

## Sujet de thèse- PhD thesis

**Modèle pronostique de la maladie de Parkinson à des stades précoces par apprentissage profond combinant des données cliniques, biologiques et d'imagerie anatomique et moléculaire**

**Unité de recherche :** Laboratoires CREATIS – Equipe MYRIAD et CRNL - Equipe PathPark

**Co-direction de thèse :** Carole LARTIZIEN Carole (DR CNRS) [carole.lartizien@creatis.insa-lyon.fr](mailto:carole.lartizien@creatis.insa-lyon.fr)

Bénédicte BALLANGER (CR CNRS) [benedicte.ballanger@cnrs.fr](mailto:benedicte.ballanger@cnrs.fr)

**Mots-clés :** Apprentissage profond, modèle pronostic, détection d'anomalies, données multimodales (Imagerie IRM, TEP, données cliniques).

### 1. Contexte scientifique

La maladie de Parkinson (MP) est une affection neurodégénérative progressive, principalement caractérisée par des symptômes moteurs (e.g. tremblements, rigidité, bradykinésie), résultant de la dégénérescence des neurones dopaminergiques (DA) dans la substance noire du cerveau. Actuellement, il n'existe pas de traitement curatif pour la MP. Les traitements pharmacologiques visent à compenser la perte de dopamine, mais ne parviennent pas à traiter tous les symptômes moteurs, (e.g. akinésie), ni les symptômes non moteurs associés (e.g. dépression, anxiété, troubles du sommeil).

La complexité de la prise en charge thérapeutique de la MP pourrait être liée à la dégénérescence d'autres systèmes de neurotransmetteurs, notamment le système noradrénergique (NA). Cette hypothèse est difficile à tester, car la dégénérescence du système NA précède celle du système DA et commence bien avant l'apparition des premiers symptômes, qui ne se manifestent qu'après environ 80% de mort cellulaire. Cette phase prodromale est cruciale pour comprendre les mécanismes physiopathologiques de la MP, mais elle est évidemment inaccessible rétrospectivement au diagnostic. De plus, l'étude du système NA a longtemps été limitée par l'absence de solutions d'imagerie moléculaire in vivo chez l'homme.

Des recherches récentes associant l'équipe de B Ballanger ont permis de développer un radiotracer, la 11C-Yohimbine, pour l'imagerie TEP des récepteurs  $\alpha_2$  du cerveau (Laurencin et al., 2021, 2023). Ces travaux ont confirmé pour la première fois in vivo le lien entre les altérations du système NA et les symptômes moteurs et non moteurs de la MP (Laurencin et al., 2024). Ces avancées ouvrent la voie à de nouvelles solutions cliniques et thérapeutiques. L'équipe de B. Ballanger a commencé à constituer une base de données pour analyser les effets du vieillissement du système NA, avec l'intention d'élargir cette cohorte au fil des projets futurs. L'analyse des marqueurs prédictifs de la MP à partir de ces données représente une opportunité précieuse pour le diagnostic

précoce et le développement de traitements personnalisés adaptés aux différentes typologies de la maladie récemment mises en évidence (Chaudhury et al., 2023; Horsager & Borghammer, 2024).

La détection et caractérisation de ces marqueurs précoces de la maladie et l'analyse de ces données pour prédire le stade et l'évolution de la pathologie est une tâche complexe pour les cliniciens, car elle nécessite de fusionner des données d'imagerie massives avec des données cliniques hétérogènes (données biologiques, compte-rendu d'examens...). Les méthodes d'analyse statistique par apprentissage profond proposent un cadre méthodologique bien adapté pour l'analyse de ces données multimodales et hétérogènes.

Dans le cadre de la maladie de Parkinson ou des pathologies neurodégénérative est très actif. L'état de l'art actuel propose des modèles de classification de patients à des stades avancés de la pathologie, avec des modèles basés sur des approches supervisées à partir d'architectures convolutionnels ou basés sur des transformers ceratines commençant à intégrer des informations cliniques.

