

Université de Montréal

Utilisation de l'intelligence artificielle pour identifier les marqueurs de la démence dans le  
trouble comportemental en sommeil paradoxal

Par

Loubna Mekki Berrada

Département de psychologie, Faculté des arts et des sciences

Thèse présentée en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)

en psychologie

Août 2023

© Mekki Berrada, Loubna

Université de Montréal

Département de psychologie, Faculté des arts et des sciences

---

*Cette thèse intitulée*

**Utilisation de l'intelligence artificielle pour identifier les marqueurs de la démence dans le trouble comportemental en sommeil paradoxal**

*Présentée par*

**Loubna Mekki Berrada**

*A été évalué(e) par un jury composé des personnes suivantes*

**Julie Carrier**

Présidente-rapporteuse

**Karim Jerbi**

Directeur de recherche

**Jean-François Gagnon**

Codirecteur

**Pierre Orban**

Membre du jury

**Stuart Fogel**

Examineur externe (pour une thèse)

## Résumé

La démence à corps de Lewy (DCL) et la maladie de Parkinson (MP) sont des maladies neurodégénératives touchant des milliers de Canadiens et leur prévalence croît avec l'âge. La MP et la DCL partagent la même pathophysiologie, mais se distinguent par l'ordre de manifestation des symptômes : la DCL se caractérise d'abord par l'apparition d'un trouble neurocognitif majeur (démence), tandis que la MP se manifeste initialement par un parkinsonisme. De plus, jusqu'à 80% des patients avec la MP développeront une démence (MPD). Il est désormais établi que le trouble comportemental en sommeil paradoxal idiopathique (TCSPi) constitue un puissant prédicteur de la DCL et la MP. En effet, cette parasomnie, marquée par des comportements indésirables durant le sommeil, est considérée comme un stade prodromal des synucléinopathies, telles que la MP, la DCL et l'atrophie multisystémique (AMS). Ainsi, la majorité des patients atteints d'un TCSPi développeront une synucléinopathie.

Malgré les avancées scientifiques, les causes du TCSPi, de la MP et de la DCL demeurent inconnues et aucun traitement ne parvient à freiner ou à arrêter la neurodégénérescence. De plus, ces pathologies présentent une grande hétérogénéité dans l'apparition et la progression des divers symptômes. Face à ces défis, la recherche vise à mieux cerner les phases précoces/initiales et les trajectoires évolutives de ces maladies neurodégénératives afin d'intervenir le plus précocement possible dans leur développement. C'est pourquoi le TCSPi suscite un intérêt majeur en tant que fenêtre d'opportunités pour tester l'efficacité des thérapies neuroprotectrices contre les synucléinopathies, permettant d'agir avant que la perte neuronale ne devienne irréversible. Le TCSPi offre ainsi une occasion unique d'améliorer la détection de la démence et le suivi des individus à haut risque de déclin cognitif. D'où l'importance cruciale de pouvoir généraliser les résultats issus de la recherche sur de petites cohortes à l'ensemble de la population.

Sur le plan de la cognition, les études longitudinales sur le TCSPi ont montré que les atteintes des fonctions exécutives, de la mémoire verbale et de l'attention sont les plus discriminantes pour différencier les individus qui développeront une démence de ceux qui resteront idiopathiques. De plus, un grand nombre de patients TCSPi souffrent d'un trouble neurocognitif mineur ou trouble cognitif léger (TCL), généralement considéré comme un stade précurseur de la démence. Les recherches actuelles sur les données cognitives chez cette population offrent des

perspectives prometteuses, mais reposent sur des approches statistiques classiques qui limitent leur validation et généralisation. Bien qu'elles offrent une précision élevée (80 à 85%) pour détecter les patients à risque de déclin cognitif, une amélioration est nécessaire pour étendre l'utilisation de ces marqueurs à une plus large échelle. Depuis les années 2000, l'accroissement de la puissance de calcul et l'accès à davantage de ressources de mémoire ont suscité un intérêt accru pour les algorithmes d'apprentissage machine (AM). Ces derniers visent à généraliser les résultats à une population plus vaste en entraînant des modèles sur une partie des données et en les testant sur une autre, validant ainsi leur application clinique.

Jusqu'à présent, aucune étude n'a évalué les apports de l'AM pour la prédiction de l'évolution des synucléinopathies en se penchant sur le potentiel de généralisation, et donc d'application clinique, à travers l'usage d'outils non invasifs et accessibles ainsi que de techniques de validation de modèles (*model validation*). De plus, aucune étude n'a exploré l'utilisation de l'AM associée à des méthodes de généralisation sur des données neuropsychologiques longitudinales pour élaborer un modèle prédictif de la progression des déficits cognitifs dans le TCSPi.

L'objectif général de cette thèse est d'étudier l'apport de l'AM pour analyser l'évolution du profil cognitif de patients atteints d'un TCSPi. Le premier chapitre de cette thèse présente le cadre théorique qui a guidé l'élaboration des objectifs et hypothèses de recherche. Le deuxième chapitre est à deux volets (articles). Le premier vise à fournir une vue d'ensemble de la littérature des études ayant utilisé l'AM (avec des méthodes de généralisation) pour prédire l'évolution des synucléinopathies vers une démence, ainsi que les lacunes à combler. Le deuxième volet vise à explorer et utiliser pour la première fois l'AM sur des données cliniques et cognitifs pour prédire la progression vers la démence dans le TCSPi, dans un devis longitudinal. Enfin, le dernier chapitre de la thèse présente une discussion et une conclusion générale, comprenant un résumé des deux articles, ainsi que les implications théoriques, les forces, les limites et les orientations futures.

**Mots-clés :** Synucléinopathies, trouble comportemental en sommeil paradoxal, maladie de Parkinson, démence à corps de Lewy, cognition, apprentissage machine, Random Forest, généralisation.

## Abstract

Lewy body dementia (LBD) and Parkinson's disease (PD) are neurodegenerative diseases affecting thousands of Canadians, and their prevalence increases with age. PD and DLB share the same pathophysiology, but differ in the order of symptom manifestation: DLB is characterized first by the onset of a major neurocognitive disorder (dementia), whereas PD initially manifests as parkinsonism. Moreover, up to 80% of PD patients will go on to develop dementia (PDD). It is established that idiopathic REM sleep behavior disorder (iRBD) is a powerful predictor of DLB and PD. Indeed, this parasomnia, marked by undesirable behaviors during sleep, is considered a prodromal stage of synucleinopathies, such as PD, DLB and multisystem atrophy (MSA). Therefore, the majority of patients with iRBD will develop synucleinopathy.

Despite scientific advancements, the causes of iRBD, PD, and DLB remain unknown and no treatment has been able to slow or halt neurodegeneration. Furthermore, these pathologies display great heterogeneity in the onset and progression of various symptoms. Faced with these challenges, research aims to better understand the early/initial stages and the progressive trajectories of these neurodegenerative diseases in order to intervene as early as possible in their development. This is why iRBD garners major interest as a window of opportunities to test the effectiveness of neuroprotective therapies against synucleinopathies, enabling action to be taken before neuronal loss becomes irreversible. iRBD thus provides a unique opportunity to improve dementia detection and monitoring of individuals at high risk of cognitive decline. Hence the crucial importance of being able to generalize results of research on small cohorts to the entire population.

In terms of cognition, longitudinal studies on iRBD have shown that impairments in executive functions, verbal memory, and attention are the most discriminating in differentiating between individuals who will develop dementia from those who will remain idiopathic. In addition, many iRBD patients suffer from a mild neurocognitive disorder or mild cognitive impairment (MCI), generally considered as a precursor stage of dementia. Current research on cognitive data in this population offers promising prospects, but relies on traditional statistical approaches that limit their validation and generalizability. While they provide high accuracy (80 to 85%) for detecting patients at risk of cognitive decline, improvement is needed to extend the use of these

markers to a larger scale. Since the 2000s, increased computational power and access to more memory resources have sparked growing interest in machine learning (ML) algorithms. These aim to generalize results to a broader population by training models on a subset of data and testing them on another, thus validating their clinical application.

To date, no study has assessed the contributions of ML for predicting the progression of synucleinopathies, focusing on the potential for generalization, and hence clinical application, through the use of non-invasive, accessible tools and model validation techniques. Moreover, no study has explored the use of ML in conjunction with generalization methods on longitudinal neuropsychological data to develop a predictive model of cognitive deficit progression in iRBD.

The general objective of this thesis is to study the contribution of ML in analyzing the evolution of the cognitive profile of patients with iRBD. The first chapter of this thesis presents the theoretical framework that guided the formulation of the research objectives and hypotheses. The second chapter is in two parts (articles). The first aims to provide an overview of the literature of studies that have used ML (with generalization methods) to predict the progression of synucleinopathies to dementia, as well as the gaps that need to be filled. The second part aims to explore and use for the first time ML on clinical and cognitive data to predict progression to dementia in iRBD, in a longitudinal design. Finally, the last chapter of the thesis presents a discussion and a general conclusion, including a summary of the two articles, as well as theoretical implications, strengths, limitations, and future directions.

**Keywords:** Synucleinopathies, Rapid Eye Movement Sleep Behavior Disorder, Parkinson's Disease, Lewy Body Dementia, cognition, machine learning, Random Forest, generalization.

# Table des matières

Résumé .....	3
Abstract .....	5
Table des matières .....	7
Liste des tableaux .....	10
Liste des figures .....	11
Liste des Sigles et Abréviations .....	12
Remerciements .....	14
Chapitre 1 – Contexte théorique.....	16
1. Le sommeil.....	17
1.1 Son étude : électroencéphalographie (EEG) .....	17
1.2 Stades du sommeil.....	17
a) Caractéristiques de chaque stade.....	17
b) Circuits nerveux sous-jacents.....	18
1.3 Parasomnie : trouble comportemental en sommeil paradoxal .....	20
a) Caractéristiques cliniques du TCSP .....	20
b) Prévalence et pathophysiologie du TCSP .....	21
2. TCSP et synucléinopathies (MP et DCL) .....	22
2.1 Évolution du TCSP.....	22
a) Épidémiologie .....	22
b) Modèles explicatifs pathophysiologiques .....	23
2.2 L'évolution redoutée : la démence à corps de Lewy.....	24
2.3 Le trouble cognitif Léger (TCL) dans le TCSP.....	26
3. Performance cognitive.....	27

3.1	Dans le TCSP .....	27
a)	Études transversales .....	27
b)	Études longitudinales .....	27
3.2	Dans la maladie de Parkinson (MP).....	29
a)	Profil cognitif .....	29
b)	Études longitudinales .....	30
4.	L'analyse quantitative de l'EEG dans les synucléinopathies.....	32
4.1	EEG chez le DCL.....	34
4.2	Anomalies EEG dans le TCSP .....	34
a)	Études transversales .....	34
b)	Études longitudinales .....	35
4.3	Anomalies EEG précédant le développement de la démence dans la MP .....	37
5.	L'apprentissage machine (AM) dans un contexte clinique.....	39
5.1	Les applications de l'AM dans le monde clinique .....	39
5.2	Statistiques classiques traditionnelles et apprentissage machine .....	40
5.3	L'apprentissage machine – supervisé.....	43
5.4	Les algorithmes .....	44
a)	Différents types .....	44
b)	Arbres de décision – vers le Random Forest.....	45
5.5	Nécessaire d'évaluer la généralisation en AM: <i>cross-validation, out-of-sample generalization</i> .....	48
5.6	Métriques et performance de classification.....	49
5.7	Et la chance dans tout ça? .....	50
5.8	Méthodes de sélection de variables.....	50
6.	Objectifs et hypothèses.....	52



Chapitre 2 – Article 1 : Prediction of cognitive decline in synucleinopathies from EEG and cognitive data: A review of machine learning approaches .....	55
Chapitre 3 – Article 2 : Predicting cognitive decline by clinical markers in isolated REM sleep behaviour disorder: a machine learning approach.....	136
Chapitre 4 – Discussion générale .....	182
Sommaire des résultats principaux de la thèse.....	183
Contributions théoriques, méthodologiques et empiriques.....	188
Implications, limites et directions futures .....	206
Conclusion.....	212
Références bibliographiques .....	213

# Liste des tableaux

## Chapitre 2

Article 1. Prediction of cognitive decline in synucleinopathies from EEG and cognitive data: A review of machine learning approaches

<b>Tableau 1.</b> – Example of a confusion matrix.....	99
<b>Tableau 2.</b> – Reviewed EEG studies.....	100-104
<b>Tableau 3.</b> – Reviewed Cognitive studies.....	105-110
<b>Tableau supplémentaire S1.</b> – Performance of all algorithms tested and presentation of all metrics used in review studies.....	129-135

## Chapitre 3

Article 2. Predicting cognitive decline by clinical markers in isolated REM sleep behaviour disorder: a machine learning approach

<b>Tableau 1.</b> – Inventory of 58 database variables categorized into three modalities.....	157-158
<b>Tableau 2.</b> – Performances for all classifications - multifeatures (mean through the 10 seed).....	159-160
<b>Tableau supplémentaire S1.</b> – Baseline performance and prevalence for all categories of patients that received a follow - up diagnosis.....	173-177
<b>Tableau supplémentaire S2.</b> – Features left after data cleaning for all classifications – used as input for the ML.....	178

# Liste des figures

## Chapitre 2

Article 1. Prediction of cognitive decline in synucleinopathies from EEG and cognitive data: A review of machine learning approaches

- Figure 1.** – Overview of the typical machine learning pipeline examined in this review.....111
- Figure 2.** – Systematic Review flow diagram of selected studies.....112
- Figure 3.** – Number of EEG and Cognitive papers for each classification .....113
- Figure 4.** – Notable features for each classification.....114-115
- Figure 5.** – Frequency of algorithms tested across studies for various classification problems using AUC and/or Accuracy metrics.....116

## Chapitre 3

Article 2. Predicting cognitive decline by clinical markers in isolated REM sleep behaviour disorder: a machine learning approach

- Figure 1.** – Participant flowchart.....161
- Figure 2** – DLB vs all iRBD classification.....162-163
- Figure 3** – DLB vs Controls classification..... 164-165
- Fig 4.** – Classifications for the conditions without cognitive impairment at follow-up (i.e., MCI) – DLB vs PD-noMCI and DLB vs iRBD-noMCI.....166-167
- Figure 5** – Comparative overview of optimal features in the RF model for classifying dementia-first progression from other trajectories at follow-up....  
.....168
- Figure Supplémentaire S1.** – Feature Appearance vs Importance through 10 seeds for the DLB vs all PD classification.....179
- Figure Supplémentaire S2.** – Feature Appearance vs Importance through 10 seeds for the A) DLB vs all iRBD classification and B) DLB vs all PD classification, without using the MCI variable.....180-181

## Liste des Sigles et Abréviations

AM : Apprentissage machine

ANOVA: Analyse de variance

AUC : *Area Under the Curve*, ou aire sous la courbe

CDR : *Clinical Dementia Rating*

DCL: Démence à corps de Lewy

EEG: *Electroencephalography*, ou électroencéphalographie

FN : faux-négatif

FP : faux-positif

MMSE : *Mini-Mental State Examination*

MoCA: *Montreal Cognitive Assessment*

MP: maladie de Parkinson

MPD: démence de la maladie de Parkinson

N-REM: *non-rapid eye movement sleep*, ou sommeil non paradoxal

RAVLT: Test d'apprentissage verbal auditif de Rey

RAVLT 1-5: Somme des essais 1 à 5 du Test d'apprentissage verbal auditif de Rey

REM: *rapid eye movement sleep*, ou sommeil paradoxal

RF : *Random Forest*, ou Forêt d'arbres de décision

SC : statistiques classiques

SVM : *Support Vector Machine*, ou machines à vecteurs de support

TCL: trouble cognitif léger

TCSP: trouble comportemental en sommeil paradoxal

TCSPi: trouble comportemental en sommeil paradoxal idiopathique

TMT: Trail Making Test

TMT A : parti A du Trail Making Test

TMT B : partie B du Trail Making Test

VN : vrai-négatif

VP : vrai-positif

*À ma famille, mon phare.*

## Remerciements

Ouf! On y est? Essayer de traverser un désert, sans savoir comment s’y prendre, en reculant parfois, en se demandant où on est et où on va, en ayant soif, pour que ce désert fasse finalement partie de soi et soit plus connu et agréable que toute autre destination. D’abord merci (et « t’exagères aussi parfois non? ») à la vie, pour ce voyage au PhD.

Je voudrais commencer en remerciant profondément mes deux directeurs de thèse, Karim et Jean-François. Vous incarnez pour moi les valeurs du scientifique et du chercheur, par votre curiosité, votre humilité et votre rigueur. Vous êtes également des êtres humains d’exception! Merci pour votre bienveillance et votre grand cœur.

Karim, merci de m’avoir offert de l’écoute et d’avoir toujours cru en moi, même si la programmation et l’intelligence artificielle m’étaient inconnues avant. Aussi, grâce à toi, il m’est maintenant évident que je veux évoluer professionnellement autour de personnes avec qui je peux être moi-même et me sentir vue, avec qui je peux rire (de choses profondes et d’autres qui le sont moins!), tout en étant très stimulée intellectuellement. Le *fun* est maintenant une des variables les plus importantes de mon modèle d’intelligence artificiel professionnel! Le CoCo Lab sera toujours une deuxième famille pour moi. Aux personnes avec qui j’ai passé le plus de nuits blanches à travailler (et à remettre en question l’existence) : Jordan, Tim, Hamza, Yann, Arthur. Un merci particulier à Arthur, Hamza et Yann, sans qui je n’aurais pas pu compléter tous les calculs et analyses de cette thèse. Vous êtes les *best*. Sans jamais oublier ma très chère Vanessa! *My giiiiir!* Le doctorat avec toi c’était comme aller à la piscine, mais avec plein de glissades d’eau et de bulles. C’était tellement plus amusant et intéressant. Merci d’avoir toujours été là pour moi, je suis heureuse et chanceuse de t’avoir dans ma vie.

Jean-François, merci pour ta gentillesse, ta transparence et tes conseils. Bien que je passais moins de temps au BRACLab, j’ai vraiment apprécié les quelques chalets entre membres du labo qu’on a pu faire ensemble. Merci pour le soutien et l’humour !

Ma famille. D’abord mes parents. Déjà enfant, je vous voyais compléter vos études supérieures, en jonglant entre la vie de couple, de famille, le travail et les études. C’est sûrement ce qui m’a donné l’impression que c’était « facile » (erreur 101). Merci pour votre sagesse et votre amour à travers mon doctorat. Leila, ma sœur, ma jumelle, celle que j’aime plus que tout. Merci de m’avoir soutenu et accompagné dans ce processus plus que quiconque. Je pense qu’on bat des

records de temps accumulés d'appels chez Vidéotron. Larbi, t'es le meilleur fréro qui existe. Merci de m'avoir fait rire comme personne alors que j'étais stressée et merci pour ta grande générosité. Lina, la meilleure petite sœur au monde, si créative et sensible déjà. Merci de me faire croire que tu veux aussi faire un doctorat quand tu seras plus grande, malgré que tu aies assisté à mon processus personnel !

Merci à mes ami(e)s, qui m'ont soutenu tout le long de ce processus et qui m'ont fait sortir de mon écran. J'ai beaucoup de chance, je suis tellement bien entourée.

# **Chapitre 1 – Contexte théorique**



## 1. Le sommeil

Depuis l'Antiquité, de nombreux philosophes et auteurs ont reconnu la valeur centrale du sommeil dans le fonctionnement d'un individu : « Qui fuit le sommeil court à la mort » (Jean-Jacques Rousseau, *Émile*, ou *De L'éducation* - 1762).

### 1.1 Son étude : électroencéphalographie (EEG)

Depuis 1929, l'EEG est devenu essentiel pour étudier les états de vigilance (éveil et sommeil). Il mesure le voltage (en microvolts) entre deux électrodes (c.à.d., fils électriques avec gel conducteur) sur le cuir chevelu, couvrant diverses régions cérébrales (F=frontal, P=pariétal, T=temporal, O=occipital, C=central) (Purves et al., 2019). L'EEG mesure principalement les courants émis lors de l'activation synaptique des dendrites des neurones pyramidaux du cortex cérébral (Bear et al., 2016). L'amplitude du signal observé sur le tracé EEG va refléter le degré de la synchronisation neuronale dans une région cérébrale spécifique (Hari, 2017; Purves et al., 2019). Les oscillations cérébrales, se distinguent aussi par leur fréquence, c'est-à-dire la rapidité avec laquelle les fluctuations de potentiel électrique s'effectuent (Klimesch, 1996). Les rythmes étudiés en EEG chez l'humain, ainsi que les bandes de fréquences qui y correspondent, sont : Delta  $\delta$  (<4 Hz), Thêta  $\theta$  (4 à 8 Hz), Alpha  $\alpha$  (8 à 13 Hz), Bêta  $\beta$  (14-30 ou 40 Hz) et Gamma  $\gamma$  (>30 ou 40 Hz) (Al-Qazzaz et al., 2014).

