diagnosed Parkinson's patients that can lead to earlier diagnosis and the classification of patients into sub-types for more tailored treatments to slow down the disease process.
To reach this objective, we have explored three different approaches. Firstly, we searched for morphometric changes in structural data using popular techniques such as VBM, DBM and SBM. Secondly, we conceived a deep-learning methodology to detect brain diffusion anomalies employing reconstruction errors of a trained auto-encoder. Finally, we developed an original statistical approach based on mixture models generated from Student distribution laws to construct a reference control model form multi-modal quantitative data of “normality” and to classify the abnormalities present in our patients.

Caractérisation d’anomalies sur les données IRM des patients Parkinsoniens ‘de novo’

La maladie de Parkinson est un trouble progressif du système nerveux caractérisé par la dégénérescence des neurones dopaminergiques trouvés dans la substantia nigra. En général, le diagnostic est posé par le neurologue lorsque les patients commencent à ressentir les symptômes moteurs bien connus de cette maladie, à savoir la raideur, l'akinésie et les tremblements au repos. À ce stade, on estime que 60 à 80% des neurones productrices de dopamine ont déjà été perdues ou altérées.
Cependant, la perte de ces neurones perturbant le fonctionnement des autres structures sous-corticales, de nombreux symptômes non moteurs (e.g. troubles olfactifs, visuels, modifications de la perfusion cérébrale) peuvent apparaitre à un stade plus précoce de la pathologie.
Dans ce contexte, notre projet vise à révéler des anomalies spécifiques chez les patients Parkinsoniens nouvellement diagnostiqués qui pourraient conduire à un diagnostic plus précoce et à un sous-typage des patients pour définir un traitement plus personnalisé et ralentir la progression de la maladie.
Pour atteindre cet objectif, nous avons exploré trois approches différentes. Premièrement, nous avons recherché des changements morphométriques dans les données structurelles en utilisant des techniques standard telles que VBM, DBM et SBM. Deuxièmement, nous avons conçu une méthodologie à base d'apprentissage profond pour détecter les anomalies cérébrales dans les IRM de diffusion. Enfin, nous avons développé une approche statistique originale basée sur des modèles de mélange générés à partir des lois de distribution de Student. Cela nous permet de pour construire un modèle de normalité à partir de données quantitatives multimodales de sujets sains, pour ensuite détecter les anomalies présentes chez nos patients.