Peu de travaux dans ce domaine s'intéressent à la caractérisation d'anomalies subtiles et difficilement discernable au premier stade de la pathologie. Cette tâche est difficile et nécessite de mettre en œuvre des solutions méthodologiques dans des domaine clés de l'apprentissage profond concernant, par exemple la fusion de données hétérogènes (images, signal, données cliniques), la construction de modèles explicables, l'apprentissage sur des bases de données incomplètes, et la prise en compte de données longitudinales acquises à différents points d'examen de suivi du patient.

## 2. Objectif de la thèse

Notre hypothèse de recherche postule que la combinaison des informations fournies par les techniques d'imagerie multimodales, telles que l'IRM et la TEP, associées à des données cliniques et neuropsychologiques, pourrait permettre une détection précoce des patients de novo (premier stade clinique) et prodromaux (présymptomatiques) atteints de la MP. Nous croyons que les techniques d'intelligence artificielle (IA), en particulier l'apprentissage automatique profond, constituent le cadre méthodologique le plus adapté pour atteindre cet objectif ambitieux.

L'objectif de cette thèse est de développer un modèle diagnostique capable de discriminer les différents sous-types de patients au stade prodromal ou peu avancé de la MP et de prédire leur évolution vers des stades plus avancées. Ce modèle sera conçu pour être robuste, même avec des bases de données de petite dimension (20 à 30 sujets par classe), et explicable. L'explicabilité est essentielle pour deux raisons: elle permettra une analyse des performances guidée par l'expertise clinique, favorisant ainsi l'adhésion des cliniciens, et elle facilitera l'analyse de l'impact des biomarqueurs sur le diagnostic, aidant ainsi les neurobiologistes à comprendre les mécanismes physiopathologiques impliqués dans la conversion des patients du stade prodromal au stade pathologique précoce.

Ce projet est ambitieux car il repose sur le développement de méthodes d'IA avancées pour explorer des données multimodales et multidimensionnelles complexes, en particulier pour des groupes de patients difficiles à discriminer en raison de la subtilité des anomalies présentes à ce stade précoce. Nous visons également à intégrer l'information du biomarqueur TEP 11C-Yohimbine, en développant une stratégie efficace pour travailler avec des données en très faible régime (30 à 50 examens). En plus de développer un modèle prédictif par IA de conversion vers la MP, nous souhaitons explorer son fonctionnement, extraire les marqueurs d'intérêt pour la prédiction et proposer des pistes d'analyse des dérèglements des systèmes NA et DA durant la phase prodromale.

Une approche transdisciplinaire est cruciale pour la réussite de ce projet. L'équipe de C. Lartzien se concentrera sur l'analyse de données de neuroimagerie par des techniques avancées d'apprentissage profond, en s'appuyant, entre autres, sur des premiers travaux en neuroimagerie de la MP (Munoz, 2021, Pinon et al 2023, Oudoumanessah

et al, 2023), tandis que l'équipe de B. Ballanger apportera son expertise en imagerie et sur les mécanismes physiopathologiques de la MP. Enfin, la collaboration avec des cliniciens spécialisés dans les pathologies neurodégénératives sera essentielle pour définir les questions cliniques, guider dans le choix des a priori à modéliser et intégrer dans le modèle d'IA, faciliter l'accès aux bases de données et évaluer le modèle tout au long de son développement.

### 3. **Méthodologie** mise en œuvre, en soulignant les verrous et enjeux ainsi que le calendrier

La première partie de la thèse sera consacrée à l'extraction de l'information pertinente contenue dans des données d'imagerie IRM massives (volume 3D > 1 million de voxels) tout en tenant compte des données cliniques de plus faible dimension, et dans un contexte d'apprentissage à faible régime.

Différentes approches pourront être explorées. Un état de l'art précis sera réalisé au début de la thèse. Une piste de recherche possible pourrait être de décomposer cette tâche en deux phases :

- Dans la première phase, nous pourrions utiliser des bases de données publiques d'images IRM de grande taille, telles que OpenBHB et UK Biobank, pour entraîner un modèle d'encodage d'images auto-supervisé. Cette stratégie permettrait d'extraire une représentation synthétique et cohérente des données. Nous prévoyons d'appliquer des méthodes d'apprentissage contrastif, guidées par des a priori cliniques, pour structurer l'espace latent et rapprocher les représentations de patients présentant des caractéristiques communes.
- La seconde phase consisterait à affiner (fine-tuning) les paramètres des encodeurs pré-entraînés sur une base de données incluant des sujets normaux, des patients atteints de la MP au stade prodromal ou de novo. La fusion des représentations extraites des images et des données cliniques serait réalisée, par exemple à l'aide de mécanismes d'attention croisée, permettant une intégration efficace des informations ou par des techniques plus récemment proposées d'apprentissage sur graphe multimodal.

