

Abstract

In this thesis, the power of Machine Learning (ML) algorithms is combined with brain connectivity patterns, using Magnetic Resonance Imaging (MRI), for classification and prediction of Multiple Sclerosis (MS). White Matter (WM) as well as Grey Matter (GM) graphs are studied as connectome data types. The thesis addresses three main research objectives. The first objective aims to generate realistic brain connectomes data for improving the classification of MS clinical profiles in cases of data scarcity and class imbalance. To solve the problem of limited and imbalanced data, a Generative Adversarial Network (GAN) was developed for the generation of realistic and biologically meaningful connectomes. This network achieved a 10% better MS classification performance compared to classical approaches.

As second research objective, we aim to improve classification of MS clinical profiles using morphological features only extracted from GM brain tissue. First, regional thickness degeneration was investigated. Compared to the current literature, which mostly extracts information from the WM connectome, our approach requires only the T1-weighted image modality, which represents the most common modality in clinical practice. Hence, these images are easily and cheaply acquired, thereby justifying the lower scores in classification performance. Furthermore, as an extension, multiple morphological features of the GM connectome were combined using a multiview kernel-based tensor factorization approach. The use of multiple morphological biomarkers describing the degeneration of the GM tissue, induced by the MS disease, enabled a 10% better discrimination between the early MS stages and the progressive MS clinical profiles.

The third objective tackles the problem of automatic and interpretable evaluation of MS progression based on cognitive and physical patients' disability status. Multiple machine learning models were combined together to generate ensemble models by following a hierarchical structure, known as stacking generalization procedure. However, due to the relevance as well as the impact on the patient's daily activities, this disability estimation analysis was also performed by adding information from the diffusion signal obtained from tractography. In fact, although very informative, structural brain connectivity does not take into account important fiber tracks such as the Cortico-Spinal Track and the Corpus Callosum which are known to be heavily involved in the process of tissue degeneration. Additionally, in order to offer more robust guidance to the neurologist's decision toward a better diagnosis and treatment of the patient's status, a new interpretability model was implemented with the aim to pin down the most important connections between pairs of GM regions, deemed to be relevant by the machine learning model for disability estimation. In conclusion, the multiple contributions proposed in this doctoral thesis, represent one of the preliminary attempts to exploit the power of connectome data analysis together with advanced machine learning algorithms for the classification and prediction of MS disability using MRI information from the white and grey matter tissue. These contributions pave the way toward a better understanding of neurodegenerative pathways being formed inside the brain of MS patients.

Résumé

L'Intelligence Artificielle (IA), et plus particulièrement, les algorithmes d'apprentissage automatique (Machine Learning (ML)) constituent aujourd'hui une approche nouvelle d'analyse et d'aide au diagnostic comme au pronostic des maladies cérébrales. Dans ce travail, la puissance des méthodes de ML est combinée aux modèles de représentation de la connectivité cérébrale à l'aide de la théorie des graphes, pour la classification des formes cliniques de sclérose en plaques (SEP) et la prédiction de l'évolution de la maladie. Grâce à l'imagerie par résonance magnétique (IRM), les images anatomiques pondérées en T1 (T1w) et les images de tenseur de diffusion (DTI) permettent de reconstruire des graphes de la Substance Grise (SG) et de la Substance Blanche (SB) permettant de caractériser respectivement la connectivité cérébrale morphologique et structurale. Ce travail de recherche est organisé autour de trois objectifs principaux.

Le premier objectif vise à générer des données de connectome cérébral réalistes en cas de manque de données et de déséquilibre des classes pour améliorer la classification des profils cliniques de SEP. Pour résoudre ce problème, un réseau de neurones de type adversarial génératif (GAN) a été développé pour la génération de données de connectomes biologiquement réalistes. Ce réseau a permis d'améliorer de 10% les performances de classification de profils cliniques de SEP par rapport aux approches classiques.

Le deuxième objectif de recherche est d'améliorer la classification des profils cliniques de SEP en utilisant des caractéristiques morphologiques extraites du tissu cérébral de SG. Notre méthode s'est focalisée sur la mesure de l'épaisseur de SG corticale, dans la mesure où elle constitue un marqueur de la neurodégénérescence. Les résultats de classification obtenus ne sont pas très performants (F1 de l'ordre de 65%) mais cette méthode présente le grand avantage de nécessiter uniquement des images IRM pondérées en T1 qui sont facilement disponibles en pratique clinique. Ensuite, dans le cadre d'une extension de cette approche à l'analyse de multiples caractéristiques morphologiques du connectome de SG, une méthode de factorisation tensorielle "multiview" a été développée. L'utilisation de plusieurs biomarqueurs morphologiques décrivant la dégénérescence du tissu de SG, induite par la maladie de la SEP, a permis d'améliorer de 10% la discrimination entre les stades précoces de la SEP et les profils cliniques de SEP progressive.

Le troisième objectif concerne la question de l'évaluation automatique et interprétable de la progression de la maladie, définie par le statut d'invalidité cognitive et physique du patient. Plusieurs modèles d'apprentissage automatique (ML) ont été combinés pour générer des modèles d'ensemble en suivant une structure hiérarchique, connue sous le nom de "Stacking Generalization Scheme" en s'appuyant sur des données d'IRM de diffusion et sur les graphes de connectivité structurale obtenus après tractographie. Les résultats montrent une bonne performance d'estimation en réduisant le RMSE (Root Mean Squared Error) d'environ 15% par rapport aux études précédentes. Néanmoins, les modèles de connectivité cérébrale structurale ne prennent pas en compte d'importants faisceaux de fibres tels que le faisceau cortico-spinal et le corps calleux qui sont connus pour être fortement endommagés dans la SEP. C'est pourquoi, il a été développé une analyse par faisceaux de fibres de SB donnant d'excellents résultats d'estimation de l'EDSS, marqueur du handicap physique. En outre, afin d'offrir une orientation plus solide à la décision du neurologue vers un meilleur diagnostic et la décision d'un meilleur traitement médical, un nouveau modèle d'interprétabilité a été mis en œuvre afin d'identifier les faisceaux de fibre ou les connexions entre les paires de régions SG les plus importants, jugées pertinentes par le modèle d'apprentissage automatique pour l'estimation du handicap.

En conclusion, ces multiples développements méthodologiques, proposés dans cette thèse de doctorat, représentent l'une des premières tentatives d'exploiter la puissance de l'analyse des données du connectome avec des algorithmes avancés d'apprentissage automatique (ML) pour la classification et la prédiction dans la SEP. Ces contributions ouvrent la voie à une meilleure compréhension des processus neurodégénératifs et inflammatoires qui caractérisent le cerveau des patients atteints de SEP.