### 1.2 Stades du sommeil

#### a) Caractéristiques de chaque stade

Le sommeil se compose de différentes phases définies principalement par des critères EEG. Il se divise en sommeil lent (ou N-REM Sleep, i.e., *non-rapid eye movement sleep*), avec trois stades (N1, N2, N3), et en sommeil paradoxal (ou REM sleep, i.e., *rapid eye movement sleep*) (Dauvilliers, 2019). Le sommeil lent se caractérise par une synchronisation dominante de l'EEG (Dauvilliers, 2019) de forte amplitude avec un rythme lent, sans mouvements oculaires rapides et en présence de mouvements musculaires occasionnels (Bear et al., 2016). L'activité mentale y est

minimale et peu complexe (pensée logique, répétitive). Le sommeil paradoxal se caractérise par une paralysie musculaire, une activité EEG désynchronisée de rythme rapide, ainsi que des saccades des mouvements oculaires rapides (Montplaisir et al., 2010; Siegel, 2011). Les rêves vivides avec un contenu émotionnel plus intense surviennent majoritairement durant cette phase (pensées vives, illogiques, étranges).

Un cycle de sommeil commence par du sommeil lent léger (stade N1 et stade N2), évolue vers du sommeil lent profond (stade N3) et se termine par le sommeil paradoxal. Un cycle dure de 60 à 120 minutes et une nuit comprend 4 à 6 cycles pour un adulte moyen (Dauvilliers, 2019).

### **b) Circuits nerveux sous-jacents**

Les états EEG d'éveil et de sommeil semblent dépendre d'un équilibre entre les systèmes qui favorisent l'éveil et ceux qui favorisent le sommeil (Siegel, 2011). Ils seraient contrôlés par différentes régions du cerveau et populations neuronales, via une régulation complexe qui n'est pas encore entièrement comprise (Gompf et Anaclet, 2020). Néanmoins, des modèles existants nous offrent certaines explications et l'étude des mécanismes neurochimiques des états de veille et de sommeil continue de progresser grâce aux modèles animaux et à la génétique moléculaire (Baghdoyan et Lydic, 2012; Gompf et Anaclet, 2020).

Plusieurs expériences ont montré que le sommeil implique des interactions entre plusieurs groupes de neurones situés dans le tronc cérébral, le thalamus et le cortex cérébral. Elles ont débuté avec des lésions cérébrales animales, révélant que les structures sous le cerveau moyen facilitent les états EEG d'éveil (Siegel, 2011). À la fin des années 1940, Moruzzi et Magoun ont identifié une région du tronc cérébral, le système activateur réticulaire (Moruzzi et Magoun, 1949). Il s'agit d'un groupe de noyaux cholinergiques situé à la jonction du pont et du mésencéphale qui projette sur les neurones thalamocorticaux. En fait, les neurones thalamocorticaux du thalamus ont 1) une activité oscillatoire spontanée de grande amplitude et basse fréquence, synchronisant les neurones corticaux et 2) une activité tonique à l'inverse de la première, quand on les dépolarise. Ce deuxième cas de figure se produit à l'éveil, alors que le

système activateur réticulaire dépolarise les neurones thalamocorticaux qui vont alors à un mode de décharge tonique provoquant une réduction des décharges oscillatoires/rythmiques des neurones thalamiques et par le fait même de l'activité synchronisée des neurones corticaux qui en dépend. Une disparition des ondes lentes est alors observée pour laisser place à une désynchronisation de l'EEG et un état de vigilance et d'éveil (España et Scammell, 2011; Siegel, 2011). D'autres systèmes neuronaux sont impliqués dans l'éveil. Les neurones thalamocorticaux reçoivent des projections ascendantes des neurones noradrénergiques du locus cœruleus, des neurones sérotoninergiques des noyaux du raphé, des neurones histaminergiques du noyau tubéro-mamillaire et des neurones oxinergiques de l'hypothalamus latéral.

Lors du sommeil lent, des neurones GABAergiques inhibiteurs de l'aire préoptique ventrolatérale, située dans l'hypothalamus antérieur, sont actifs (Szymusiak et al., 1998; Takahashi et al., 2009) et par des projections vers les systèmes favorisant l'éveil, inhibent l'activité cholinergique, dopaminergique, sérotoninergique, noradrénergique, histaminergique et orexinergique (Sherin et al., 1998; Siegel, 2011). Cette réduction de l'activité excitatrice va libérer les neurones GABAergiques du noyau réticulaire thalamique, qui pourront agir efficacement pour hyperpolariser les neurones thalamocorticaux, les ramenant à leur activité oscillatoire spontanée et synchronisant les neurones corticaux. C'est ce qui permettra une forme de déconnexion du cortex cérébral par rapport au monde extérieur (Sherin et al., 1998; Siegel, 2011).

Au cours du sommeil paradoxal, le taux de sérotonine et celui des neurotransmetteurs monoaminergiques diminuent fortement, tandis que le taux d'acétylcholine augmente jusqu'à atteindre à peu près le niveau de l'état d'éveil. Lors du sommeil paradoxal, il y a une levée de l'inhibition aminergique, permettant aux neurones cholinergiques de certains noyaux du système activateur réticulaire d'avoir une activité importante, comme à l'éveil, et leur permettant d'exciter les neurones thalamocorticaux et d'inhiber l'activité des neurones réticulaires thalamiques responsables de générer les fuseaux de sommeil lors du sommeil lent (Hu et al., 1989; Steriade et al., 1987). Cela a donc pour effet de générer l'activité rapide et désynchronisée du sommeil paradoxal. En ce qui concerne le contrôle moteur (l'atonie musculaire), ce sont les neurones glutamatergiques excitateurs du noyau subcoeruleus du pont dans le tronc cérébral qui projettent aux noyaux réticulaires gigantocellulaires ventraux et alpha. Ces derniers contiennent des neurones

GABAergiques et glycinergiques qui innervent les motoneurones de la moelle épinière, les hyperpolarisant et provoquant l'atonie musculaire (Brooks et Peever, 2012; Krenzer et al., 2011; St Louis et al., 2017).

Il y a ainsi une multiplicité des systèmes et de neurotransmetteurs impliqués dans les différentes phases du sommeil. La neurodégénérescence d'un système neuronal peut ainsi générer un changement de l'activité EEG et motrice, ainsi que des symptômes cliniques associés, comme dans le cas de certaines parasomnies.

### **1.3 Parasomnie : trouble comportemental en sommeil paradoxal**

Une parasomnie est définie comme étant : « des événements physiques ou des expériences indésirables qui se produisent à l'endormissement, pendant le sommeil ou lors d'éveils partiels » (Medicine, 2014). On regroupe les parasomnies en trois catégories : 1) du sommeil lent profond, 2) du sommeil paradoxal et 3) autres (Dauvilliers, 2019). Dans le cadre de cette thèse, nous nous pencherons sur le trouble comportemental en sommeil paradoxal (TCSP). L'étude des parasomnies se fait principalement avec la polysomnographie, soit une procédure qui utilise l'EEG, l'électro-oculogramme (c.à.d., mouvements oculaires), l'électromyogramme (c.à.d., mouvements menton), l'électrocardiogramme (c.à.d., cardiaque) et l'oxymétrie (c.à.d., mesure de la saturation en oxygène du sang), ainsi que le débit d'air et l'effort respiratoire, pour évaluer les causes sous-jacentes des troubles du sommeil (Rundo).

#### **a) Caractéristiques cliniques du TCSP**

Le TCSP se caractérise par une perte de l'atonie musculaire lors du sommeil paradoxal et la survenue de vocalisations ou de comportements indésirables mimant le contenu des rêves (Sateia, 2014). On peut observer, chez les patients, de simples secousses brèves des membres et des cris. Toutefois, les comportements peuvent être plus complexes et inclure des actes agressifs et violents tels que des combats, des poursuites et des injures. Cette agitation se manifeste généralement en fin de nuit, où la proportion de sommeil paradoxal est la plus élevée (Dauvilliers, 2019). Le TCSP

idiopathique (TCSPi) fait référence au TCSP sans condition médicale concomitante (ex : parkinsonisme ou démence) (Miglis et al., 2021; Postuma et al., 2019).

Selon les critères de la troisième édition de la Classification internationale des troubles du sommeil (ICSD-3), le diagnostic de TCSP nécessite les éléments suivants : 1) la présence répétée de vocalisations ou de mouvements complexes pendant le sommeil; 2) la manifestation de ces comportements pendant la phase de sommeil paradoxal, qui peut être établie à l'aide d'une polysomnographie ou de l'historique des rêves du patient; 3) la présence de sommeil paradoxal sans atonie musculaire sur la polysomnographie; 4) l'exclusion d'autres troubles du sommeil, troubles médicaux ou neurologiques, troubles mentaux ou troubles liés à l'utilisation de substances pouvant expliquer les perturbations du sommeil (Medicine, 2014).

### **b) Prévalence et pathophysiologie du TCSP**

Le TCSP affecte généralement les adultes de plus de 50 ans, avec une prévalence estimée d'environ 0.5% à 2% dans la population générale (Frauscher et al., 2010; Kang et al., 2013; Ohayon et Schenck, 2010) et touchant plus souvent les hommes (85%) (Bodkin et Schenck, 2009; Boeve, 2010; Peever et al., 2014).

La pathogenèse du TCSP n'est pas tout à fait claire (Fantini et al., 2003). Un dysfonctionnement dopaminergique striatal peut être présent dans le TCSPi, tel que démontré dans des études d'imagerie cérébrale par une diminution de l'innervation dopaminergique nigrostriatale (Albin et al., 2000) et une réduction de la liaison du transporteur présynaptique de la dopamine dans le striatum (Eisensehr et al., 2000).

Les études provenant de modèles animaux (Hendricks et al., 1982; Jouvet, 1965) et d'observations neuropathologiques chez l'homme (Arnulf et al., 2000; Schenck et al., 1996; Turner et al., 2000; Uchiyama et al., 1995) montrent une implication de plusieurs structures du tronc cérébral, principalement situées dans le pont (jonction mésopontine ventrale, le noyau pédonculo-pontin, le noyau tegmental latérodorsal, le locus cœruleus et la zone du péri-locus cœruleus. Par

ailleurs, le noyau pédonculo-pontin et le locus cœruleus présentent de grands groupes de neurones cholinergiques et noradrénergiques (Jones, 2000). Ces réseaux neuronaux auraient une implication dans l'éveil, ainsi que dans certaines fonctions cognitives (Aston-Jones et al., 1999; Berridge et al., 1993; Kleiner et Bringmann, 1996; Sleekier et al., 1994; Winn, 1998).

Quant à l'agitation nocturne spécifiquement, elle résulterait d'altérations structurales ou fonctionnelles des systèmes neuronaux impliqués dans la régulation de l'atonie musculaire lors du sommeil paradoxal (Peever et al., 2014), soit les neurones glutamatergiques du tronc cérébral. Plus précisément, des lésions du noyau sous-latéral dorsal / sous-coeruleus résultant d'atteintes dans le pont dorsal provoqueraient une perte d'atonie lors du sommeil paradoxal et un TCSP (St Louis et al., 2017). Certains auteurs suggèrent que la propagation de l' $\alpha$ -synucléine débiterait à cet endroit, avant de s'étendre à d'autres régions cérébrales (Del Tredici et al., 2002).

## **2. TCSP et synucléinopathies (MP et DCL)**

### **2.1 Évolution du TCSP**

#### **a) Épidémiologie**

Des recherches longitudinales (Iranzo et al., 2006; Postuma et al., 2009; Schenck et al., 1996) ont mis en évidence que le TCSP représente un stade prodromal des synucléinopathies, soit la maladie de Parkinson (MP), la démence à corps de Lewy (DCL) et l'atrophie multisystématisée. La MP est définie par un parkinsonisme qui comprend la bradykinésie, accompagnée de tremblements au repos et/ou de rigidité (Postuma et al., 2015). Quant à la DCL, il s'agit d'un trouble neurocognitif majeur altérant le fonctionnement quotidien. Une étude longitudinale multicentrique récente a montré que le taux de conversion du TCSPi atteint 10,6 % à 2 ans, 17,9 % à 3 ans, 31,3 % à 5 ans, 51,4 % à 8 ans, 60,2 % à 10 ans et 80% par la suite (Postuma et al., 2019). Les patients ont évolué vers une démence dans environ 43,5% des cas, tandis que 56,5% ont développé d'abord un parkinsonisme (Postuma et al., 2019).

## **b) Modèles explicatifs pathophysiologiques**

Une explication possible du lien entre le TCSP et le développement des synucléinopathies provient d'études portant sur la progression pathologique de ces dernières (Braak et al., 2003; Halliday et al., 2011). L'alpha-synucléine subit des modifications structurales qui entraînent la formation d'agrégats, perturbant ainsi le fonctionnement normal des cellules et pouvant même provoquer leur mort (Jellinger, 2014). Des recherches ont montré que la pathologie à corps de Lewy peut être détectée chez des individus asymptomatiques, suggérant une évolution préclinique lente de la pathologie bien avant l'agrégation des  $\alpha$ -synucléines dans la substance noire lors des stades prodromiques de la MP (Berg et al., 2021). De plus, la propagation toxique de l'alpha-synucléine semble suivre une séquence spécifique, conceptualisée dans le modèle de stadification proposé par Braak, en six stades bien définis (Braak et al., 2003).

Au stade initial (stade 1), la pathologie débute dans le noyau dorsal moteur du nerf vague et dans le bulbe olfactif, ce qui peut expliquer la perte olfactive et la dysautonomie. La pathologie évolue ensuite (stade 2) vers la partie caudale du tronc cérébral (noyaux du raphé, noyaux réticulaires gigantocellulaires, complexe coeruleus–subcoeruleus), notamment dans des régions impliquées dans la physiopathologie du TCSP, tels que le noyau subcoeruleus, qui joue un rôle important dans l'atonie musculaire lors du sommeil paradoxal (McKenna et Peever, 2017). C'est à ce stade que le TCSP et des symptômes anxiodépressifs peuvent apparaître. Au stade 3, la pathologie progresse dans le mésencéphale, particulièrement dans la substance noire, entraînant l'apparition de symptômes moteurs. Au stade 4, la neurodégénérescence est suffisamment importante dans la substance noire pour permettre un diagnostic clinique de MP, avec une atteinte du prosencéphal basal et du mésocortex (région transentorhinal). Enfin, aux stades 5 et 6, les structures limbiques et le néocortex sont touchés, pouvant provoquer l'apparition de symptômes cognitifs. Il est important de souligner que bien que le modèle de Braak représente un schéma de progression courant (Halliday et al., 2008; Jellinger, 2019), certains patients peuvent présenter des variantes dans l'ordre et la séquence des symptômes (ex : TCSP après les atteintes cognitives, motrices ou autonomiques; trouble cognitif léger (TCL) en même temps que le diagnostic de MP; aucun trouble cognitif aux derniers stades malgré une pathologie importante) (Parkkinen et al., 2008). Cela met en évidence la variabilité interindividuelle dans l'évolution des synucléinopathies

et met en lumière une première limite du modèle de Braak (St Louis et al., 2017). Ainsi, les variations dans les différentes séquences de propagation sont encore un sujet d'étude important à développer.

Des alternatives au modèle de Braak ont été proposées pour mieux refléter la variabilité des maladies à corps de Lewy. Par exemple, le *Unified Staging System for Lewy Body Disorders* classe les pathologies à corps de Lewy à partir de la distribution de la pathologie en quatre stades : 1) bulbe olfactif uniquement, 2a) prédominance du tronc cérébral ou 2b) prédominance limbique, 3) tronc cérébral et limbique (distribution équivalente dans les deux), et 4) aires néocorticales. L'avancement au sein de ces stades a été associé à une élévation des scores de densité de la synucléinopathie de type Lewy d'un point de vue pathologique et à une aggravation des signes cliniques de dysfonctionnement cognitifs et moteur (Adler et Beach, 2016; Beach et al., 2009).

## **2.2 L'évolution redoutée : la démence à corps de Lewy (DCL)**

La DCL est principalement définie par un déclin cognitif progressif d'une ampleur suffisante pour perturber les fonctions sociales ou professionnelles, ou les activités quotidiennes habituelles (McKeith et al., 2017). Elle se manifeste généralement par des déficits attentionnels, exécutifs et de traitement visuel importants, alors que la mémoire et la dénomination demeurent relativement préservées au début de la maladie (Ferman et al., 2006; Hamilton et al., 2012; McKeith et al., 2017). La DCL débute habituellement après l'âge de 50 ans (McKeith et al., 2004). Elle constitue la deuxième forme de démence après la maladie d'Alzheimer (McKeith et al., 2005).

Le diagnostic de DCL est posé lorsque la démence apparaît avant ou en même temps que le parkinsonisme (McKeith et al., 2017), alors que la démence associée à la MP est diagnostiquée lorsque l'atteinte cognitive se développe dans le contexte d'une MP établie (Emre et al., 2007). La DCL se caractérise par une combinaison variable de caractéristiques cliniques principales (*core clinical features*), telles que le parkinsonisme (bradykinésie, tremblement de repos ou rigidité), le TCSP, les fluctuations de la cognition/vigilance et les hallucinations visuelles (McKeith et al., 2017; Weisman et McKeith, 2007). Un diagnostic de DCL « probable » peut être établi si au moins



deux de ces caractéristiques cliniques principales sont présentes, ou si une seule caractéristique clinique principale est associée à un biomarqueur indicatif (*indicative biomarker*). Les biomarqueurs indicatifs incluent une diminution de la capture de la dopamine dans les ganglions de la base, une scintigraphie anormale au 123I-MIBG et un sommeil paradoxal sans atonie musculaire. La DCL « possible » peut être diagnostiquée si une seule caractéristique clinique principale est présente, ou en présence d'au moins un biomarqueur indicatif (McKeith et al., 2020).

La phase prodromique de la DCL correspond à des symptômes annonciateurs de l'évolution ultérieure vers la DCL, pouvant se manifester 15 ans ou plus avant l'apparition de la démence. Cette phase inclut le TCL, le délirium (*delirium-onset*) et des manifestations psychiatriques (*psychiatric-onset*) (McKeith et al., 2020). Pour le TCL, le taux de conversion vers la DCL varie entre 5.6% et 22% (Bombois et al., 2008; Fischer et al., 2007; Palmqvist et al., 2012). La DCL prodromique peut se manifester par n'importe quel sous-type de TCL, les domaines visuospatial et exécutif étant les plus touchés (Donaghy et al., 2015). Une étude indique que la présence de signes extrapyramidaux légers constitue le meilleur prédicteur de la DCL chez les patients atteints de TCL (Cagnin et al., 2015).

Le TCL est connu comme étant un stade intermédiaire entre le fonctionnement cognitif normal et la démence (Gauthier et al., 2006), mais n'y conduit pas toujours (Fischer et al., 2007; Ganguli et al., 2004; Gauthier et al., 2006). Effectivement, certains patients atteints de TCL peuvent revenir à une cognition normale, mais sont toujours à risque de démence. On distingue quatre sous-types de TCL, définis selon nombre et le type de domaines cognitifs atteints : 1) amnésique à domaine unique (c.à.d., une atteinte isolée de la mémoire, sans remplir les critères de démence), 2) amnésique à domaines multiples (c.à.d., une atteinte de la mémoire ainsi que d'autres domaines cognitifs), 3) non-amnésique à domaine unique (c.à.d., une atteinte d'un domaine unique autre que la mémoire, telles que les fonctions exécutives, le langage ou les aptitudes visuospatiales), 4) non-amnésique à domaines multiples (c.à.d., une atteinte de plusieurs domaines, à l'exception de la mémoire).

Les preuves disponibles sont suffisantes pour proposer des critères diagnostiques uniquement pour la DCL débutant par un TCL, désignée TCL avec corps de Lewy (Armstrong,

2021; McKeith et al., 2020). Le TCL avec corps de Lewy se caractérise par des déficits dans un ou plusieurs domaines cognitifs, plus importants que ceux attendus lors du vieillissant normal, accompagnés d'une plainte cognitive de la part du patient, de l'informateur ou du clinicien, sans perturbation du fonctionnement quotidien (McKeith et al., 2020). Le diagnostic de TCL avec corps de Lewy « probable » peut être posé si deux caractéristiques cliniques principales de la DCL ou plus sont présentes, ou si une caractéristique clinique principale est associée à un biomarqueur. Le diagnostic de TCL avec corps de Lewy « possible » peut être établi en présence d'une caractéristique clinique principale de la DCL ou d'un biomarqueur (McKeith et al., 2020). Parmi les patients atteints de TCL qui développent ultérieurement une démence, la présence du TCSP constitue un prédicteur important d'une transition future vers la DCL (Cagnin et al., 2015; Sadiq et al., 2017).