Dans la deuxième moitié de la thèse, l'accent sera mis sur l'intégration des données d'imagerie TEP utilisant le traceur 11C-Yohimbine. Cette phase présente des défis en raison de la taille réduite de la base de données (30-50 patients) et de la spécificité du radiotracer, rendant le pré-entraînement auto-supervisé décrit précédemment impossible. Pour surmonter ces limitations, nous explorerons des pistes alternatives, par exemple l'intégration hiérarchique de caractéristiques extraites manuellement de la TEP ou l'utilisation de modèles génératifs multimodaux combinant la synthèse de données TEP manquantes à partir des IRM et données cliniques et l'apprentissage joint de représentations multimodales.

Dans la dernière partie, nous déploierons un protocole d'évaluation des performances des modèles conçus dans les différentes phases du projet, en collaboration avec l'équipe de cliniciens. Nous évaluerons différentes stratégies d'explicabilité, par exemple en explorant les approches contrefactuelles basées sur des graphes causaux, pour déterminer les régions de l'image et les données cliniques les plus discriminantes pour la détection des différentes catégories de patients.

### 4. **Contributions originales attendues :**

D'un point de vue méthodologique, cette thèse vise à proposer des contributions méthodologiques originales dans le domaine de modèles prédictifs explicables par apprentissage sur des données hétérogènes. Ces modèles

pourront s'appuyer sur des modèles de détection d'anomalies subtiles qui pourront également faire l'objet de contributions méthodologiques

D'un point de vue applicatif, le formalisme développé permettra de détecter des patients atteints de la maladie de Parkinson à des stades précoces ou prodromaux, et extraire des marqueurs d'imagerie et clinique de cette pathologie, permettant ainsi de mieux comprendre ses mécanismes physiopathologiques, et in fine, de développer des thérapies plus performantes.

### 5. Environnement de travail et encadrement

L.e.a doctorant.e sera co-encadré par Carole Lartizien, Directrice de Recherche CNRS dans l'équipe MYRIAD du laboratoire CREATIS et Bénédicte Ballanger, CR CNRS responsable de l'équipe PathPark du CRNL. Carole Lartizien apportera son expertise méthodologique en apprentissage profond et en analyse d'images cérébrales. Bénédicte Ballanger apportera son expertise dans les domaines de l'imagerie et des mécanismes physiopathologiques sous-jacents au développement de la maladie de Parkinson, en particulier concernant le rôle du système noradrénergique.

Il ou elle sera hébergé.e au laboratoire CREATIS et bénéficiera d'un accès au CRNL. La proximité géographique des deux collaboratrices garantira un suivi proche et régulier. Le matériel informatique nécessaire sera fourni par le laboratoire, de même que des ressources de calcul pour les expériences, en particulier l'accès au centre de calcul national Jean-Zay.

Les collaborateurs cliniciens des HCL impliqués dans ce projet sont le Dr C. Laurencin et le Pr S. Thobois spécialiste de l'imagerie du Parkinson et membre du comité exécutif du réseau national NS-Park (<https://parkinson.network/>) de recherche clinique sur la MP.

L.e.a candidat.e aura accès à plusieurs bases de données :

- La base de données libre PPMI (<https://www.ppmi-info.org/access-dataspecimens/download-data>) contenant des examens d'imagerie multi-modalités et des données cliniques associés de sujets contrôles et de patients parkinsoniens à différentes phases de la MP, dont des patients prodromaux.
- La base de données libre OpenBHB (<https://ieee-dataport.org/open-access/openbhb-multi-site-brain-mri-dataset-age-prediction-and-debiasing>) qui agrège 10 bases de données de neuroimagerie et servira à l'entraînement auto-supervisé.
- La base privée collectée par l'équipe de B. Ballanger, dans le cadre du projet ANR INSyst et qui regroupe les examens IRM (T1, FLAIR) et TEP de 30 sujets contrôle et 30 patients. Un projet ANR en cours d'évaluation vise à compléter cette base de données par l'acquisition de 165 examens supplémentaires (dont 45 incluant des patients parkinsoniens).
- Nous solliciterons également l'accès à des bases de données collectées dans le cadre de NS-PARK (eg PREDISTIM).