### **2.3 Le trouble cognitif Léger (TCL) dans le TCSP**

Plus d'un tiers des patients atteints de TCSPi présentent un TCL (Gagnon et al., 2009; Zhang et al., 2016). Le type de TCL le plus fréquemment retrouvé dans le TCSP est le TCL non-amnésique à domaine unique avec des troubles de l'attention et des fonctions exécutives (Gagnon et al., 2012). Des preuves récentes montrent que les patients TCSPi atteints de TCL présentent des altérations importantes de la matière grise corticale et sous-corticale par rapport à ceux sans TCL (Fereshtehnejad et al., 2019). Le TCL chez les patients TCSPi est également lié à des modifications cérébrales plus sévères, notamment un amincissement cortical dans le cortex temporal antérieur gauche, une perfusion réduite et une activité EEG plus lente (Miglis et al., 2021; Rahayel et al., 2021). De plus, tout sous-type de TCL associé à un TCSPi confirmé par polysomnographie serait susceptible d'indiquer une maladie sous-jacente à corps de Lewy (McKeith et al., 2020).

La présence à la fois de TCL et de TCSPi fournit des informations précieuses sur la progression vers la démence. Plusieurs études ont montré que le TCL chez les patients avec un TCSPi au temps de base est un marqueur de phénoconversion (Ferini-Strambi et al., 2019; Postuma et al., 2019). En effet, le TCL dans le TCSPi est associé à une probabilité plus élevée de phénoconversion vers un phénotype de démence-*first*, comparativement à un phénotype

parkinsonien-*first* (Genier Marchand et al., 2017; Postuma et al., 2019; Sateia, 2014). Le développement d'une DCL dans le TCSPi est associé aux deux sous-types de TCL, soit amnésique et non-amnésique (Genier Marchand et al., 2017; Genier Marchand et al., 2018; Postuma et al., 2019; Terzaghi et al., 2019). Toutefois, certaines études suggèrent que le TCL non amnésique à domaine multiple combiné au TCSPi pourrait être plus prédictif de la progression vers la DCL (Jicha et al., 2010; Kondo et al., 2016).

### **3. Performance cognitive**

#### **3.1 Dans le TCSP**

##### **a) Études transversales**

Plusieurs études ont montré des atteintes cognitives chez les patients atteints de TCSPi, avec un profil similaire à celui des patients atteints de DCL (Gagnon, 2019), soit des performances déficitaires au niveau de l'attention, des fonctions exécutives et de la mémoire épisodique (Ferini-Strambi et al., 2004; Ferini-Strambi et al., 2019; Gagnon et al., 2009; Marques et al., 2010; Massicotte-Marquez et al., 2008; Terzaghi et al., 2008; Zhang et al., 2016). L'atteinte des habiletés visuo-spatiales demeure plus controversée, en particulier dans les stades initiaux du TCSPi (Ferini-Strambi et al., 2004; Ferini-Strambi et al., 2019; Gagnon et al., 2009; Marques et al., 2010) (Li et al., 2016; Plomhause et al., 2014; Youn et al., 2016). Des performances réduites dans les parédoles, la mémoire prospective et de la prise de décision ont aussi été documentées (Bezdicek et al., 2018; Delazer et al., 2012; Honeycutt et al., 2020; Sasai et al., 2012). Par ailleurs, le TCSPi est également associé à des modifications de l'humeur et à des altérations cérébrales, telles que révélées par l'imagerie cérébrale (Bourgouin et al., 2019; Fereshtehnejad et al., 2019; Hogl et al., 2018).

##### **b) Études longitudinales**

Des études se sont penchées sur l'évolution de la cognition dans le TCSPi à travers le temps. Par ailleurs, deux études (Fantini, 2011; Terzaghi et al., 2013) ont pu démontrer une altération de la

performance à travers le temps, sans toutefois pouvoir prédire une phénoconversion à partir d'une performance au temps de base. Fantini et al. (2011) ont suivi pendant une moyenne de 2,4 ans des patients avec un TCSPi. Ils ont trouvé qu'au suivi ces patients avaient une performance significativement plus faible que 12 patients contrôles à une tâche d'apprentissage visuo-spatial (Test d'empan de Corsi). Terzaghi et al. (2013) ont suivi pendant une moyenne de 3,5 ans 20 patients avec un TCSPi et ont montré une baisse marquée des habiletés visuospatiales chez 45% des patients. En résumé, la cognition se détériore dans le temps chez les patients avec un TCSP, particulièrement aux tâches d'attention visuelle et d'habiletés visuospatiales.

Youn et al. (2016) ont suivi 84 patients TCSPi pendant une moyenne de 4.2 ans. Au suivi, les 18 patients qui ont développé un trouble neurodégénératif présentaient une performance plus basse en attention visuelle et fonctions exécutives (Trail Making Test, partie A) en comparaison à ceux qui n'ont pas converti. Une recherche de notre laboratoire (Genier Marchand et al., 2017) a suivi 76 patients atteints de TCSPi, dont 19 ont développé la MP et 15 la DCL après une moyenne de 3.6 ans. Les patients qui développaient une DCL présentaient des déficits dans tous les domaines cognitifs (attention et fonctions exécutives, apprentissage épisodique et habiletés visuospatiales), par rapport à ceux qui développaient une MP. Deux tests évaluant l'attention et les fonctions exécutives permettaient de détecter de manière fiable (aire sous la courbe  $\geq 85$ ) ceux qui développeraient une démence : le Stroop Color Word Test et la partie B du Trail Making Test. En examinant l'évolution des déficits cognitifs dans la même cohorte, en fonction du nombre d'années précédant la conversion en démence (Genier Marchand et al., 2018), il est constaté que la mémoire (évaluée par le Test d'apprentissage verbal auditif de Rey) et la fluidité verbale sont significativement altérées deux ans avant la conversion vers la DCL. En revanche, les fonctions exécutives (mesurées par le Trail Making Test, partie B) présentent une détérioration significative six ans avant la conversion en DCL. En résumé, des atteintes attentionnelles et exécutives sont associées à une conversion en DCL et les indicateurs à surveiller quant à l'évolution des troubles au fil du temps sont des tests d'apprentissage épisodique verbale et de mémoire (Ferini-Strambi et al., 2019). Les tests de dépistage tels le Montreal Cognitive Assessment semblent aussi avoir un potentiel d'utilisation pour prédire la conversion de la DCL chez les patients avec un TCSPi (Fereshtehnejad et al., 2019). Cependant, des études supplémentaires sont nécessaires pour valider ses propriétés psychométriques dans cette population (Miglis et al., 2021).

Les principales méthodes d'analyse de données utilisées jusqu'à présent sont des comparaisons de groupes (Fantini, 2011; Fereshtehnejad et al., 2019; Genier Marchand et al., 2017; Genier Marchand et al., 2018; Terzaghi et al., 2013; Youn et al., 2016), des corrélations (Fantini, 2011; Genier Marchand et al., 2017) des régressions (Fereshtehnejad et al., 2019; Genier Marchand et al., 2018; Terzaghi et al., 2013; Youn et al., 2016) ou encore des régressions logistiques (Genier Marchand et al., 2017). Il convient de souligner l'absence d'études en apprentissage machine (AM) utilisant des méthodes de généralisation pour examiner les prédicteurs neuropsychologiques de la démence dans le TCSPi. Cela s'avère crucial étant donné le caractère prodromique du TCSPi, pouvant précéder le diagnostic de démence de plusieurs années. En effet, pour cibler efficacement cette pathologie, il est essentiel d'utiliser des modèles d'AM qui permettent une certaine généralisation. En résumé, bien que certains marqueurs aient été identifiés, une certaine hétérogénéité existe et leur utilisation clinique demeure limitée en raison des difficultés de généralisation. De plus, la valeur prédictive des tests cognitifs pour la démence chez les TCSPi est bonne mais peut être améliorée.

### **3.2 Dans la maladie de Parkinson (MP)**

Comme mentionné précédemment, le TCSPi représente un facteur de risque majeur pour le développement de la démence, mais également d'autres synucléinopathies telles que la MP. Dans cette section, nous aborderons succinctement le profil cognitif observé dans la MP, ainsi que les marqueurs prédictifs d'une conversion vers la démence.

#### **a) Profil cognitif**

Dans les premiers stades de la MP, il y a déjà près de 24 à 36% des patients avec MP qui ont des troubles cognitifs (Foltynie et al., 2004; Muslimovic et al., 2005), alors que 3 à 8 ans après le diagnostic de MP, cette proportion est de 50% à 80% (Aarsland et al., 2009; Muslimovic et al., 2009; Williams-Gray et al., 2007). L'atteinte cognitive la plus précoce chez les patients avec MP est une dysfonction exécutive (Troster, 2008; Zgaljardic et al., 2003). Apparaît ensuite une perturbation de la mémoire épisodique verbale, principalement caractérisée par un déficit dans le

rappel libre des informations, ainsi que des anomalies visuo-spatiales qui se développent plus tardivement (Troster, 2008). Cela résulterait de la pathologie primaire dans la MP, l'épuisement de dopamine, qui perturberait le bon fonctionnement des circuits neuronaux fronto-striataux (Alexander, 1986; Owen, 2004). La dysfonction exécutive se manifeste par des problèmes de mémoire de travail (Lewis et al., 2005), de planification, d'initiation et de contrôle (Morris et al., 1988), une atteinte de l'inhibition (Gauggel et al., 2004) et des difficultés de flexibilité mentale (Monchi et al., 2004). Les patients ont ainsi des performances altérées dans diverses tâches mesurant les fonctions exécutives, telles : la fluence verbale phonémique et sémantique, le Wisconsin Card Sorting Test, le test de Stroop (interférence), la partie B du Trail Making Test (Kudlicka et al., 2011), ainsi que des problèmes de récupération de l'information en mémoire (Dujardin et al., 2001; Pillon et al., 1997; Stebbins et al., 1999). Les patients avec MP démontrent également des troubles au niveau des fonctions visuospatiales (Kerai et al., 2012) et au niveau de la vitesse de traitement des informations (Muslimovic et al., 2009). Lorsque les patients avec une MP ont également un TCSP, ils performent moins bien que les patients avec une MP sans TCSP à la plupart des tâches cognitives (Gagnon et al., 2009; Naismith et al., 2011; Sinforiani et al., 2006; Vendette et al., 2007).

### **b) Études longitudinales**

Les études prospectives révèlent qu'entre 75 et 90% des patients atteints de MP développeront une démence sur une période de plus de 10 ans au cours de l'évolution de la maladie (Aarsland et al., 2001; Buter et al., 2008; Hely et al., 2008; Hobson et Meara, 2015). De plus, le TCSP est présent chez 25 à 58% des patients atteints de la MP (Hu, 2020) et représente un facteur de risque important au développement de la démence dans la MP (Anang et al., 2014; Moore et Barker, 2014). Plusieurs études se sont penchées sur l'utilisation des tests cognitifs comme marqueurs de la démence dans la MP, cependant les résultats sont variables (De Roy et al., 2020). Ceci peut être expliqué par la taille de l'échantillon, le taux de suivi, l'absence de groupe contrôle, les tests cognitifs utilisés, ainsi que les critères diagnostiques de démence utilisés (De Roy et al., 2020).

La visualisation spatiale, mesurée par le sous-test Block Design, a été identifiée, à travers l'utilisation d'un modèle de régression de Cox comme un marqueur important du développement de la démence dans la MP (De Roy et al., 2020). D'autres études soutiennent également que des atteintes des habiletés visuospatiales sont de bons marqueurs du développement de la démence (Williams-Gray et al., 2009; Williams-Gray et al., 2007; Williams-Gray et al., 2013). Dans l'étude de Williams-Gray (2013), les auteurs ont suivi 180 patients atteints de la MP pendant 10 ans et ont constaté que l'un des tests cognitifs prédictifs de la conversion vers la démence (n=41) était une faible fluence verbale sémantique (<20 HR 3.05, p=0.005) ou une difficulté à copier deux pentagones enchevêtrés (HR 2.55, p=0.001), lié au fonctionnement des aires corticales postérieures. Ces résultats ont également été rapportés dans d'autres études (Broeders et al., 2013; Stern et al., 1993; Williams-Gray et al., 2007).

Toutefois, d'autres études ont identifié d'autres marqueurs prédictifs tels que les déficits en mémoire épisodique verbale et aux fonctions exécutives (De Roy et al., 2020; Levy et al., 2002; Pedersen et al., 2013). L'étude menée par Latreille (2015), a suivi 68 patients atteints de la MP pendant près de 4,5 ans, parmi lesquels 18 ont développé une démence tandis que 50 n'en ont pas développé. Au temps de base, les patients atteints de la MP qui ont développé une démence ont présenté une performance significativement plus faible aux tests de fonctions exécutives ( $Z=2.13 \pm 1.92$ ), de mémoire épisodique ( $Z=-0.99 \pm 0.67$ ) et de mesures visuospatiales composites ( $Z=-1.06 \pm 0.70$ ) que ceux qui n'ont pas développé de démence ou les sujets contrôles. D'après l'étude de De Roy et al. (2020), la partie B du Trail Making Test (temps) ainsi que la fluence verbale (sémantique) sont les tests cognitifs les plus performants pour identifier les patients atteints de la MP à risque de démence, avec une aire sous la courbe supérieure à 0,90. Cela pourrait s'expliquer par le fait que ces tests mobilisent une grande variété de ressources cognitives telles que les fonctions exécutives et motrices, les capacités visuospatiales et sémantiques, qui font appel à de vastes réseaux cérébraux (ex : frontales/exécutives, temporelles/sémantiques et pariétales/visuospatiales (Strauss et al., 2006). En résumé, la littérature scientifique sur la cognition dans la MP rapporte donc un profil cognitif hétérogène, compliquant l'identification de marqueurs cognitifs de la démence. Enfin, un facteur clinique correspondant à un facteur de risque important pour la démence dans cette population est le TCL (Aarsland, 2016; Litvan et al., 2012), qui a une

prévalence estimée à travers les études transversales entre 17 et 38% dans la MP (Dalrymple-Alford et al., 2011; Litvan et al., 2011).

Dans les études neuropsychologiques longitudinales, les algorithmes de ML ont souvent été utilisés pour différentes classifications cognitives de la maladie de Parkinson, sans avoir été testés pour leur capacité de généralisation (notamment via une méthode de validation croisée) : MP → démence (Anang et al., 2014; Hogue et al., 2018; Janvin et al., 2005; Jurcau et Nunkoo, 2021; Kim et al., 2019; Lichter et al., 2018), MP → déclin cognitif (ou TCL ou progression de la maladie) (Galtier et al., 2019; Hong et al., 2014; Kataoka et Sugie, 2021; Oikonomou et al., 2021; Tsiouris et al., 2017; Yu et al., 2020) et MP-TCL → démence (Cholerton et al., 2018; Galtier et al., 2016; Tajiri et al., 2020). Les algorithmes d'AM sans généralisation ont également été utilisés avec des données neuropsychologiques pour différentes classifications dans des études transversales : MP avec TCL (ou MP avec TCL amnésique) vs MP sans TCL (Brennan et al., 2017; Galtier et al., 2017; Isella et al., 2013; Yu et al., 2020), MP avec démence vs MP avec TCL (Brennan et al., 2017; Isella et al., 2013), MP vs MP avec démence (Dujardin et al., 2010; Llebaria et al., 2008), MP avec démence vs DCL (Bougea et al., 2022; Petrova et al., 2015), et DCL vs témoins (Ferman et al., 2006). Quelques-unes ont présenté des méthodes de sélection de variables robustes (ex : (Dujardin et al., 2010), et d'autres ont testé une variété d'algorithmes pour permettre la comparaison des performances (ex : (Bougea et al., 2022)). Dans l'ensemble, les résultats de ces études sont très variables, tout comme dans les études statistiques classiques présentées précédemment dans cette section. Les marqueurs identifiés sont principalement les tests mesurant les fonctions exécutives (Janvin et al., 2005), la mémoire verbale (Hogue et al., 2018) et le test de dépistage Montreal Cognitive Assessment (Kim et al., 2019).

#### **4. L'analyse quantitative de l'EEG dans les synucléinopathies**

L'analyse de l'EEG s'effectue en mesurant divers indices. Il est entre autres possible de mesurer des changements globaux des différentes activités oscillatoires composant l'EEG. Cela peut se faire par plusieurs méthodes, dont : l'analyse fréquentielle (ex. analyse de densité spectrale via la transformée de Fourier), l'analyse temps-fréquence (ex. via l'analyse en ondelettes), qui permet de décomposer le contenu du signal EEG à travers les dimensions temps et fréquences (Gonthier,



2013). L'analyse spectrale, fournit de l'information sur la puissance relative des hautes et basses fréquences, qui reflètent la synchronisation locale des assemblées neuronales (Hornero et al., 2008). L'analyse temps-fréquence, revient à appliquer l'analyse spectrale « à chaque instant », ce qui permet d'observer les variations du contenu fréquentiel de l'EEG au cours du temps (Nunez, 2006). Cela est typiquement implémenté via un calcul d'ondelettes ou de transformée d'Hilbert. La cohérence EEG, quant à elle, est utilisée pour évaluer la connectivité fonctionnelle corticale et quantifier les connexions cortico-corticales ou cortico- sous-corticales. La fonction de cohérence peut également être utilisée pour quantifier la corrélation linéaire et détecter la synchronisation linéaire entre deux canaux EEG (Dauwels et al., 2010; Escudero et al., 2011). Une baisse dans la cohérence est interprétée comme une réduction dans le « couplage » fonctionnel entre deux aires corticales (Rossini et al., 2007). La communication entre différentes populations de neurones se fait à travers cette synchronisation oscillatoire. Ainsi, un dysfonctionnement de ce système de réseau (c.à.d., synchronisation) peut également se refléter à travers des changements de l'activité électrique dans des zones spécifiques du cerveau (Fries, 2005). Certains suggèrent qu'il pourrait s'agir d'une caractéristique potentielle de la neurodégénérescence (Pievani et al., 2011).

Plusieurs études se sont penchées sur le ralentissement de l'EEG. Il est mesuré à travers différentes métriques, mais est principalement reflété par une augmentation de la puissance Delta ( $\pm 0,5-4$  Hz) ou Thêta (4–8 Hz), ou par une diminution de la puissance Alpha (8–13 Hz) ou Beta (13–  $\pm 30$  Hz) (Geraedts et al., 2018). D'autres mesures utilisées incluent : le rapport de puissance lent-rapide  $[(\text{delta} + \text{theta}) / (\text{beta1} + \text{beta2})]$ , reflétant le ralentissement cortical global; les changements de la fréquence occipitale dominante pendant l'éveil (c.à.d., alpha), calculé à travers la moyenne du pic dominant des spectres de puissance de quelques électrodes occipitales (Rodrigues Brazete et al., 2016); la puissance globale (%) relative calculée pour chaque bande de fréquence (avec plusieurs électrodes) et qui correspond au pourcentage de la puissance EEG totale entre 1,5 et 30 Hz (via la moyenne de la transformée de Fourier rapide).

L'EEG est un outil très prometteur pour la prédiction du développement futur de la démence (Hampel et al., 2014), en raison de sa facilité d'usage, de son aspect non-invasif et de son faible coût (Al-Qazzaz et al., 2014). De plus, il a été démontré que la quantification des rythmes EEG pourrait fournir un biomarqueur important pour plusieurs troubles neuropsychiatriques (Gandal et

al., 2012; Leuchter et al., 2009), et faciliterait l'identification des signes précoces de déclin cognitif dans des pathologies tels le TCL, la maladie d'Alzheimer, la DCL ou la MP (Bonanni et al., 2015; Klassen et al., 2011; Poil et al., 2013). Ceci est crucial pour les interventions précoces, ainsi que l'évaluation de l'efficacité des traitements développés pour ces maladies (Han et al., 2013). Il s'agira d'un des outils que nous examinerons dans le cadre de cette thèse.

#### **4.1 EEG chez le DCL**

La condition clinique de la DCL a été caractérisée par des anomalies EEG, telles qu'une augmentation de l'activité à ondes lentes dans les régions postérieures (temporales et occipitales) (Bonanni et al., 2016; Bonanni et al., 2015; Bonanni et al., 2008; McKeith et al., 2017). Effectivement, le rythme alpha dominant habituel semble être ralenti, adoptant un rythme pre-alpha/theta rapide et devenant plus variable à travers le temps (Bonanni et al., 2016; Bonanni et al., 2008; Stylianos et al., 2018). Ainsi, la fréquence pré-alpha domine, soit de façon stable, soit entremêlée d'activité alpha, theta ou delta (Bonanni et al., 2008). Il convient de noter que l'augmentation des bandes delta et theta a également été identifiée comme étant une caractéristique des patients atteints de la DCL qui permet principalement de les différencier de ceux atteints de la maladie d'Alzheimer (Kai et al., 2005).

#### **4.2 Anomalies EEG dans le TCSP**

##### **a) Études transversales**

La première étude montrant une activation corticale compromise chez les patients avec un TCSPi est celle de Fantini et al. (2003). Un ralentissement global de l'activité EEG à l'éveil a été démontré chez les patients ayant un TCSPi, et ce par une augmentation de la puissance theta aux régions frontales, temporales et occipitales et une diminution significative de la fréquence dominante (beta) occipitale (Fantini et al., 2003). Le theta/beta2 a été utilisé comme index de ralentissement cortical et permet de distinguer les TCSPi des contrôles, à partir d'analyse de variance (ANOVA) (Fantini et al., 2003). À noter, qu'une étude a révélé que ce ralentissement général est présent autant chez

les patients TCSP jeunes que âgés, quand comparés à des sujets contrôles (Sasai et al., 2013). Toutefois, d'autres études qui ont suivi celle de Fantini et al. (2003), ont plutôt montré une puissance absolue alpha plus élevée dans les régions frontales chez les TCSPi, en comparaison aux sujets contrôles (Bang et al., 2017). Ensuite, Brazète et al. (2013) ont plutôt démontré qu'un ralentissement de la fréquence dominante occipitale (c.à.d., alpha) était uniquement trouvé chez les patients ayant un TCSPi + TCL, en comparaison avec des sujets contrôles et non entre les TCSPi sans TCL et les sujets contrôles. À noter que les divergences de résultats entre les études pourraient être attribuable au statut cognitif des patients TCSPi étudiés (c.à.d., avec ou sans plaintes cognitives) (Ferini-Strambi et al., 2019).

Certaines études transversales ont montré une relation entre ce ralentissement EEG et la performance à des tâches neuropsychologiques d'attention, de fonctions exécutives et de mémoire verbale (Massicotte-Marquez et al., 2008). D'autres ont montré une association (stats) entre le ralentissement EEG et le score au MoCA par une corrélation négative entre le score au MoCA et l'augmentation de la puissance dans les bandes beta et theta (Sasai et al., 2013).

La connectivité fonctionnelle a récemment commencé à être investiguée dans l'identification de la progression clinique dans le TCSPi. Une étude a évalué la connectivité fonctionnelle chez les patients atteints de TCSPi à travers l'EEG à l'état de repos (Sunwoo et al., 2017). Une analyse de comparaison de différentes connexions fonctionnelles, à partir d'un test de permutation non-paramétrique a démontré que les TCSPi présentaient une baisse de la connectivité fonctionnelle (mesuré par le Weighted Phase Lag Index) uniquement pour la bande de fréquence delta entre les régions frontales et centropariétales. Par ailleurs, le Weighted Phase Lag Index entre les régions frontales et centrales était corrélé à la performance en fluence verbale (Sunwoo et al., 2017).

## **b) Études longitudinales**

Les études longitudinales ont identifié le ralentissement de l'EEG comme biomarqueur de la neurodégénérescence chez les patients atteints de TCSPi (Ferini-Strambi et al., 2019). L'étude de Brazète et al. (2016) rapporte par une analyse de corrélation qu'un ralentissement global de l'EEG pendant l'éveil chez les TCSPi est associé à l'apparition de la DCL (Rodrigues Brazete et al., 2016).

Ils ont suivi pendant trois ans 54 patients TCSPi. Au suivi, 28 ont développé une synucléinopathie (c.à.d., MP, DCL ou l'atrophie multisystématisée). Toutes les régions cérébrales ont été investiguées (frontale, centrale, pariétale, temporale, occipitale) pour les bandes de fréquences delta, theta, alpha, beta1 et beta2. Le ralentissement EEG a été objectivé à partir de mesures de puissances spectrales absolues et relatives, de fréquence dominante occipitale et de ratios lents à rapides  $[(\text{delta} + \text{theta})/(\text{beta1} + \text{beta2})]$ . Les patients qui ont développé une synucléinopathie ont présenté une puissance absolue delta et theta plus élevée dans toutes les régions corticales par rapport aux patients sans synucléinopathie et aux contrôles. Par ailleurs, la puissance relative des ondes theta dans les régions occipitales permettait de discriminer de manière précise les patients TCSPi qui convertiraient en une synucléinopathies des sujets sains. Le rapport de puissance lent à rapide était plus élevé dans toutes les régions chez les patients qui ont développé une synucléinopathie. De plus, ces patients avaient également une fréquence occipitale dominante plus lente par rapport aux témoins. Ensuite, ils identifient que les patients qui développent une MP au suivi se distinguent par un ralentissement de l'EEG limité aux régions postérieures (temporales et occipitales), alors que ceux qui développent une DCL présentent plutôt un ralentissement généralisé de l'EEG lors de la première visite. D'autres études ont identifié une relation entre le ralentissement EEG et le déclin cognitif à court-terme, principalement par l'étude du TCL dans le TCSPi. L'étude d'Iranzo et al. (2010) ont suivi 23 patients ayant un TCSPi, ainsi que 10 sujets contrôle pendant près de deux ans. À travers des tests de comparaison de groupe (test T et test  $\chi^2$ ), ils ont démontré que les patients atteints de TCSPi qui ont développé un TCL (n=10) au suivi présentaient au temps de base un ralentissement marqué de l'EEG (une activité élevée des ondes delta et theta) dans les régions centrales et occipitales pendant l'éveil et le sommeil paradoxal, en particulier dans l'hémisphère droit, par rapport aux témoins et par rapport aux patients qui sont demeurés idiopathiques (Iranzo et al., 2010).

Une étude transversale ayant comparé 23 patients TCSPi avec un TCL de 19 TCSPi sans TCL ont démontré une corrélation entre la présence du TCL et le ralentissement EEG, la décélération de l'EEG étant principalement observée dans les régions corticales postérieures (Rodrigues Brazete et al., 2013). Effectivement, un ratio lent à rapide plus élevé est retrouvé aux régions pariétales, temporales et occipitales chez les TCSP+TCL. Ces derniers présentent également puissance relative plus élevée en theta dans les régions pariétale, temporelle et occipitale

et une puissance relative plus faible en alpha dans la région occipitale par rapport aux patients TCSPi sans TCL et aux sujets contrôles (Rodrigues Brazete et al., 2013).

Ainsi, une activité anormale de l'EEG révèle un processus neurodégénératif avant même la présence de la détérioration cognitive (Ferini-Strambi et al., 2019). La connectivité fonctionnelle a récemment commencé à être considérée comme utile dans l'identification de la progression clinique dans le TCSPi. Une étude longitudinale récente a suivi 144 patients avec un TCSPi pendant 4 ans. Ils ont calculé la complexité globale de chaque sujet à partir de spectrogrammes EEG à l'état de repos (14 électrodes), générés à partir de segments sans artéfacts de 20 secondes. Ils ont révélé qu'une complexité plus faible mesurée lors de la première visite est associée à une phénoconversion au suivi (c.à.d., MP ou DCL). De plus, ils démontrent que les scores de complexité de bande complète sont plus discriminants que les analyses de bandes uniques ou encore les ratios spectral (c.à.d., ratio theta/beta2) (Ruffini et al., 2019).

#### **4.3 Anomalies EEG précédant le développement de la démence dans la MP**

L'EEG a été démontré comme un outil utile dans l'étude des marqueurs précoces du déclin cognitif dans la MP (de Weerd et al., 1990; Neufeld et al., 1994; Pezard et al., 2001; Stoffers et al., 2007; Tanaka et al., 2000). La revue (Geraedts et al., 2018) a examiné les corrélations entre les symptômes cognitifs de la MP et les métriques d'EEG quantitatifs, à travers 19 études transversales. Les études transversales révèlent que la sévérité des atteintes cognitives est corrélée au ralentissement EEG (Caviness et al., 2015; Caviness et al., 2007; Guner et al., 2017; He et al., 2017; Helkala et al., 1991; Kamei et al., 2010; Latreille et al., 2016; Morita et al., 2011; Mostile et al., 2015; Stanzione et al., 1996). Pour la bande de fréquence bêta, des résultats divergents ont été trouvés (Geraedts et al., 2018). D'autres mesures de ralentissement de l'EEG qui sont corrélées avec la cognition dans la maladie de Parkinson incluent le rapport spectral de la puissance de l'EEG rapide/lent (Eichelberger et al., 2017; Guner et al., 2017; Kamei et al., 2010; Latreille et al., 2016; Morita et al., 2011) et une fréquence dominante plus élevée (Caviness et al., 2007; Latreille et al., 2016; Soikkeli et al., 1991; Stanzione et al., 1996). Pour identifier la relation entre le ralentissement de l'EEG et le déclin cognitif, certaines des études transversales ont utilisé des comparaisons de groupe (ANOVA, ANCOVA, test T, test  $\chi^2$ ) (Fonseca et al., 2009; Guner et al.,

2017; Hassan et al., 2017; Helkala et al., 1991; Latreille et al., 2016; Neufeld et al., 1994; Pozzi et al., 1994; Soikkeli et al., 1991; Tanaka et al., 2000) ou des corrélations (Caviness et al., 2007; Eichelberger et al., 2017; Guner et al., 2017; He et al., 2017; Kamei et al., 2010; Latreille et al., 2016; Morita et al., 2011; Mostile et al., 2015; Stanzione et al., 1996; Tanaka et al., 2000), d'autres ont utilisé une régression sans généralisation (Caviness et al., 2015; Caviness et al., 2016; Helkala et al., 1991; Mostile et al., 2015; Stanzione et al., 1996; Tanaka et al., 2000), ou une régression logistique sans généralisation (Caviness et al., 2016; Kamei et al., 2010; Latreille et al., 2016).

La revue (Geraedts et al., 2018) a également examiné des études longitudinales, concluant que la diminution de la fréquence dominante (Caviness et al., 2015; Klassen et al., 2011; Latreille et al., 2016) et l'augmentation de la puissance theta (Caviness et al., 2015; Klassen et al., 2011), reflétant le ralentissement de l'EEG, seraient des biomarqueurs du déclin cognitif. Les études longitudinales ont utilisé soit une régression linéaire (Caviness et al., 2015; Klassen et al., 2011) ou une régression logistique (Latreille et al., 2016) sans méthodes pour valider la capacité de généralisation du modèle hors de cet échantillon. D'autres études se sont intéressées aux mesures de puissance globale relative et à l'activité de base (c.à.d., éveillée, à l'état de repos, les yeux fermés), suggérant qu'à l'EEG à l'éveil, une plus grande puissance relative dans les basses fréquences (delta ou theta) seraient des marqueurs du développement de la démence dans la MP (Caviness et al., 2015; Klassen et al., 2011). Plus spécifiquement, des corrélations ont été trouvées entre l'augmentation de la puissance delta globale et un déclin de performance à cinq tâches neuropsychologiques, avec la corrélation la plus élevée pour le test d'apprentissage verbal auditif de Rey (Caviness et al., 2015). L'étude de Klassen et al. (2011) montre un risque de développer la démence 13 fois plus élevé chez les personnes dont l'activité de base est faible (<8.5 Hz) plutôt qu'élevé ou présentant une puissance relative élevée dans la bande de fréquence theta (Klassen et al., 2011). Depuis 2018, la corrélation entre le ralentissement de l'EEG et la détérioration cognitive dans la MP a été répliquée (Yilmaz et al., 2020).

D'autres mesures EEG portant sur la communication cérébrale fourniraient des informations pertinentes et seraient sensibles aux altérations cérébrales menant à la démence. Par exemple, l'indice de décalage de phase (ou *Phase Lag Index*), le réseau pondéré (ou *weighted network*) et les arbres couvrants minimaux (ou *minimum spanning trees*) ont été utilisés pour

distinguer les patients ayant une MP avec démence de ceux ayant une cognition normale (Utianski, 2014). Les patients ayant une MP avec démence se distinguent entre autres par une complexité de signal plus élevée ou une valeur de verrouillage de phase (ou *Phase-Locking-Value*) plus faible dans les basses fréquences (Bertrand et al., 2016). Toutefois, d'autres études montrent que les mesures de réseau global (ou *global network*) basées sur la valeur de verrouillage de phase (Hassan et al., 2017), ainsi que l'asymétrie de signal (Mostile et al., 2015) ne se révèlent pas toujours corrélées aux résultats des tests cognitifs. D'autres marqueurs de détérioration cognitive comprennent une baisse de connectivité fonctionnelle, une perte de connectivité fronto-temporale (Hassan et al., 2017), une baisse de fréquence dominante et une augmentation de puissance theta (Geraedts et al., 2018). Les études présentées ont utilisé des méthodes d'analyses des statistiques classiques, ainsi que la régression linéaire (Mostile et al., 2015). D'autres études sont nécessaires pour clarifier les métriques de connectivité (Geraedts et al., 2018). Peu d'études ont abordé les corrélats EEG de la cognition dans la MP.

## **5. L'apprentissage machine (AM) dans un contexte clinique**

### **5.1 Les applications de l'AM dans le monde clinique**

Au cours des dernières années, l'essor fulgurant de l'AM en neurosciences s'est traduit par une multitude d'applications diverses (Belic et al., 2019; Myszczyńska et al., 2020; Tautan et al., 2021). Cette évolution a été largement propulsée par l'augmentation de la puissance de calcul et l'émergence de l'ère des données massives (Big Data) (Zhou, 2017). D'ores et déjà, l'AM apporte un soutien précieux au diagnostic médical, notamment grâce au traitement des signaux biomédicaux et à l'analyse d'images médicales (Magoulas, 1999). De plus en plus sollicitée dans le domaine médical, l'AM s'est révélée particulièrement efficace pour identifier des patrons généralisables dans les données soumises, fournissant des prédictions éclairées pour guider les décisions cliniques, telles que les choix de traitements (Ij, 2018), mais permettant également d'améliorer la compréhension de l'évolution de nombreuses maladies. Par exemple, l'AM est employé pour évaluer l'importance de différents paramètres cliniques et de leur combinaison pour le pronostic, tel que la prédiction de la progression d'une maladie (Magoulas, 1999). À ce sujet, des progrès considérables ont été réalisés, notamment dans l'étude des biomarqueurs liés à la maladie

d'Alzheimer (Chang et al., 2021; Li et al., 2021), l'épilepsie, la neuro-oncologie et la schizophrénie (English et al., 2022).

## 5.2 Statistiques classiques traditionnelles et apprentissage machine

La frontière entre l'inférence statistique et l'AM (ou la prédiction) est sujette à débat, puisque certaines méthodes tombent nettement dans l'un ou l'autre des domaines, mais beaucoup sont utilisées dans les deux. Effectivement, les statistiques classiques (SC) et l'AM se situent sur un continuum d'outils quantitatifs, présentant tous deux des méthodes qui peuvent être utilisées à la fois pour faire des prédictions et des inférences (Bzdok et Meyer-Lindenberg, 2018). On note toutefois une utilisation courante des tests de significativité contre une hypothèse nulle (ANOVA, test T,  $\chi^2$ , *false discovery rate* ou l'inférence bayésienne à postériori) lorsque l'objectif est l'inférence, permettant le calcul de valeurs P pour des variables cibles spécifiques. D'ailleurs, dans le cadre de cette thèse, nous utiliserons le terme « statistiques classiques » (SC) pour faire référence à ces analyses de tests d'hypothèses, aux analyses explicatives (*explanatory*) et autres analyses des statistiques inférentielles. Lorsqu'il s'agit de faire des prédictions en apprentissage automatique, divers outils sont couramment employés. Ceux-ci incluent les machines à vecteurs de support, les algorithmes de forêts aléatoires, d'autres méthodes d'apprentissage en ensemble (ou *ensemble learning*), les algorithmes d'apprentissage profond, ainsi que la régression linéaire ordinaire et la régression linéaire pénalisée (Bzdok et Ioannidis, 2019).

Ces utilisations divergentes sont entre autres reliées à leurs trajectoires historiques et au fait qu'ils sont ancrés dans des contextes mathématiques différents (Bzdok et Ioannidis, 2019). À travers l'histoire, les outils des SC ont été très axés sur l'inférence. La question de recherche précède souvent le choix du modèle statistique et le but est généralement de modéliser la boîte noire en faisant un ensemble d'hypothèses formelles sur son contenu. On veut vérifier les effets de variables « compréhensibles » (Bzdok et Ioannidis, 2019). Ainsi, les modèles reposent souvent sur des hypothèses de modélisation prédéterminées intégrant notre connaissance du système - les processus qui ont généré les données sont donc considérés comme partiellement connus. En AM, la visée est généralement de modéliser la sortie de la boîte noire (ex : distinguer des personnes en bonne santé de personnes schizophrènes) à partir de variables d'entrée (ex : mesures volumétriques



du cerveau) en faisant un minimum d'hypothèses possibles (Abu-Mostafa et al., 2012) pour évaluer à quelle précision un modèle construit peut prédire les résultats futurs (Bzdok, 2017; Bzdok et Meyer-Lindenberg, 2018) - le processus sous-tendant ou générant les données est considéré comme complexe, largement inconnu et en partie impossible à connaître. Ainsi, un avantage des SC est l'accès à un récit explicatif, alors qu'en AM, plusieurs algorithmes, entre autres d'apprentissage profond, sont plutôt qualifiés de « boîtes noires opaques ». Toutefois, la reconnaissance des relations complexes entre les variables est un domaine dans lequel de nombreux algorithmes d'AM excellent. Ils sont donc plus abordables pour traiter des données plus complexes (ex : lorsqu'un sujet présente plusieurs observations différentes pour différentes variables). D'autres avantages de l'AM incluent une meilleure capacité à traiter un grand nombre de variables (ex : variables > des sujets) et une efficacité en présence d'interactions non-linéaires (Bzdok et Meyer-Lindenberg, 2018).

Étant donnée les visées différentes des SC et de l'AM, les variables ont des rôles différents à jouer. L'inférence se concentre généralement sur l'isolement de variables considérées comme individuellement importantes au-dessus d'un certain niveau de chance, souvent basé sur les valeurs P, alors que la prédiction vise généralement à identifier des ensembles de variables qui permettent ensemble de deviner avec précision les résultats basés sur d'autres données ou des données futures (Breiman, 2001; Bzdok et Meyer-Lindenberg, 2018; Donoho, 2017). Ainsi, les variables qui s'avèrent être les plus discriminantes dans les tests d'hypothèses nulles peuvent être moins pertinentes pour maximiser la performance prédictive chez de nouveaux individus ou dans de nouveaux contextes (Bzdok et Ioannidis, 2019). Et à l'inverse, il est également possible de construire un modèle prédictif utile et précis en se basant sur des mesures qui sont peu susceptibles d'avoir une corrélation élevée avec le résultat d'intérêt (Bzdok et Ioannidis, 2019). **C'est ce qui nous intéresse dans le cadre de cette thèse : au-delà des mesures qui sont corrélées avec le résultat d'intérêt, quelles mesures permettent de construire un modèle prédictif précis ? Soit une question qui a une visée très clinique. Par ailleurs, il est important de préciser que le fait de se concentrer sur les prévisions cliniques peut compléter la quête visant à identifier les causes biologiques des maladies (Bzdok et Ioannidis, 2019) et une telle visée n'est pas en opposition ni ne sous-estime l'importance de l'effort à long terme visant la compréhension de ce qui sous-tend les pathologies cérébrales.**

Ensuite, on peut noter qu'avec l'AM, contrairement aux SC, il est possible de procéder à une évaluation pour extrapoler le modèle développé au-delà des données en main. En statistiques classiques le modèle n'est généralement pas évalué sur des données futures/inconnues (généralisation hors de l'échantillon d'entraînement), alors qu'il s'agit d'une pratique standard avec l'AM, qui se fait en deux étapes : 1) L'algorithme est entraîné/ajusté sur une grande quantité de données et 2) est testé sur des données indépendantes (Bzdok et Meyer-Lindenberg, 2018; Bzdok et Yeo, 2017). Avec les SC, les chercheurs visent à rejeter l'hypothèse nulle en considérant l'ensemble des données (Wasserstein, 2016). Par ailleurs, les p-values ou les tailles d'effet dans les tests d'hypothèses nuls ne peuvent pas être utilisés pour évaluer d'autres données dans une étape ultérieure (c.à.d., une p-value significative  $\neq$  une haute précision de prédiction sur de nouvelles données; les valeurs P ne mesurent pas l'exactitude prédictive d'un modèle) (Bzdok et Meyer-Lindenberg, 2018). La pertinence souhaitée d'une relation statistique dans la population est plutôt validée par des preuves mathématiques formelles et n'est pas généralement vérifiée par des évaluations explicites sur de nouvelles données (Breiman, 2001; Wasserstein, 2016).