Nous rechercherons des financements supplémentaires (e.g. France Parkinson, Fondation de France), en particulier pour la participation à des conférences/colloques internationaux ou nationaux.

### 6. Financement

Ce sujet bénéficie d'un financement doctoral du CNRS alloué par le CNRS dans le cadre de l'appel d'offre interdisciplinaire 2025 de la « Feuille de route Santé du CNRS ».

### 7. Compétences

Nous sommes à la recherche d'un étudiant ou d'une étudiante motivé et autonome avec un fort intérêt dans la recherche pluridisciplinaire (Apprentissage statistique et traitement d'image dans un contexte médical). De l'expérience avec des bibliothèques de machine learning (Tensorflow, Pytorch ou Scikit learn) et les systèmes d'exploitation sous noyau Linux seront appréciées.

### 8. Contact

Merci d'envoyer votre CV accompagné d'une lettre de motivation et des bulletins de notes les plus récents à [carole.lartizien@creatis.insa-lyon.fr](mailto:carole.lartizien@creatis.insa-lyon.fr)

### 9. Références

Laurencin C, Lancelot S, Brosse S, Mérida I, Redouté J, Greusard E, Lamberet L, Liotier V, Le Bars D, Costes N, Thobois S, Boulinguez P, **Ballanger B.** (2024) Noradrenergic alterations in Parkinson's disease: a combined [11C]yohimbine PET/neuromelanin study. *Brain*. 147(4):1377-1388.

Laurencin C, Lancelot S, Mérida I, Costes N, Redouté J, Le Bars D, Boulinguez P, **Ballanger B.** (2023) Distribution of  $\alpha$ 2-Adrenergic Receptors in the Living Human Brain Using [11C]yohimbine PET. *Biomolecules*. 13, 843.

Laurencin C, Lancelot S, Gobert F, Redouté J, Mérida I, Lecker T, Liger F, Irace Z, Greusard E, Lamberet L, Bars DL, Costes N, **Ballanger B.** (2021) Modeling [11C]yohimbine PET human brain kinetics with test-retest reliability, competition sensitivity studies and search for a suitable reference region. *Neuroimage* 240:118328.

Chaudhuri, RK., Leta, V., Bannister, K., Brooks, D.J., Svenningsson, P (2023). The noradrenergic subtype of Parkinson disease: from animal models to clinical practice. *Nat Rev Neurol* 19, 333–345.

Horsager J, Borghammer P (2024). Brain-first vs. body-first Parkinson's disease: An update on recent evidence. *Parkinsonism Relat Disord*. 122:106101.

Pinon N, Robin Trombetta R and **Lartizien C**, One-Class SVM on siamese neural network latent space for unsupervised anomaly detection on brain MRI white matter hyperintensities, *6th International Conference on Medical Imaging with Deep Learning (MIDL)*, Jul 2023, Nashville, USA, <https://openreview.net/forum?id=c9r6-HCEaN>

Oudoumanessah G, **Lartizien C**, Dojat M, Forbes F. Towards frugal unsupervised detection of subtle abnormalities in medical imaging. 26th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Oct 2023, Vancouver (BC), Canada. pp.1-13.

Munoz-Ramirez V, Pinon N, Forbes F, **Lartizien C**, and Dojat M (2021). Patch vs. global image-based unsupervised anomaly detection in mr brain scans of early parkinsonian patients. In Abdulkadir, A., Kia, S. M., Habes, M., Kumar, V., Rondina, J. M., Tax, C., and Wolfers, T., editors, *Machine Learning in Clinical Neuroimaging*, pages 34–43, Cham. Springer International Publishing.