Bzdok et al. (2019) nous éclairent sur les valeurs épistémiques différentes des SC et de l'AM : une compréhension mécanistique avec les SC (apporter de l'*insight*) et l'optimisation d'une précision prédictive avec l'AM (= modéliser le monde avec précision). Ils suggèrent de se demander laquelle de ces motivations est centrale dans notre analyse quantitative et ensuite de choisir les outils d'analyses y répondant.

À noter que tel que mentionné au début de cette section, il y a une zone grise entre l'inférence et la prédiction. Un bon exemple de ce fait est la régression linéaire, qui peut être utilisée autant pour faire de l'exploration (c.à.d., regarder les relations entre des variables; par corrélation de Pearson sans p value ou par régression linéaire), de l'inférence (c.à.d., discerner individuellement chaque variable qui contribue à la variable cible au-delà du niveau de la chance; tests d'hypothèse nulle, taux de fausses découvertes, etc.) ou de la prédiction (c.à.d., bâtir un modèle à tester sur des données indépendantes pour vérifier sa capacité à la capacité à estimer les variables cibles à partir de nouvelles données) (Bzdok et Ioannidis, 2019). Pareillement, plusieurs outils algorithmiques peuvent être utilisés pour explorer des données ou faire des inférences. Toutefois, bien que plusieurs outils puissent accomplir différents objectifs, certains sont plus

efficaces que d'autres pour répondre à certains objectifs, selon le cas en question. Par exemple, bien que les modèles de régression linéaire simples puissent fonctionner raisonnablement bien en termes de puissance prédictive, s'il s'agit de maximiser la précision de la prédiction, il est souvent bénéfique d'exploiter les associations complexes et non additives dans les données, par exemple à travers des algorithmes de réseaux de neurones profonds (Bzdok et Ioannidis, 2019).

Les modèles d'AM complètent les avantages établis des SC en neurosciences. De plus, les modèles d'AM ont augmenté le potentiel de la recherche clinique qui vise à améliorer le diagnostic et suivre la trajectoire de progression des maladies neurodégénératives, favorisant ainsi la découverte de nouveaux biomarqueurs (Khatri et Kwon, 2020; Trambaiolli et al., 2011). Cela est particulièrement important, étant donné que les mécanismes physiopathologiques, la présentation clinique et la progression de la plupart des maladies neurodégénératives sont encore inconnus, entravant ainsi le développement d'interventions neuroprotectrices. Dans le cadre de cette thèse, nous nous intéresserons aux outils qui permettent au mieux d'optimiser la précision prédictive - soit les algorithmes d'AM. Toujours dans une visée clinique, la rigueur dans les méthodes de validation de ces modèles algorithmiques nous sera importante, puisque cela favorise la généralisation de leurs résultats sur des données futures inconnues. Ces techniques de validations de modèle seront expliquées dans une des prochaines sections de cette thèse.

### **5.3 L'apprentissage machine – supervisé**

L'AM a été défini par Mitchell (1997) comme suit : « Un programme informatique apprend d'une expérience E en rapport à une tâche T et d'une mesure de performance P si sa performance à la tâche T, telle que mesurée par P, augmente avec l'expérience E » ((Mitchell, 1997); traduction libre). L'AM permet de découvrir des principes généraux sous-tendant un ensemble d'observations sans instructions explicites (Breiman, 2001; Bzdok et Meyer-Lindenberg, 2018; Goodfellow, 2016). Les algorithmes d'AM sont utilisés pour des tâches de classification, recevant typiquement en entrée des caractéristiques discrètes et/ou continues (c'est-à-dire des « variables indépendantes »; ex : âge, éducation) et produisant en sortie une classe (c'est-à-dire la « variable dépendante »; ex : patient dément ou non-dément). Une tâche de classification peut se faire en utilisant deux types d'algorithmes différents, soit les deux catégories d'algorithmes en AM : l'apprentissage supervisé

et non supervisé (Mirzaei et al., 2016). L'apprentissage supervisé correspond aux algorithmes qui identifient des règles/schémas et les généralisent, à partir d'un ensemble de données ou d'attributs (ex. enregistrement en EEG) en ayant les étiquettes correspondantes (ex : un diagnostic). Ainsi, l'algorithme s'entraîne d'abord sur les données d'apprentissage étiquetées (échantillon d'entraînement), pour trouver la meilleure manière (c.à.d., fonction) de prédire l'étiquette. Ensuite, sa capacité à généraliser sera mise à l'épreuve en testant sa capacité à prédire les étiquettes, à partir de données *naïves* jamais utilisées par l'algorithme auparavant (échantillon test), soit un nouvel échantillon de données sans étiquettes ; la comparaison des étiquettes prédites par l'algorithme aux vraies étiquettes de ces nouvelles données détermine la performance de classification (Kotsiantis, 2007; Mohri et al., 2012). En résumé, l'apprentissage supervisé consiste en des algorithmes qui apprennent une fonction à partir d'une entrée vers une sortie, en se basant sur des données étiquetées (entrées + variables cibles). En revanche, pour les algorithmes d'apprentissage non supervisé, il n'y a pas de "bonne réponse", car ils découvrent une structure dans les variables à travers les sujets qui eux n'ont pas de variable cible/étiquette (Mahesh, 2020).

## **5.4 Les algorithmes**

### **a) Différents types**

Parmi les algorithmes d'apprentissage supervisé on retrouve l'analyse discriminante linéaire, les arbres de décision, les Bayésiens naïfs, ou encore les machines à vecteurs de support. Les algorithmes d'apprentissage non supervisé incluent l'analyse par composantes principales, les K-Means, et l'analyse de clustering. Il existe également d'autres catégories d'algorithmes d'AM, comme l'apprentissage par renforcement (c.à.d., non supervisé, mais qui reçoit un feedback appelé "récompense") et l'apprentissage profond (c.à.d., les réseaux neuronaux multicouches et réseaux de neurones profonds) (Mahesh, 2020). Nous présentons ci-bas un aperçu succinct des méthodes les plus fréquemment employées dans ce domaine, notamment l'analyse discriminante linéaire, la machine à vecteurs de support, la Régression Logistique et les réseaux neuronaux.

L'algorithme de l'analyse discriminante linéaire est un algorithme linéaire qui tente de trouver une projection d'hyperplan qui maximise la distance moyenne entre deux classes, tout en minimisant la variance interclasse (Fisher, 1936). Le principal inconvénient de cet algorithme est

sa linéarité, qui peut être problématique pour des données non-linéaires complexes (Lotte et al., 2007). La machine à vecteurs de support effectue une classification linéaire, même avec des classes non linéairement séparables, en trouvant un hyperplan (ou un ensemble) dans un espace de plus haute dimension à l'aide de fonctions de noyau (astuce du noyau) (Burges, 1998). Lorsqu'il tente de séparer les classes, il maximise spécifiquement la distance par rapport au point de l'échantillon le plus proche à l'aide de vecteurs de support (Bhardwaj et al., 2016). La régression logistique est utilisée pour prédire la probabilité d'une variable cible en ajustant une courbe logistique à l'ensemble de données (Xanthopoulos, 2013). Les réseaux de neurones profonds est un réseau neuronal (imitant la microarchitecture du cerveau) qui tente de trouver des relations sous-jacentes dans les données en utilisant plusieurs couches de non-linéarités (Mahesh, 2020). Plus il y a de couches cachées incluses, plus le niveau d'apprentissage des représentations abstraites d'une entrée est élevé (Arisoy, 2012, June).

## **b) Arbres de décision – vers le Random Forest**

Les arbres de décision sont une des méthodes les plus intuitives et populaires de la fouille de données (*data mining*), d'autant plus qu'ils supportent bien les données hétérogènes (qualitatives et quantitatives) et les effets non linéaires entre les variables explicatives et les variables à expliquer (Vemuri, 2020). Il existe plusieurs arbres de décisions, développés par Léo Breiman : l'arbre CART (Breiman, 1984), l'arbre CART élagué (Breiman 1984) et la forêt aléatoire (Breiman, 2001). Un arbre de décision comme l'arbre CART est assimilable à une fonction  $f$  qui prend comme valeurs d'entrée  $k$  variables (ex : longueur et largeur des pétales) et comme valeur de sortie la prédiction de la variable explicative (ex : déterminer si la fleur est une iris de l'espèce *Viricolor* ou *Setosa*; soit un problème de classification à deux classes). Le nœud de départ s'appelle la « racine » de l'arbre et les nœuds terminaux sont les « feuilles » de l'arbre. L'arbre se construit de façon récursive, c'est-à-dire qu'il partitionne les éléments d'un nœud en deux nœuds descendants, de telle sorte que ces derniers soient le plus homogène possible. L'hétérogénéité d'un nœud peut être mesurée par l'indice de Gini ou encore l'entropie de Shannon. Ainsi, à chaque scission d'un nœud en particulier, la variable explicative qui maximisera la variation d'impureté sera sélectionnée. Quant à la forêt aléatoire, son mécanisme consiste à introduire du hasard dans la construction de plusieurs arbres CART (dans le choix de l'échantillon et/ou des variables

explicatives) qui constitueront la forêt et à ensuite agréger ces arbres par la moyenne des valeurs prédites ou par la majorité de la classe prédite (c.à.d., aller avec le vote majoritaire des arbres) (Singh, 2016, March).

Un des objectifs fondamentaux de l'apprentissage automatique est la volonté de construire de « meilleurs » classifieurs (Vemuri, 2020). Plusieurs éléments permettent de qualifier la performance d'un classifieur, dont sa fiabilité (ex : son erreur de généralisation) et sa complexité (ex : temps d'exécution, le réglage de ses paramètres et lisibilité des résultats). Par ailleurs, différents facteurs peuvent affecter ces mesures de performance. D'abord, la taille de l'échantillon accessible, soit un enjeu réel dans les milieux de recherche cliniques, peut affecter le biais et la variance. Par exemple, pour l'arbre CART et la forêt aléatoire, le biais correspond à l'ajustement de l'arbre (l'erreur d'ajustement), alors que la variance est la capacité de généralisation (l'erreur de généralisation). Lorsqu'on compare l'effet de différentes tailles d'échantillons sur les performances de ces algorithmes, en analysant l'évolution de cette erreur lorsque l'algorithme est testé sur l'échantillon sur lequel il s'est construit (courbe de resubstitution; « erreur théorique »), puis lorsqu'il est testé sur toute la population (« erreur réelle »), deux phénomènes importants sont observés : le sur-apprentissage (*overfitting*) et le sous-apprentissage. Commençons par l'arbre CART. On remarque que la courbe de resubstitution va toujours diminuer en erreur (erreur théorique), ce qui n'est pas toujours le cas pour la courbe sur la population (erreur réelle). À travers les différentes tailles d'échantillon, on remarque un grand écart entre ces deux courbes : lorsque l'échantillon est très petit, on observe un **sous-apprentissage** (c.à.d., le modèle n'a pas beaucoup appris et a donc un pouvoir prédictif faible, soit un grand écart avec l'erreur réelle) et lorsque l'échantillon est plus grand, cet écart entre les deux courbes reflète plutôt un **sur-apprentissage**, signifiant que le modèle est trop collé aux données d'entraînement et se généralise moins bien à la population. Ainsi, on voit un fort biais optimiste avec le modèle CART, c'est-à-dire qu'en moyenne, l'erreur de resubstitution sous-estime l'erreur réelle (« dit en moyenne qu'elle est plus basse que la vraie »). Toutefois, l'introduction de l'élagage à l'arbre CART (c.à.d., réduire sa profondeur) permet de résoudre en partie le problème de sur-apprentissage, c'est-à-dire qu'il réduit l'écart entre les deux courbes, mais seulement pour les échantillons de grande taille. Ainsi, pour l'arbre CART élagué, l'estimateur de l'erreur par resubstitution présente un très faible biais d'optimisme sur les échantillons de grande taille. Maintenant, si on revient à la forêt aléatoire, il

est démontré que cet algorithme permet de mieux contrer les phénomènes de sous-apprentissage et de sur-apprentissage, sur de grands échantillons, mais également sur de petits échantillons. En résumé, l'objectif de la forêt aléatoire est d'agrèger un ensemble d'arbres CART (= prédicteurs plutôt faibles) pour en faire un prédicteur fort (« l'union fait la force ») présentant un meilleur compromis biais-variance.

D'autres éléments sont à considérer dans le choix et la modélisation d'un algorithme d'AM. Un arbre de décision, tout comme d'autres algorithmes possèdent des paramètres, qu'on appelle les « hyperparamètres ». Il s'agit de caractéristiques de l'algorithme qui peuvent être optimisés pour améliorer la performance de prédiction. Par exemple, pour un arbre CART, il peut s'agir des critères d'arrêt de l'élaboration de l'arbre (ex : pureté du nœud, nombre d'effectif minimal dans le nœud, taille maximale de l'arbre et profondeur de l'arbre, car un arbre trop profond peut sur-ajuster les données à partir desquelles il a été construit et mal s'appliquer à de nouvelles données). Enfin, la complexité et transparence d'un modèle d'AM est importante. Plus tôt dans cette thèse, nous avons abordé des différences entre les modèles plus efficaces pour faire de la prédiction vs de l'inférence. Les considérations présentées plus haut à ce sujet encouragent à faire des compromis entre la transparence du modèle pour une interprétabilité facile et la complexité du modèle qui permettrait de prédire des relations particulièrement complexes (Bzdok et Ioannidis, 2019). Bzdok et al., (2019) analysent les différents algorithmes, à travers un graphique à deux axes : 1) la transparence du modèle (c.à.d., qui permet une compréhension scientifique) et 2) la capacité théorique du modèle à permettre des prédictions sophistiquées. L'algorithme de la forêt aléatoire se retrouve tout juste au mieux de ces deux axes, étant assez transparent tout en ayant un bon niveau de complexité. Ainsi, cet algorithme présente de nombreux avantages, ce qui le rend très attrayant pour cette thèse. Prenons l'exemple de la génétique pour illustrer l'interprétabilité offerte par le RF. Si nous essayons de prédire un phénotype à partir de plusieurs gènes, le classifieur de forêt aléatoire prendra en compte simultanément tous les gènes et fera croître plusieurs arbres (ex : 100 arbres) de décision pour prédire le phénotype et nous permettra de connaître l'importance des variables/caractéristiques (gènes) sélectionnés et utilisés dans sa construction. Ce score d'« importance » quantifie la contribution d'un gène donné à l'amélioration moyenne de la classification au sein d'une partition lorsque l'arbre est divisé en sélectionnant ce gène.

## 5.5 Nécessaire d'évaluer la généralisation en AM: *cross-validation, out-of-sample generalization*

Les approches prédictives sont de plus en plus adoptées et recommandées. Cependant, plusieurs exigences doivent être remplies pour qu'elles puissent être considérées comme adaptées à une application à grande échelle dans des environnements cliniques réels. L'extrapolation des résultats des approches prédictives sur de nouvelles données en pratique clinique se fait à travers différentes techniques de validation de modèle hors échantillon.

Cela se fait en deux étapes. D'abord, l'algorithme s'entraîne et tente d'identifier un certain modèle (il est "ajusté" ou *fitted*) sur une partie des données disponibles (c'est-à-dire les données d'entraînement) (Bzdok et Meyer-Lindenberg, 2018; Domingos, 2012). Ensuite, ce modèle "déjà ajusté" est utilisé pour la prédiction sur un sous-ensemble distinct des données disponibles (c.à.d., les « données de test » ; les données qui n'ont pas été utilisées pour ajuster le modèle) (Bzdok et Meyer-Lindenberg, 2018; Domingos, 2012). Il s'agit d'une "validation interne", si cela est effectuée à l'aide d'une technique de rééchantillonnage/division d'échantillon en utilisant l'échantillon original qui a servi à la construction du modèle (Steyerberg et Harrell, 2016). Il s'agit de "validation externe" si le modèle (c.à.d., les associations prédictives identifiées à partir de l'échantillon original) est testé sur un échantillon totalement distinct recueilli à un moment ou dans un lieu différent (Austin et al., 2016; Djulbegovic et Ioannidis, 2019; Manrai et al., 2018). La validation externe est plus robuste (Bzdok et Ioannidis, 2019; Fusar-Poli et al., 2018), le test de généralisation étant plus fort lorsque les prédictions du modèle sont évaluées sur de nouveaux individus provenant d'un échantillon distinct de l'échantillon de sujets original (Naci et Ioannidis, 2015; Steyerberg et Harrell, 2016). Par ailleurs, qu'il s'agisse de validation interne ou externe, lorsque le modèle sera testé sur les « données test », la précision des prédictions sera généralement inférieure à celle des étapes précédentes (Bzdok et Ioannidis, 2019).

Une méthode standard de validation de modèle est la procédure de validation croisée (ou *cross-validation*). La validation croisée consiste à diviser les données en un ensemble d'entraînement (pour ajuster le modèle) et de test (hors échantillon pour évaluer les performances). Ce processus est répété pour différentes divisions des données. S'il est effectué 5 ou 10 fois, on



parle de validation croisée 5-plis (ou *5-fold*) ou 10-plis, respectivement (Bzdok et Meyer-Lindenberg, 2018). La précision de prédiction moyenne à travers tous les ensembles de test fournit une estimation de la performance de classification hors échantillon du modèle (Domingos, 2012). De plus, en agrégeant par la moyenne les erreurs estimées sur les échantillons test on obtient l'estimateur de l'erreur (biais) et sa variance. L'étude de l'évolution du biais et de la variance dans l'estimation de l'erreur avec l'arbre CART a révélé que plus le nombre de plis ( $k$ ) est élevé, plus le biais et la variance du modèle en sont réduits de façon importante (Adler, 2010; Rouvière, 2010). Ainsi, le nombre de plis d'une procédure de validation croisée peut être sélectionné en fonction de l'algorithme et des données en main. À noter que lorsqu'on souhaite également optimiser les hyperparamètres du modèle, la méthode de validation croisée imbriquée (ou *nested*) est souvent utilisée. Il y a alors une double boucle, deux boucles indépendantes, une boucle externe (qui servira à évaluer la performance de classification hors échantillon du modèle) et une boucle interne (qui servira à construire le modèle et sélectionner ses hyperparamètres) (Cawley, 2010).

## 5.6 Métriques et performance de classification

Prenons l'exemple d'une classification binaire où on tente de prédire si un sujet appartient à la classe « démence » ou « pas de démence ». Plusieurs types de prédictions et d'erreurs qui peuvent être faites : faux-positif (FP), faux-négatif (FN), vrai-positif (VP), vrai-négatif (VN). Ces mesures d'erreur sont souvent utilisées pour calculer les métriques de performance. Une multitude de métriques peuvent être utilisées pour mesurer la performance des modèles d'AM sur leur capacité à prédire une classe. Les plus utilisées dans le domaine sont l'aire sous la courbe (AUC;  $\int_0^1 ROC$  qui montre le taux de FP et le taux de VP pour différents seuils de probabilité), la justesse (*accuracy*;  $\frac{|VP+VN|}{|VP+VN+FP+FN|} \times 100\%$ , le nombre de cas de test correctement classés par rapport au nombre absolu de cas de test), la sensibilité ( $\frac{|VP|}{|VP+FN|} \times 100\%$ , le *taux de VP*) ou la spécificité ( $\frac{|VN|}{|VN+FP|} \times 100\%$ , le *taux de vrais négatifs*). À noter que plusieurs études utilisent la justesse ou encore l'AUC, mais une étude de notre laboratoire montre que ces métriques mesurent des aspects distincts de la performance d'un classifieur et, par conséquent, ne peuvent pas être comparées directement, car l'AUC sera légèrement plus élevée que la justesse lorsque les données sont

équilibrées dans un problème de classification, tandis que la justesse surestimera les performances lorsque les données sont déséquilibrées (Tholke et al., 2023).

### **5.7 Et la chance dans tout ça?**

Une fois qu'on a une performance de classification, il faut vérifier sa significativité, car pour un problème de classification à deux classes, si l'échantillon est petit, une performance de classification de 60-70% pourrait théoriquement provenir du hasard ! Cette vérification peut se faire avec la loi binomiale en vérifiant la performance de classification minimale nécessaire en fonction de la taille de l'échantillon utilisé (voir (Combrisson et Jerbi, 2015)), ou encore à travers un test de permutation. Le test de permutation consiste à comparer la « vrai » performance de classification (c.à.d., sur les données bien étiquetées ; ex : démence/pas de démence) aux performances de classification d'un autre modèle construit à partir de données permutées (c.à.d., sur les données mal étiquetées/permutées ; les étiquettes démence/pas démence sont mélangées aléatoirement). Cette comparaison permettra de déterminer si la performance de classification provient du hasard ou pas. Il s'agit d'une méthode non-paramétrique qui ne fait aucune hypothèse sur la distribution des données (Cummins et al., 2004).

### **5.8 Méthodes de sélection de variables**

Dans de nombreux cas, certaines variables d'entrée peuvent être non pertinentes à une classification efficace. Par conséquent, la sélection des variables/attributs (ou *feature selection*) est souvent effectuée en tant qu'étape de prétraitement des données avant la classification en AM (ou comme méthode pour déduire un modèle plus efficace pour utilisation sur de nouvelles données - à partir des résultats de classification sur un premier ensemble de données). La "sélection d'attributs" est une technique d'AM qui améliore les performances de classification en éliminant les attributs jugés moins pertinents de l'analyse, réduisant ainsi les risques de sur-apprentissage ainsi que la dimensionnalité des données (Elkhani et Muniyandi, 2017). Elle aide également à mieux comprendre le modèle d'apprentissage ou les données (Cai et al., 2018).

La sélection des attributs est une étape cruciale dans le processus de l'apprentissage automatique. Elle permet de choisir les attributs les plus pertinents pour le modèle à construire, en évitant ceux qui sont redondants ou peu informatifs. Il existe plusieurs méthodes pour faire la sélection des attributs :

***Filtre (Filter methods)*** : Ces méthodes évaluent chaque attribut individuellement et le classent en fonction de ses statistiques. Les méthodes de filtre n'utilisent pas le modèle de classification final, mais utilisent d'autres mesures comme le coefficient de corrélation de Pearson, le test du chi carré (Rahman et al., 2020). Les attributs qui n'atteignent pas un certain seuil sont éliminés.

***Emballage (Wrapper methods)*** : Ces méthodes utilisent le modèle de classification final pour évaluer l'utilité d'un ensemble d'attributs. Elles construisent une série de modèles à partir de différentes combinaisons d'attributs et sélectionnent ceux qui donnent les meilleurs résultats. Les exemples courants de ces méthodes sont la sélection séquentielle en avant (ou *forward selection*), la sélection séquentielle arrière (ou *backward elimination*) et la méthode exhaustive.

***Méthodes intégrées (Embedded methods)*** : Ces méthodes intègrent la sélection des attributs dans le processus de formation du modèle. Par exemple, les méthodes de régularisation comme la régression Lasso ou Ridge sont des exemples courants de méthodes intégrées. Ces méthodes ajoutent un terme de pénalité à la fonction de coût du modèle pour réduire le nombre d'attributs.

***Méthodes basées sur l'apprentissage ensembliste (Ensemble methods)*** : Ces méthodes, telles que les forêts aléatoires ou le boosting de gradient, effectuent implicitement une sélection de fonctionnalités en attribuant des poids ou des importances aux fonctionnalités. Les fonctionnalités avec des poids plus élevés sont considérées comme plus importantes.

## 6. Objectifs et hypothèses

L'objectif général de cette thèse est d'étudier l'apport et d'évaluer l'efficacité de l'AM avec validation croisée pour prédire le développement de la démence dans les synucléinopathies.

Dans le premier article de cette thèse, une revue systématique de la littérature a été conduite pour examiner les recherches transversales et longitudinales utilisant l'AM avec validation croisée, à partir de données EEG au repos ou de données cognitives pour classifier/prédire le statut cognitif dans les synucléinopathies (MP, DCL, TCSPi).

Les objectifs spécifiques sont :

- 1) Effectuer une recherche exhaustive de la littérature, afin d'évaluer le potentiel de l'AM pour la prédiction du statut cognitif (ou son évolution) dans les synucléinopathies, par une analyse qualitative approfondie des méthodologies utilisées et des performances obtenues pour ces classifications à travers différentes métriques.
- 2) Identifier les variables les plus importantes pour optimiser la prédiction dans les différentes classifications du statut cognitif des synucléinopathies, telles que sélectionnées par les modèles d'AM.
- 3) Évaluer quantitativement les types d'algorithmes qui montrent une performance supérieure en fonction de la modalité et de la métrique utilisée.

Dans le deuxième article de cette thèse, un algorithme d'AM, la forêt aléatoire, a été entraîné sur des données sociodémographiques, cognitives et cliniques de patients ayant un TCSPi, avec et sans TCL, recueillies en moyenne 5,6 ans avant leur dernier suivi. Certains de ces patients ont par la suite développé une synucléinopathie (MP ou DCL), tandis que d'autres sont restés à la phase du TCSPi (n'ont pas encore développé une maladie neurodégénérative).

Les objectifs spécifiques sont :

- 1) Déterminer si la classification par la forêt aléatoire avec validation de modèle distinguera de façon significative les patients avec un TCSPi qui progresseront vers une démence, de ceux qui :
  1. resteront idiopathiques ;
  2. développeront une MP ;
  3. sont des sujets sains (contrôles) sans TCSPi ni trouble cognitif.
- 2) Identifier quelle combinaison d'attributs cognitifs, sociodémographiques et cliniques permettent les meilleures précisions de prédiction quant aux patients avec un TCSP qui progresseront vers une démence, de ceux qui
  1. resteront idiopathiques ;
  2. développeront une MP ;
  3. sont des sujets sains (contrôles) sans TCSPi ni trouble cognitif.

Les prédictions sont que :

- 1) La classification distinguera de manière significative et avec une  $AUC > 80\%$  les patients avec un TCSP qui progresseront vers une démence, de ceux qui
  - a. resteront idiopathiques ;
  - b. développeront une MP ;
  - c. sont des sujets sains (contrôles) sans TCSPi ni trouble cognitif.
- 2) Pour l'ensemble des classifications, la combinaison de données issues des tests cognitifs et des mesures cliniques optimisera la performance de classification (AUC).

Avec la résurgence de l'AM, ainsi que l'intérêt récent porté à l'utilisation de ces outils dans l'investigation des synucléinopathies, notre première étude offrira une mise à jour nécessaire, tout en introduisant les cliniciens non-initiés à l'AM. Étant donnée l'orientation clinique de cette thèse, qui transparait par les méthodes de validation de modèles (pour la généralisation des résultats) et l'utilisation d'outils non-invasifs et accessibles, notre première étude devrait d'autant plus révéler un nombre restreint de recherches répondant à ces conditions spécifiques, mais importantes. L'article 1, le relevé systématique de la littérature, permettra donc pour la première fois de mettre en lumière les manquements (ex : le nombre limité d'études, les classifications avec nombre restreint d'études et les variables limitées), ainsi que le potentiel des algorithmes d'AM dans la prédiction du statut cognitif dans les synucléinopathies – nous prévoyons trouver des performances significatives (et performances  $>80$  AUC ou *accuracy*) dans les études observées. Ainsi, cette thèse permettra de favoriser le début d'un transfert de connaissances, de la théorie à la pratique, en soulignant la rigueur nécessaire pour y arriver et outillant les cliniciens souhaitant contribuer à ce champ de recherche.

Par ailleurs, l'article 2 devrait montrer qu'il est possible d'utiliser une large base de données, incluant plusieurs données cognitives, avec des outils d'AM pour prédire l'évolution vers la démence dans le TCSPi et de distinguer cette trajectoire des autres conditions cliniques possibles. Cette étude permettra également de valider que les variables qui ont démontré individuellement une importance dans un contexte d'inférence n'auront pas nécessairement d'importance dans un modèle d'AM, dans le cadre d'interactions complexes avec d'autres variables dans une visée d'optimisation de la prédiction. Cette thèse reconnaît la variabilité de la présentation clinique et de la progression des symptômes de ces maladies et souligne que c'est cette complexité qui peut être capturée par des algorithmes d'AM – elle contribuera à l'élaboration de modèles performants, pouvant ouvrir la voie à l'identification des individus avec un TCSPi à haut risque de développer une démence. Le dépistage précoce de cette condition permettra une prise en charge plus adaptée et personnalisée des personnes atteintes et facilitera le développement d'approches thérapeutiques visant à ralentir la neurodégénérescence. Ceci pourrait également permettre d'aider à mieux comprendre les mécanismes déclencheurs de cette neurodégénérescence, qui demeurent à ce jour incompris.

**Chapitre 2 – Article 1 : Prediction of cognitive decline in  
synucleinopathies from EEG and cognitive data: A review of  
machine learning approaches**

## Article 1

### **Prediction of cognitive decline in synucleinopathies from EEG and cognitive data: A review of machine learning approaches**

Loubna Mekki Berrada<sup>1</sup>, Golnoush Alamian<sup>2</sup>, Jean-François Gagnon<sup>\*3,4</sup>, Karim Jerbi<sup>\*1,5</sup>

<sup>1</sup>Cognitive and Computational Neuroscience Lab, Psychology Department, University of Montreal, QC, Canada

<sup>2</sup>Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

<sup>3</sup>Department of Psychology, Université du Québec à Montréal, Montreal, QC, Canada

<sup>4</sup>Centre for Advanced Research in Sleep Medicine, Centre intégré universitaire de santé et de service sociaux de Nord-de-l'Île-de-Montréal – Hôpital du Sacré-Coeur de Montréal, Montreal, QC, Canada

<sup>5</sup>MILA (Quebec Artificial Intelligence Institute), Montreal, QC, Canada

\*These authors share senior authorship.



## Abstract

Dementia with Lewy bodies (DLB) and Parkinson's disease (PD) with dementia (PDD) are two synucleinopathies characterized by a prodromal phase during which cognitive decline progresses insidiously. Numerous studies have investigated markers of disease progression to dementia in PD and idiopathic REM sleep behavior disorder (iRBD), a prodromal stage of DLB, using classical statistical approaches. However, the rise of machine learning (ML) applications in virtually all scientific domains and clinical practice suggests that it might help research improve upon the predictors of cognitive decline in these pathologies.

This systematic review examines the contribution of studies that have applied model-validated ML algorithms to accessible resting state electroencephalography (EEG) and cognitive data to classify different synucleinopathy conditions. A selection of thirty-three studies based on inclusion criteria reveals various classifications, mainly using EEG data (59%). The main categories include 1) PD vs controls, 2) DLB vs controls, 3) PD vs PD with dementia or DLB, 4) PD vs PD with prodromal state, 5) PD leading to cognitive decline or dementia, and 6) iRBD evolving into PD or DLB, with most research focusing on distinguishing PD from controls.

The findings reveal significant challenges in precise trend extraction and analysis. The main difficulty lies in the diversity of binary classifications studied, as well as the variety of modalities and variables employed. However, we note that EEG features often emerge as crucial for various multimodal classifications, with specific features improving predictive performance in PD diagnosis (PD vs controls) - notably spectral characteristics across different brain regions, inter-frequency connectivity, and complexity measures – shedding light on the interest of neural networks.

Our review also indicates the value of integrating various data modalities - cognitive, clinical and demographic data - in classifying cognitive states in PD (MCI or dementia). Cognitive measures are consistently identified as key in optimizing these multivariate ML models, just after

EEG data. Screening test's predictive performance is often improved when combined to cognitive tests.

Regarding ML algorithms, ten categories are predominantly used, with Random Forest (RF) and Support Vector Machine (SVM) leading. All reviewed algorithms show significant performance, highlighting the importance of considering publication bias and the metrics used. Notably, a fifth of the studies employ accuracy metrics with imbalanced data, which may affect the generalizability of these findings. We discuss the limitations of directly comparing algorithmic performances across studies due to these issues. However, the analysis of within-study comparisons reveals that no single algorithm consistently outperformed the others across classifications.

This exhaustive literature review study uniquely maps the continuum of ML models identified throughout the progression of iRBD and PD, linking them to findings from previous non-ML research. It also discusses methodological limitations and future directions for ML in discriminating and prognosticating synucleinopathies.

**Keywords:** Synucleinopathies, Parkinson Disease, Lewy Body Disease, Machine Learning, Electroencephalography, Neuropsychological Tests.

## 1. INTRODUCTION

### *Problematic*

Dementia (or major neurocognitive disorder) is an extremely debilitating condition, affecting approximately 46.8 million individuals globally, including 1.7 million across the Americas. Annually, it is estimated that more than 9.9 million new instances of dementia emerge worldwide [1]. Dementia with Lewy bodies (DLB), a synucleinopathy, is the second most common form of degenerative dementia, accounting for 10 to 30% of all dementia cases [2, 3]. Parkinson's disease (PD), another synucleinopathy, affects 2% of the population aged 65 years and over [4]. Dementia reaches as high as 90% of PD patients (PDD) over the course of the disease [5]. Several studies have attempted to predict the onset of dementia in synucleinopathies to better direct patient care and preventive interventions [6, 7]. These studies have in large part used classical statistical approaches with some success, yet clear prognostic biomarkers are still lacking, invoking the need for new quantitative approaches in predicting dementia in synucleinopathies.

### *Machine Learning and classical statistics*

The boundary between statistical inference and ML is much debated, but null hypothesis significance tests (ANOVA, t-tests, chi-squared tests, false discovery rate, or Bayesian inference) are commonly used for inference to calculate P-values for specific target variables. In this review article, "classical statistics" (CS) refers to these tests, explanatory analyses, and other inferential statistical analyses. In contrast, ML algorithms such as support vector machines, random forest algorithms, deep learning algorithms, and ordinary and penalized linear regression, are often used to build a model that is capable of out-of-sample generalization [8]. ML algorithms each have their strengths and weaknesses, making them suitable for different types of data and tasks. Researchers and practitioners often choose among these methods based on the nature of the problem, available data, and desired model complexity. It has been argued that CS may allow for a more mechanistic understanding of the patterns present in the entirety of a given data set, while ML aims for precise

modeling of a phenomenon which would be valid in a new dataset, beyond the data used to build the model. Although the borders are not clear cut, CS often isolates individually key variables based on P-values and often relies on predetermined modeling hypotheses, incorporating domain-knowledge. Meanwhile, ML aims to identify sets of variables that accurately predict outcomes generally, based on minimal assumptions, in some cases relying also on "black box" algorithms. However, recognizing complex relationships between variables is an area where many "black box" algorithms excel, making them more suitable for handling complex data. The aim is to generally model the output of the black box (e.g. distinguishing dementia individuals from those with no dementia) from input variables (e.g. neuroimaging measures) with minimal assumptions [9], in order to evaluate how accurately a built model can predict future results on independent datasets [10, 11]. That said, the notion of a black box is sometimes contested, as the weights can in theory always be accessed (even in the case of deep learning approaches, for instance). However, it is more difficult to associate clinical interpretation to neural network model weights than to handcrafted features derived by domain experts.

More globally, variables found to be most important by CS may not necessarily maximize out-of-sample predictive performance in an ML context, and conversely, useful and accurate predictive ML models may include variables that, individually, exhibit non-significant relationships with the outcome. The pursuit of clinical predictions complements efforts to understand the underlying biology in these neurodegenerative disorders and does not contradict or underestimate the importance of CS methodology [8].

### ***Applications of Machine Learning in clinical research***

Recent years have witnessed an exponential rise in the use of machine learning (ML) methods in neuroscience for a variety of applications [12-14]. This rapid rise is a testament of the strengths and utility of ML in the context of data-driven research. ML tools are often compared to "classical" statistical tools, i.e. to explanatory or inferential statistics, or to hypothesis-testing analyses. In fact, classical statistics and ML occur along a continuum of quantitative tools [11]. Both are relevant

but often serve different purposes (inference vs prediction) [15]. Classical statistical tools (e.g., explaining variance, testing for group effects) are often based on predetermined modeling assumptions (e.g., the central limit theorem), and do not generally evaluate the fitted model on new data [16]. For example, p-values or effect sizes in null-hypothesis testing cannot be used to judge new data at some later step (i.e., significant p-value  $\neq$  high prediction accuracy on new data) [11]. In contrast, ML uncovers general principles underlying a set of observations without the need for preset instructions or strong assumptions about data distributions [11, 17, 18]. More importantly, ML has the ability to quantify the model's extrapolation capacity to independent data through various model-validation techniques [11]. Indeed, instead of implementing a set of hardwired rules, ML algorithms learn the rules from the data. In one of the most widely used ML frameworks, known as supervised learning, the algorithm learns the rules from a training dataset (with known labels). The input to a classifier is typically a set of discrete and/or continuous features and the output is a class. For example, a ML algorithm can be used to classify patients into "with dementia" or "without dementia", based on sociodemographic input variables (e.g., age, education, etc.). To identify discriminative patterns in the data, the classification algorithm is trained – a process sometimes referred to as model fitting - on a subset of the available data (i.e., the training data) [11, 19]. Next, the fitted model is used for prediction on the rest of the data, which was excluded from the training (i.e., the test data) [11]. This process of evaluating a trained model on test data is known as model validation, and it assesses the generalizability of the model. One of the most prominent validation technique is an are out-of-sample estimate, where one evaluates the performance of the trained model on new data [19]. A standard out-of-sample model validation technique is the cross-validation procedure.

ML models have enhanced the potential of clinical research in neurodegenerative diseases by supplementing, rather than completing with, the established advantages of classical statistics. As a result, diagnosis and tracking of disease progression have improved, and new biomarkers have been identified [20, 21]. These are important advances given that the causes, pathophysiological mechanisms, and clinical presentation and progression of most neurodegenerative diseases, are still unknown, thus hindering the development of neuroprotective interventions. Moreover, most studies to date have been conducted on relatively small samples, thereby limiting extrapolation and

validation of the results. Consequently, “big data” analytics, such as ML-driven frameworks, are expected to gain more traction as a tool to (or to be a useful tool in identifying) identify prognostic biomarkers of clinical symptoms [22].

### ***Overview of neurodegenerative disease***

Neurodegenerative diseases can be classified by the type of abnormal proteins associated to the neuronal alterations. For instance, the tau protein is altered in Alzheimer’s disease, progressive supranuclear palsy, and fronto-temporal dementia (tauopathies), while  $\alpha$ -synuclein protein is altered in PD, multiple system atrophy (MSA), and DLB (synucleinopathies). The present review specifically addresses synucleinopathies and the cognitive decline observed in these pathologies, particularly in PD and DLB, and in one of their prodromal stages, isolated/idiopathic rapid-eye-movement (REM) sleep behavior disorder (iRBD).

PD is mainly defined by its motor syndrome, parkinsonism, characterized by bradykinesia accompanied by rest tremor and/or rigidity [23]. Non-motor symptoms are also very common in PD [24]. One of the most deleterious is cognitive decline, which can take the form of mild cognitive impairment (MCI; or mild neurocognitive disorder) or dementia [25, 26]. The prevalence of MCI among PD patients is estimated to be between 17 and 38%, according to cross-sectional studies [27, 28]. The presence of MCI is therefore a well-known risk factor for dementia in this population [29, 30]. Moreover, prospective studies reveal that 75-90% of PD patients will eventually progress to a dementia state over a period of more than 10 years [5, 31-33].

DLB is primarily defined by a progressive cognitive decline of sufficient magnitude, interfering with normal social or occupational functions, or usual daily activities [34]. DLB is diagnosed when dementia occurs before or concurrently with parkinsonism [34], while PDD is diagnosed when a dementia syndrome develops in the context of established PD [35]. DLB can appear with a varying combination of its core clinical features: parkinsonism, RBD, fluctuating cognition/alertness, and visual hallucination [34]. Probable DLB can be diagnosed if two or more

of these core clinical features are present or if one core feature is present with an indicative biomarker, while possible DLB can be diagnosed if only one core feature or at least one indicative biomarker is present [5]. The prodromal phase of DLB (i.e., a prodementia stage with symptoms indicating that DLB will subsequently develop) can manifest 15 years or more before dementia onset, and includes MCI, delirium-onset, and psychiatric-onset [36]. MCI due to Lewy bodies (MCI with LB) has recently been defined as deficits in one or more cognitive domains that are greater than would be expected from normal aging, accompanied by cognitive complaint from the patient, informant or clinician, not sufficient to interfere with typical daily functioning [36]. Probable MCI with LB can be diagnosed if two or more core clinical features of DLB or a single core clinical feature along with one biomarker are present, and possible MCI with LB can be diagnosed if one core clinical feature of DLB or one biomarker is present [36]. RBD strongly predicts later transition to DLB in patients with MCI who later develop dementia [37, 38].

iRBD, a parasomnia, is characterized by repeated episodes of behavior or vocalization during REM sleep with evidence of REM sleep without atonia [39]. A diagnosis of iRBD can be established in the absence of other significant neurological manifestations, such as parkinsonism or dementia [6, 40]. iRBD is a major risk factor for synucleinopathies; a multicenter longitudinal study reported that the estimated conversion rate increases steadily with time after RBD diagnosis, reaching 10.6% at 2 years, 17.9% at 3 years, 31.3% at 5 years, 51.4% at 8 years, and 60.2% at 10 years [40]. In patients who developed a synucleinopathy, approximately 43.5% developed dementia first, while 56.5% developed parkinsonism first [30]. Moreover, about one third of iRBD patients have MCI [6], which is associated with a greater likelihood of phenoconversion to a dementia-first compared to a parkinsonism-first phenotype [39-41].

### ***Affordable modalities and model validation: paving the way for clinical application***

Multiple medical tools have been tested and studied on prodromal or clinical synucleinopathies to investigate disease progression. This review focuses on data from (i) electroencephalography (EEG) and (ii) cognitive testing. These tools were chosen for their non-invasiveness, relatively low cost and demonstrated value in evaluating cognitive impairments in prodromal and clinical

synucleinopathies [6, 22, 42]. The relevance of using EEG and cognitive measures to evaluate and predict prodromal dementia in synucleinopathies has previously been demonstrated in several systematic reviews [6, 22, 42]. These markers were mainly identified through studies that used classical statistical approaches or simple ML algorithms (e.g., linear regression or logistic regression) and have allowed major advances in the field. However, since such studies generally do not use model validation methods, it is not possible to know whether or not the fitted models developed are valid or useful for predicting clinical outcome in patients outside of the study cohorts. This is problematic since many of these studies had relatively small sample sizes. Given the importance of generalizable results for clinical prognosis, there is a real need for a systematic assessment of the findings of strictly model-validated ML studies of cognitive decline progression in neurodegenerative disease.

***Objective:***

The current review aims to fill this gap by focusing on synucleinopathies that involve cognitive impairment (DLB and PDD) and their prodromal stages (MCI with LB, PD with MCI, and iRBD). This focus is motivated by the high occurrence and diverse nature of cognitive decline in patients with these diagnoses. Critically, early detection of prodromal dementia in synucleinopathies and better follow-up of the progression of disease in individuals at higher risk of cognitive decline could improve the intervention procedures of clinical trials and ultimately slow down or prevent cognitive decline [4, 33, 43]. This article presents a comprehensive review of EEG (resting, awake, and sleep) and cognitive studies that have applied ML approaches along with model validation techniques (in both cross-sectional and longitudinal data) to study prodromal dementia in synucleinopathies. Classifications between different synucleinopathy conditions, and healthy subjects (controls), are also examined. Furthermore, this article delves into the common obstacles that arise when using ML for dementia prediction, along with the constraints and potential advances that are needed in the future.



## 2. METHODS

### 2.1 Study selection

For study inclusion, PubMed was first systematically searched for potentially relevant studies up to July 8, 2022 (when the search was conducted), using combinations of several keywords. For clinical categories: “Dementia with Lewy bodies”, “Lewy body dementia”, “Lewy body disease”, “prodromal”, “Parkinson”, “Parkinson’s”, “Parkinson's disease dementia”, “pre dementia”, “REM sleep behavior disorder”, and “RBD”. For biomarkers: “electroencephalography”, “EEG”, “cognition”, “cognitive”, “neuropsychology”, “neuropsychological”, “neuropsychological test”, and “risk factor”. For machine learning: “Machine learning”, “ML”, “prediction”, “predict”, “predictor”, “predictive”, “algorithm”, “Linear discriminant analysis”, “Random forest”, “Decision Tree”, “support vector machine”, “SVM”, “LDA”, “RF”, “DT”, “neural networks”, “deep learning”, “logistic regression” and “linear regression”. All relevant articles were retrieved by placing restrictions on fields (free-text terms searched exclusively in the title or abstract of the articles).

Eligibility was initially assessed by the first author (LMB) by screening titles, abstracts, and entire articles, based on the following inclusion criteria:

- (1) Original research (only longitudinal and cross-sectional studies were considered);
- (2) Clinical category of patients studied: iRBD, PD or DLB with or without a prodromal state or cognitive impairment;
- (3) Analysis of either cortical (surface) EEG at resting-state (no task) or cognitive data (cognitive assessments or screening tests);
- (4) Use of ML algorithms with a model validation technique (e.g., cross-validation). This follows from the discussion above on the added value of model validation obtained through cross-validation methods;
- (5) Article in English.

For the second criterion, classification with control groups was also included, since the scientific literature has shown that cognitive and EEG alterations appear already at the RBD and PD stage [22, 41, 44, 45]. Regarding the fourth criterion, an overview of relevant findings from studies that used classical statistical methods outside the ML framework is also provided in the discussion section to ensure completeness. Treatment efficiency studies were excluded.

## **2.2 Data extraction**

The methodologies and outcomes of ML studies generally involve a common set of processes and variables (or features), which are summarized in Figure 1. In order to provide a comparative overview, the following seven elements were extracted from each paper: (i) Input variables (or features) examined, (ii) Feature selection method and best features, (iii) Algorithm(s) used, (iv) Decoding performance obtained for all metrics used, (v) Cross-validation method (explained in the introduction), (vi) Classification addressed, and (vii) Population (number of participants). In the following subsections, each of these elements is examined more closely.

### ***2.2.1 Input variables (or features) – EEG and cognitive testing***

EEG-derived features were a first type of input variable. The EEG records the electric potential signals on the scalp associated with synchronized synaptic activity [46]. Traditionally, EEG electrodes are placed in standard positions using the International 10-20 system involving 21 electrodes at different locations on the scalp (e.g., F=frontal, T=temporal, P=parietal, C=central, and O=occipital) [47, 48]. Usually, following pre-processing of the raw EEG (e.g. noise removal), spatial, spectral (frequency) or temporal information is extracted using one of many existing feature extraction methods [49]. Among the most widely used feature computation methods, the Discrete Wavelet Transform (DWT) and Discrete Fourier Transform (DFT) allow for the assessment of the power spectrum (or power spectral density) in the frequency domain [50]. The main frequency bands in which EEG signals are commonly decomposed are delta ( $\delta$ , 0.5–4 Hz), theta ( $\theta$ , 4–8 Hz), alpha ( $\alpha$ , 8–13 Hz), beta ( $\beta$ , 14–26 Hz), and gamma ( $\gamma$ , 30–100 Hz) [51, 52]. These EEG rhythms have been shown to be modulated by changes in behavior. For example, relaxed wakefulness,

drowsiness and eyes-closed conditions have been reliably linked to heightened alpha-band power [53], peaking over the occipital cortex [54]. A related and commonly examined EEG feature is oscillatory coupling, a measure of functional connectivity thought to capture the communication between brain regions. Commonly used metrics for inter-areal EEG coupling include coherence [55], directed transfer function [56], synchronization likelihood [57] and phase lag index [58].

Cognitive test score represented a second category of input variable. Such tests, which are often primarily conducted through paper-pencil assessments, are designed and built on the premise that they measure the behavioral expression of different cognitive functions [59] (e.g., executive functions, attention, memory, working memory, visuospatial, language; definitions in the Glossary Table). Cognitive tests developed to assess specific functions differ from cognitive screening tests. Screening tests are generally brief and narrow in scope. They can be administered during a routine clinical visit to help identify individuals in need of a more comprehensive assessment [60]. Some of the most frequently used screening tests are the Mini-Mental State Examination (MMSE; [61]) and Montreal Cognitive Assessment (MoCA; [62]). Comprehensive cognitive testing are multidimensional and used to determine the nature and severity of a person's cognitive difficulties, functional limitations and identifying primary and secondary diagnoses [60]. Cognitive data used as ML features typically consist of scores on cognitive tests assessing specific function or screening cognitive tests.

### ***2.2.2 Feature selection methods***

Feature selection is often executed during data preprocessing to enhance subsequent ML classification performance. This step removes input features deemed irrelevant or redundant from the analysis, thus reducing the dimensionality of the data [63]. By doing so, feature selections allow for a better understanding of the learning model and data [64].

### 2.2.3 Algorithms

Broadly speaking, ML algorithms can be split up into two categories: supervised and unsupervised. Supervised learning implements algorithms that learn a function that maps an input to an output, based on training on labeled data. In contrast, in unsupervised learning, the algorithms seek to uncover distinct patterns coherently present in the variables across subjects, without relying on data labeling [65]. Examples of supervised learning algorithms include Linear Discriminant Analysis (LDA), Decision Trees (DT), naïve Bayes (NB), or Support Vector Machines (SVM) [66], while unsupervised learning approaches include Principal Component Analysis (PCA), K-Means, and clustering analysis [67]. There are also other categories of ML algorithms, such as Reinforcement Learning (i.e., not supervised, but receives feedback called a ‘reward’) and Deep Learning (i.e., multi-layer neural networks) [65].

Short definitions of some of the most frequently encountered algorithms are:

- *Random Forest (RF)*: trains multiple decision trees as an ensemble and chooses the class based on the majority vote of the trees [68].
- *Linear Discriminant Analysis (LDA)*: is a linear algorithm that attempts to find a hyperplane projection that maximizes the average distance between two classes, while also minimizing the inter-class variance [69].
- *Support Vector Machine (SVM)*: performs linear classification, even with classes that are non-linearly separable, by finding a hyperplane (or a set) in high dimensional feature space using kernel functions (kernel trick) [70]. When trying to separate the classes, it specifically maximizes the distance from the nearest sample point with the help of support vectors [71].
- *Logistic Regression (LR)*: used to predict the probability of a target variable by fitting a logistic curve to the dataset [72].
- *Deep Neural Network (DNN)*: is an artificial neural network that tries to find underlying relationships in the data using multiple layers of nonlinearities [65]. The more hidden layers are included, the higher-level the learning of abstract representation of an input [73].

#### **2.2.4 Cross-validation**

Briefly, cross-validation is a statistical technique used to assess the performance and generalizability of a predictive model by partitioning data into subsets for training and testing, enabling reliable evaluation of its effectiveness. The process is repeated for different splits of the data. In n-fold cross-validation, the data is divided into n-subsets (or n folds), to iteratively train and test a model, ensuring comprehensive evaluation and reducing bias [11]. The average prediction accuracy across all test sets provides an estimate of the model's out-of-sample classification performance [19].

#### **2.2.5 Performance metrics**

A wide variety of metrics can be used to measure different performance aspects of ML models, regarding their ability to correctly classify data. Here, our focus is on the metrics most commonly used in clinical research studies. Table 1 is an illustrative example of a confusion matrix, which depicts the types of predictions and errors that can be made in the context of binary classification. These metrics are often used to compute performance metrics.

**Area Under the Curve (AUC)** quantifies the overall performance of a binary classification model by calculating the area under the receiver operating characteristic (ROC) curve, which shows the relationship between false positive rate and true positive rate for different probability thresholds [74].

**Accuracy** quantifies the proportion of correct classifications on the test set [75].

**Sensitivity**, also referred to as the true positive rate, measures the proportion of correctly identified positive instances in a binary classification model, indicating its ability to correctly classify true positives [75].

**Specificity**, also known as the true negative rate, measures the proportion of correctly identified negative instances in a binary classification model, indicating its ability to correctly classify true negatives [75].

The various performance metrics measure different aspects of algorithm performance. Which performance metric is appropriate for clinical applications can depend on the objective of the study and may have to take imbalance in the data into account [76].

## 2.3 Data analysis

### *Feature analysis:*

The studies were systematically categorized based on the classification question, for all modalities (EGG and cognition). We qualitatively analyzed the features studied for each classification question and highlight the notable variables identified in the research.

### *Algorithm analysis:*

From the included studies, we compiled all tested algorithms (i.e., including those of lower performance) and organized them in categories of algorithms. When counting the occurrence of each category of algorithm within a single study, we considered multiple instances of algorithms from the same category as a single occurrence (e.g.: SVM-RBF, SVM-P2 and SVM-P3 counted as 1 occurrence of the SVM category for that study). We also compiled the associated performance metrics and sample sizes (i.e., data balance/imbalance for the classification).

### 3. RESULTS

#### 3.1 Search results and classification conditions

The initial search yielded 3579 studies, all of which were examined for eligibility. Based on our exclusion criteria, 33 articles were retained for the purpose of this review (Figure 2). A substantial portion (90%) of the research articles included were published between 2016 and 2022, with a peak of publications occurring in 2020.

Figure 3 shows that, among the 33 studies retained, 26 employed a cross-sectional design, while 7 used a longitudinal design. The classification conditions examined in these articles can be organized into six primary study categories, some cross-sectional (1-4) and others longitudinal (5-6): 1) PD vs. control, 2) DLB vs. control, 3) PD vs. PDD/DLB, 4) PD vs. PD with a prodromal state (i.e., RBD or MCI), 5) iRBD patients progressing into either PD or DLB, and 6) PD patients that either developed cognitive decline/dementia or remained cognitively normal.

Most studies tested PD patients vs. controls (36,36%), then PD patients vs. PD patients with a prodromal state (21,21%), PD patients that developed cognitive decline/dementia (18,18%), PD vs. PDD/DLB (15,15%), DLB vs. control (3,03%) and iRBD patients that developed PD/DLB (3,03%). We can see great discrepancy in the number of studies in each category. Indeed, our review reveals that the longitudinal prediction of progression towards PD or DLB in individuals with iRBD, as well as the cross-sectional classification of DLB vs control only make up 3% of our studies (See Figure 3). These categories may be underrepresented and may be less reliable than other categories with greater number of studies. The majority of the studies aimed to distinguish PD patients from controls using EEG data during the resting state (37%).

It also shows that on average, more studies used EEG data (59%; n=20) than cognitive data (41%; n=14). It is worth noting that one study [77] examined both longitudinal EEG and cognitive data, and consequently appeared twice in the analysis.

Regarding the specific classification conditions, the distribution of studies is as follows:

1. *PD vs. control*: predominantly EEG studies (EEG=12, cognitive=1).
2. *DLB/PDD vs. control*: only one EEG study (EEG=1, cognitive=0).
3. *PD vs. PDD/DLB*: exclusively cognitive studies (EEG=0, cognitive=5).
4. *PD vs. PD with a prodromal state*: a balanced distribution (EEG=3, cognitive=4).
5. *iRBD evolving into either PD or DLB*: only one EEG study (EEG=1, cognitive=0).
6. *PD cases that either progress to cognitive decline or dementia*: slightly more cognitive studies (EEG=2, cognitive=4).

### **3.2 Classifications of cognitive status in PD:**

It's important to note that it is not possible to directly compare the performance of different classifications or modalities, as factors such as problem complexity, data quality, sample size, and hyperparameter optimization efforts can vary significantly, rendering direct comparisons inappropriate. For instance, the comparison of a 95% AUC score in the "PD vs controls" classification to an 87% score achieved in a classification "PD vs PDD/DLB" can be quite misleading. Therefore, we present the observations alongside each other to illustrate their respective performance and provide descriptive insights.

Tables 2 and 3 provide comprehensive summaries of the EEG and cognitive data studies, respectively. These tables detail the classification performance, encompassing all examined features, sample sizes, and highlight the most significant features identified in the machine learning models. Figure 4 specifically summarizes the most notable features for each classification problem, when the information was available.



***PD with prodromal state (MCI, iRBD, or cognitive decline):***

Resting state EEG studies that classified PD patients, with or without a prodromal stage, using EEG data have investigated a variety of features. These range from spectral power analyses (both relative and absolute) across all frequency bands to measures of cerebral connectivity, and further to micro-sleep during wakefulness. It should be noted that in the context of ML, since the algorithms capture and integrate complex patterns and interactions in the data without requiring a predefined direction of effect, many studies do not specify the directionality of the feature's effect. That is, they may reference "relative theta power in the frontal region", for example, but without specifying whether it is increasing or decreasing.

Two studies tried to classify PD with different cognitive impairments. In the first study, conducted by Betrouni et al. (2019), a cluster analysis was performed and revealed five distinct PD subtypes, according to the severity and nature of their cognitive impairments, based on neuropsychological assessments [78]. They obtained the best overall classification accuracy using KNN algorithm (87%) with the relative powers ( $\theta$  and  $\beta_1$  from frontal, temporal, central and parietal) at a low-density electrode configuration. The use of demographic data, high density electrode configuration, absolute power or all regions' electrodes (global) did not improve the accuracy. These results show the interest of region-specific spectral analysis. It is also possible to study the degree of synchronization between pairs of signals between specific brain regions (ex: the Phase Lag Index). This is what the Chaturvedi et al. (2019)'s study analyzed. They tried to distinguish PD with vs without MCI considering the Phase Lag Index functional connectivity measure, as well as spectral power ( $\delta$ ,  $\theta$ ,  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha$ ,  $\beta$ ). Using a RF algorithm, they found the most informative feature to be Phase Lag Index, yielding an AUC score of 74% when used alone [79]. When Phase Lag Index was used in combination with the frequency measures, classification performance was lower (73%). The best ranked features in this model were Phase Lag Index in  $\delta$  power between left frontal and right temporal regions, followed by Phase Lag Index in  $\beta$  power between left temporal and right parietal regions.

The investigation of another prodromal condition, iRBD, in PD patients shows distinct findings. Cesari et al. (2021) performed sleep and awake EEG on three groups of PD patients: with definite RBD; with REM behavioral events without definite RBD; and without RBD. After a feature reduction, 159 features were selected, belonging to these categories: spectral, coherence complexity, micro-sleep structures, and electrooculographic measures [80]. Using a RF, they achieved 81% accuracy in classifying the three groups. The most important features were defined as those used in at least half of the 50 RF models obtained through the 50 training and 50 validation datasets. The best performing features proposed were, in decreasing order: 1) micro-sleep instability, 2) EEG spectral features (relative power in  $\delta$ 1 band and maximum peak  $\delta$  band in REM sleep), 3) EEG complexity features (multi-scale entropy in REM sleep; scale. 8), 4) electrooculographic energy features, and 5) EEG coherence. Note that for the Cognition study from Chong-Wen et al. (2022) that classified PD patients with RBD vs PD patients without RBD, the sleep variable (i.e., the Pittsburgh Sleep Quality Index) was also the most important factor of the RF model, using 20 other variables, including cognitive screening tests, clinical (motor and anxiety/depression) and sociodemographic measures (see Table 2 for detailed features) [81].

In summary, to distinguish iRBD in PD, the features of micro-sleep instability are important. For MCI in PD, the noteworthy measures are the interhemispheric temporal-frontal and temporal-parietal connectivity in delta and beta frequencies, as well as the relative powers in theta and beta bands. The beta frequency band appears to be particularly relevant for machine learning models.

All of the cognitive studies that tried to classify MCI in PD patients used a combination of cognitive screening tests, with either sociodemographic data or clinical measures. Byeon et al. (2020) used a RF algorithm to distinguish PD patients with MCI and PD with normal cognition (n=51) [82]. They achieved a 65% accuracy using a large number of sociodemographic, life style (smoking, coffee), clinical (motor, sleep, depression), screening (MoCA, MMSE) and activities of daily living (see Table 2 for detailed features). The top four variables were 1) Clinical Dementia Rating sum of boxes, 2) Unified Parkinson's Disease Rating Scale motor score, 3) Korean MMSE total score, and 4) Korean MoCA total score. In another study two years later (Jeon et al. 2022

[83]), the MoCA score was combined with much less features (only combined to cognitive complaints) and allowed for the classification of PD with normal cognition from PD with cognitive impairment (MCI or dementia) with 87% accuracy using RF. The same year, Ortelli et al. (2022) used a new tool they named CoMDA (Cognition in Movement Disorders Assessment), merging the MoCA with two other cognitive screening tests: MMSE and Frontal Assessment Battery items. They obtained 81% AUC with QDA, using seven pre-screened variables: CoMDA score, gender, age, disease, disease duration, years of education, and level 1 score [84]. CoMDA score and age\*CoMDA and Education\*CoMDA were found to be the most important predictors. We see that the MoCA seems interesting in multivariable models, but also when used with few other variables.

### **3.3 Parkinson's Disease vs healthy control (HC)**

As mentioned in the first result section, PD vs HC was the most studied classification, making up to 36% of the studies in this review paper, mainly with the EEG modality. Only one cognitive study was found for the classification of PD. Souillard-Mandar et al. (2016) suggest the use of multiple features from a single cognitive test, the Clock Drawing Test [85]. They obtained the best performance with regularized LR (AUC of 91%), using five sets of digital-pen features from the Clock Drawing Test: clinician features (i.e., from the manual scoring systems, used by clinicians), operationalized clinician features (i.e., computed versions of the clinician features) and simplest features (i.e., features easy to evaluate by hand).

Different type of resting state EEG features were investigated, from spectral features, ratios to complexity measures and to no hand-picked feature extraction from raw EEG signal using neural network modes.

#### *3.3.1 Spectral (or frequency domain) analysis studies:*

Spectral (or frequency domain) analysis studies have shown a variety of results. Lee et al. (2022) used measures of the EEG signal's time-domain characteristics (estimating the band power across

wide frequency bands), the Hjorth parameters (activity, mobility, complexity) with the Gradient Boost Decision Tree algorithm (CatBoost; 91% AUC) with low  $\theta$  (best feature) and high  $\delta$  [86]. The subsequent studies were more focused on specific brain regions, introducing the concept of source reconstruction. This approach allows for the determination of the signal's origin, providing valuable information. It indicates the particular region for which a specific frequency band is more discriminative.

In the study by Waninger et al. (2020), they used a feature vector comprising 3060 variables, including absolute and relative Power Spectral Density and wavelet coefficients [87]. The most discriminative features, selected via step-wise discriminant analysis, were primarily in the theta, beta, and gamma frequency bands. These key features were observed in specific brain regions: theta and beta in the left and central frontal regions, theta, alpha, and gamma in bilateral temporal regions, and beta and gamma in central and right occipital regions. Using these targeted features, the linear discriminant function analysis (IDFA) model demonstrated a sensitivity and specificity of 84.84% and 95.24% respectively, effectively distinguishing between PD patients and HC.

Chaturvedi et al. (2017) incorporated in their ML model Global power (each band:  $\delta$ ,  $\theta$ ,  $\alpha_1$ ,  $\alpha_2$ ,  $\beta$ ), power in every region (each band),  $\alpha_1/\theta$  ratios (all regions), median and peak frequencies (I.e., the discrete frequency with the highest magnitude within that frequency range [88]) [89]. The most important variables (i.e., here minimizing prediction error) in the LASSO penalized logistic regression model (AUC of 63%) and RF model were the theta power in temporal left region and the  $\alpha_1/\theta$  ratio in the central left region.

Anjum et al. (2020) computed Power Spectral Density (Linear Predictive Coding coefficients) to conduct a PD vs HC single-channel (2 min-long eyes open + eyes closed) classification using the linear-predictive-coding EEG algorithm, a novel EEG-based signal processing approach that they developed [90]. They obtained 85% AUC using the 6 top-performing channels: TP8, P6, FC5, PO8, O2, and CP5.

Vanneste et al. (2018) used a SVM model for classifying PD patients versus HC, using frequency bands from EEG data [91]. The selected frequency bands were theta, alpha, beta, and gamma, derived from various cortical areas. The study achieved an accuracy of approximately 95% (AUC) using this approach. The most effective features were power across multiple frequency bands and in different cortical regions: left and right parahippocampus  $\theta$  and  $\gamma$ ; motor  $\theta$ ,  $\beta$ ,  $\gamma$ ; dorsal anterior cingulate  $\theta$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$ ; subgenual anterior cingulate  $\gamma$ ; right insula  $\theta$ ; posterior cingulate  $\theta$ ,  $\beta$ ,  $\gamma$ . By integrating these spectral features, ML models like SVM can leverage the nuanced changes in brain activity associated with PD.

Theta power is consistently identified as an important feature in various studies, but it is noted across different brain regions. It is predominantly reported in the temporal regions, especially on the left side, and also observed bilaterally.

### *3.3.2 Connectivity measures:*

Chang et al. (2022) achieved a 90% AUC in classifying PD patients from HC using a Bagging algorithm that focused on connectivity measures derived from EEG signals [92]. The model leveraged Holo-Hilbert Spectral Analysis (HHSA) to select the three most informative features: 1) FZ channel with AM frequency 1–2 Hz and Carrier frequency (Cf) 128–256 Hz; 2) F8 channel with AM frequency 2–4 Hz and Cf 4–8 Hz; and 3) C3 channel with AM frequency 8–16 Hz and Cf 32–64 Hz. These features encapsulate the degree of connectivity between specific frequency bands at different EEG channels – FZ (frontal cephalic), F8 (frontal right), and C3 (central left) – highlighting the importance of inter-frequency relationships in PD detection. The study's findings highlight the significance of connectivity measures in capturing the altered neural network dynamics in PD in a ML classification problem.

### *3.3.3 Complexity measures:*

Some studies used complexity measures (i.e., irregularity and unpredictability of brainwave patterns). The first study to do so was Liu et al. (2017), using Sample Entropy values with a three-way decisions model based on an optimal center constructive covering algorithm (O\_CCA) [93]. They achieved 93% accuracy and showed that the performance could be improved (+ 4.76%) by preprocessing the EEG signal with a discrete wavelet transform (i.e., decomposes the EEG signal into sub bands). A year later, another study by Yuvaraj et al. (2018) also examined complexity features, more specifically bispectrum feature (see Table 2 for details). A total of thirteen bispectrum features were extracted from 2-second EEG segments using higher-order spectra analysis. These features capture complex, nonlinear characteristics of the EEG signals. An SVM with radial basis function achieved optimum accuracy (99%) using only three ranked bispectrum features: H1 (i.e., bispectrum moment, as the sum of logarithmic amplitudes of bispectrum), Ent1 (i.e., bispectral entropy), and H2 (i.e., sum of logarithmic amplitudes of diagonal elements in the bispectrum) [94].

### *3.3.4 Neural networks:*

As noted above, there are many interesting features to consider for the classification of PD. With some ML models, specifically neural networks, it is possible to let the model autonomously determine which features are more important, based and the raw EEG signal. The field has seen significant advancements through the adoption of deep learning techniques, which effectively harness the rich and intricate data present in EEG studies.

Oh et al. (2018) used a convolutional neural network (CNN) model on raw EEG signal to accomplish the PD vs HC classification with 88% accuracy [95]. Xu (2020) also automatically learned features from the raw EEG data (Emotiv EPOC neuroheadset; 14 channels) for the same classification problem. They used a deep recurrent neural network (PDRNN) and report 88% precision, 85% sensitivity and 92% specificity [96]. Such neural network models (CNN and PDRNN here) obviate the need for pre-selecting features, instead directly learning from the raw data to identify patterns/features indicative of PD. This method represents a significant leap in EEG

analysis, allowing for a more organic discovery (i.e., no hand-picked feature extraction stages) of features that underlie the complex neural signatures of the disease.

In the study by Lee et al. (2021), a convolutional-recurrent neural network (CRNN) approach was used to classify resting-state EEG data for PD vs HC [97]. This deep learning model uses convolutional neural network (CNN) layers for extracting spatiotemporal features from the EEG signals. The model achieved an accuracy of 99.2% in classifying PD from HC. The CRNN model was particularly effective in detecting changes due to dopaminergic medication in PD patients, suggesting its sensitivity to clinically relevant features in EEG data. A study by Khare et al. (2021) specifically focused on the use or absence of medication in PD patients and the ML capacity to classify these clinical conditions from HC [98]. They classified with 97% AUC HC vs PD ‘off’ medication and 98% for HC vs PD ‘on’ medication using a least-squares SVM with five brain complexity and fractality related features across all sub-bands.

In summary, the relevance of complexity measures highlights the crucial role of complex, nonlinear brain signal dynamics in PD. This concept is further reinforced by power frequency studies, which identified a spectrum of spectral power features encompassing a wide array of neural dynamics, including the combined power of various frequency bands across multiple brain regions. Notably, in PD, neural dynamics are already significantly altered prior to the onset of dementia, presenting a considerable challenge in pinpointing the specific changes in PD that are predictive of dementia's development. The following section will delve into addressing this challenge.

### **3.4 Cognitive decline and dementia in Parkinson’s Disease**

#### *3.4.1 PD → cognitive decline/dementia:*

Two EEG studies employed longitudinal designs to investigate predictors of cognitive decline in PD patients (with or without MCI). First, Cozac et al. (2016) followed PD patients for a mean interval of 3 years. Cognitive outcome was assessed using an index of change in overall cognitive

score between follow-up and baseline [99]. Numerous baseline variables were considered as predictive features: motor, cognitive, psychiatric, and EEG resting state variables. The RF algorithm showed EEG global relative median power in the theta band to be the most important feature (i.e., the median power in the theta band measured relative to the power in the other bands), showing 75% AUC when used alone. This finding suggests the importance of the theta band power in distinguishing PD patients evolving to different levels of cognitive impairments. Second, Arnaldi et al. (2017), followed de novo PD patients for an average of 4.8 years. At baseline, patients were identified as mainly motor, intermediate, or diffuse/malignant (i.e., according to Fereshtehnejad et al. categories [100]) and as cognitively normal or impaired (i.e., either MCI or PDD) [77]. At follow-up, cognitive decline was reported in 21 patients, while 33 remained stable. The use of different clinical measures (neuropsychology, single-photon emission computed tomography and resting state EEG) in a stepwise LR allowed for the identification of the EEG posterior mean frequency at baseline as the best variable to predict cognitive outcome at follow-up. Posterior regions correspond to (P3, Pz, P4, T5, T6, O1, and O2) and mean frequency relates to the mean of frequency value in the 2-to-16-Hz band. Linear discriminant analysis was used with baseline EEG posterior mean frequency to classify PD with cognitive impairment vs PD without cognitive impairment with 82% accuracy.

Four studies used cognitive data to attempt to classify the cognitive outcome of patients. The first study, from Arnaldi et al. (2017) found that more cognitively impaired patients among their cognitive dataset showed lower baseline phonological verbal fluency and Trail Making Test (part A and part B) scores than cognitively stable patients (n=33) [77]. When classifying the cognitive outcome of patients using a linear discriminant analysis algorithm on the baseline data, they obtained an accuracy of 72% for the Trail Making Test part B, 65% for phonetic verbal fluency, and 63% for the Trail Making Test part A. This shows the value of using measures of executive control to predict cognitive decline in PD.

Hogue et al. (2018) followed PD patients for a mean of 3 years, during which 11% developed early cognitive decline [101]. Explanatory variables were demographic, disease history, motor, non-motor (Hopkins Verbal Learning Test, Benton Judgement of Line Orientation,



semantic fluency (animals), Wechsler Adult Intelligence Scale–3rd Edition Letter-Number sequencing, Symbol Digit Modalities Test). Based on bivariate associations with a LR, five variables from each category (demographic, motor and non-motor) were selected to be included jointly in a binary LR model: verbal memory retention, right-sided bradykinesia, years of education, subjective report of cognitive impairment and RBD. Bootstrap resampling was used to validate the model and a 75% accuracy was obtained.

Salmanpour et al. (2019) selected PD patients from the Parkinson’s Progressive Marker Initiative, using 93 features from demographic, motor and cognitive domains [102]. Ten classifiers were tested to predict the MoCA score four years after baseline data collection (year (y) 0 and 1). The study found that (Error: 1.65) using a non-dominated sorting genetic algorithm for feature selection yielded the best performance, combined with Local Linear Model Trees. Local Linear Model Trees selected MoCA years 0 and 1, RBD Questionnaire y1, Letter Number Sequencing Number 4 y0 and Number 3 y1, and State Trait Anxiety Inventory for Adults y0.

Horne et al. (2021) followed PD patients, of whom 25% developed dementia after four years. Each predictor of dementia risk was examined separately in a Bayesian prediction model: three neuro-psychiatric measures obtained from a feature selection conducted with a receiver operating characteristic curve (1. neuropsychiatric inventory total score; 2. hallucination; 3. anxiety), global cognitive score (expressed as an aggregate z score derived from an average of performance on measures taken across the five cognitive domains), age, and motor score [103]. Global cognitive score (delta ELPD = 36, SD 8) and age (delta ELPD = 11, SD 5) achieved highest performance.

These studies collectively indicate the utility of a range of cognitive assessments - including tests for executive function, working memory, and broader cognitive screening instruments, in the ML evaluation of the progression of cognitive decline or dementia in individuals with PD. They also indicate the classification importance of demographic and clinical data when used in conjunction with cognitive testing in ML models to improve the classification accuracy.

### 3.4.2 PD with and without dementia:

The following five cognitive studies used ML algorithms to discriminate between different cognitive conditions in PD or between PDD and DLB. No EEG studies addressed these classification condition.

Filoteo et al. (2009) used a discriminant function analysis, training on PDD, DLB and controls, to achieve a 79% correct classification rate when testing on a new sample of PDD [104]. Based on effect size, redundancy or distinct impairment, three variables from the California Verbal Learning Test were included in the model: short delay cued recall, percentage of perseverance errors, and List B recall.

Lee et al. (2016) developed a brief PDD assessment tool, the PDD-5S, a 30-item questionnaire (performance-based + brief informant interview) [105]. They built a classifier to distinguish patients with PDD vs PD without dementia using five screening questions (score= 1–2) from the questionnaire that were selected using a binary LR with a forward selection model, combined with a cognitive test (a 5-item questionnaire (score  $\geq 3$ ) and a verbal fluency test (category; score  $\leq 10$  items)) selected from literature. They obtained 87% AUC when using LR with these three variables on unrelated PD patients (i.e., not from the original study population).

Dujardin et al. (2010) obtained an AUC of 93% in distinguishing idiopathic PD vs from PDD patients [106]. Using a stepwise selection, five variables were selected for a LR model: the MMSE ( $\leq 26$ ), word generation ( $\leq 7$ ), number of free recalled words on the Five Word Test, score on the Pill Questionnaire, and age ( $\geq 70$ ).

Byeon et al. (2020) aimed to distinguish early-onset PDD vs PD with normal cognition using RF algorithm, discriminant analysis and the naïve Bayesian model [107]. Explanatory variables consisted of 36 different measures: sociodemographic, environmental, health, disease history, cognitive tests, RBD, and motor signs. Cognitive tests included: Korean MMSE, Korean

MoCA, Clinical Dementia Rating (CDR) scale global score and Sum of Boxes scores, Korean IADL, UPDRS (motor and total scores), Hoehn and Yahr staging, Schwab & England Activities of Daily Living. Using bootstrap samples for validation, best results were obtained with RF (AUC of 89%). Interestingly, the features with the highest weights for predicting EOPD only consisted of neuropsychological characteristics: #1) the Korean MMSE, #2) Korean MoCA, #3) CDR, #4) UPDRS, and #5) Korean IADL.

Byeon et al. (2021) studied PDD patients with varying severity of cognitive impairment measured by the CDR [108]. Explanatory variables included several sociodemographic, sleep, motor, cognitive, and PD-related clinical variables. After testing 8 SVM regression models, they obtained the lowest mean-squared error (0.078) with the highest convergence rate using the Nu-SVR-RBF model. The best predictors were found to be Korean IADL, total UPDRS, motor UPDRS, tremor, postural instability, age, age at diagnosis of PD, education level above high school graduation, a family history of PD, packs of cigarettes/year (21–40), coffee drinker, traumatic brain injury, atrial fibrillation, RBD, and depression.

In summary, the optimal features for differentiating PD with and without dementia in the discussed ML models encompass multiple cognitive assessments measures (focusing on memory and verbal tasks), evaluations of motor function, along with a comprehensive set of demographic, clinical, lifestyle, and functional variables. The combination of diverse features appears important for accurately distinguishing between PD with and without dementia.

### **3.5 Multimodality:**

This section examines the few studies that evaluated the efficacy of various modalities within identical cohorts and employing the same ML model. These studies focused on differentiating between DLB and control groups in a cross-sectional design and classifying cognitive states in PD through both longitudinal (PD leading to cognitive decline or dementia) and cross-sectional comparisons (PD-MCI versus PD without MCI).

In their study, Dauwan et al. (2016) classified DLB patients from controls (RF; 94% accuracy) by integrating various features: EEG characteristics (such as power across different frequency bands, peak frequency, minimum spanning tree, phase transfer entropy, and  $\theta/\alpha$  ratio), clinical indicators (hallucinations and extrapyramidal signs), biomarkers (MRI and CSF), and neuropsychological tests (MMSE score, visual association test, Trail Making Test, and Digit Span). EEG features (mean delta power, lowest and highest theta power) were identified as the most performant in this comprehensive model - EEG characteristics were the top 22 out of 30 key features, followed by three neuropsychological tests (MMSE, TMT-A, and Digit Span backward) and MRI/CSF biomarkers [109]. Moreover, when each modality was used independently, EEG showed the best results with 92% accuracy. Cozac et al. (2016) conducted a 3-year study on PD patients, aiming to predict cognitive outcomes such as decline or dementia [99]. They used an index of change in cognitive scores from baseline to follow-up for assessment. Their ML model incorporated various predictors: Global Relative Median Power (GRMP) across EEG frequency ranges, cognitive domains (attention, executive functions, fluency, memory, and visual-spatial abilities), and demographic/clinical variables (age, sex, education level, disease duration, observation period, Levodopa dosage, and UPDRS-III score). The model achieved a 75% Area Under Curve (AUC) in classifying cognitive outcomes, with theta range GRMP emerging as the most significant feature. The RF algorithm showed EEG global relative median power in theta to be the most important feature, presenting 75% AUC, when used alone. In a separate study conducted by Arnaldi et al. (2017), the importance of using EEG data was once again confirmed as it yielded the highest accuracy when used alone, compared to other modalities. It did not involve comparisons with other types of neuroimaging data, focusing solely on cognitive, demographic, and clinical information. The study used a stepwise logistic regression approach with different modalities, including neuropsychological tests, single-photon emission computed tomography (SPECT), and resting-state EEG. The posterior quantitative EEG mean frequency (qEEG MF; 82% accuracy) and the more affected hemisphere caudate specific to non displaceable binding ratios (MAH caudate SBRs; 80% accuracy) were identified as the most important predictors. The addition of other potential predictors did not significantly improve the outcome prediction, indicating that these two features were the most effective in the model. In comparison, the prediction accuracy

was lower for age (68%), phonological verbal fluency (65%), Trail Making Test (TMT) A (63%), and TMT B (72%) scores compared to qEEG MF and MAH caudate SBRs [77].

The findings from these studies indicate that EEG features consistently outperform other modalities in accuracy and importance within multimodal frameworks, whether in conjunction with neuroimaging data, cognitive assessments, or sociodemographic information.

Regarding the studies that used cognitive data, the majority of them adopted a multimodal approach, incorporating various features, mainly cognitive screening tests with sociodemographic and clinical measures. These studies shared a common objective of classifying MCI or dementia in PD patients [82-84, 107]. While there isn't a consistent trend in terms of which category of variables holds more importance in the multivariable models, we identified that neuropsychological or screening test features tend to be more important in the multivariable model than clinical and demographic features.

### **3.6 Machine Learning algorithms in synucleinopathy research**

In this review, only studies using model-validation techniques were selected, which reduces the risks of overfitting. Detailed descriptions of the model validation process, feature selection methodologies, and performance metrics used in the studies can be found in Tables 2 and 3. These tables serve as a resource for researchers aiming to apply similar algorithms and methodologies in their work. The predominant approach for model evaluation was 10-fold cross-validation, which ensures that the model is tested on 10 distinct subsets of the dataset. Feature selection, a step preceding the machine learning classification, was conducted using inferential statistics to determine the most relevant features to include in the ML model. Our analysis also revealed that the accuracy metric was the most employed metric across the 33 studies, closely followed by the Area Under the Curve (AUC).

### *3.6.1 Algorithms, metrics and data imbalance*

The analysis of all ML algorithms (i.e., including those of lower performance) tested in the 33 studies revealed multiple types of algorithms, which we organized into ten categories: Gradient Boosting (GB), Neural Networks (NN), Decision Trees (DT), K-nearest Neighbors (KNN) and K-means, Linear and Quadratic Discriminant Analysis (LDA and QDA), Logistic Regression (LR), Bayesian Models, Random Forest (RF), Support Vector Machine (SVM), and the LEAP algorithm. This process revealed 65 distinct evaluations of these categories of algorithm, offering insights into their frequency of use and the associated metrics and data balance/imbalance.

In Figure 5, we selected twenty-seven studies – only including those using the AUC or the accuracy metrics and that provided information on class balance or imbalance for their binary classification problem. Our analysis focused on these two metrics, given their predominant use in the literature (29/33 studies). Note that studies that reported both metrics were classified under ‘AUC’, while those reporting only accuracy were placed in the 'Accuracy only' category. The Arnaldi et al. (2017) study was counted once [77]. Figure 5’s legend details the exclusion criteria.

Our findings highlight: 1) a mean number of 2,53 +/- 2,13 categories of algorithms were tested in each study; 2) a significant reliance on RF (18,56%) and SVM (15,38%) algorithms, despite the exclusion of some studies; 3) more studies using only accuracy (15 studies; 47.69% of algorithms) than employing AUC (12 studies; 52.31% of algorithms); 4) an important incidence of imbalanced datasets, more precisely 67,69% of the algorithms were tested on imbalanced data (15 studies). Finally, our study reveals that 32,31% of the algorithm’s performance were evaluated with the accuracy metric using imbalanced dataset (6 studies).

### *3.6.2 Comparative analysis of algorithm performance within individual studies*

In this review, we identified a subset of studies that performed within-study comparisons of multiple machine learning algorithms. This approach is advantageous for evaluating algorithmic performance, as it ensures uniformity in terms of the cohort and features used across different algorithms. Note that all studies addressed the PD vs HC classification.

Chaturvedi et al. (2017) tested four algorithms (AUC) with spectral features: RF (80%), SVM (73%), J48 DTrees (67%), LR - LASSO (63%). They obtained best performance with RF, but reported the LASSO penalized logistic regression model [89]. This focus was due to LASSO's ability to select a small subset of uncorrelated features from a dataset with high multicollinearity, which is essential in reducing overfitting and improving model interpretability. Yuvaraj et al. (2016) tested six different algorithms with brain complexity measures (accuracy – balanced dataset): decision tree (90%), KNN (90%), fuzzy KNN (93%), naïve Bayes (88%), probabilistic neural network (93%) and SVM (P2-91%, P3-95%, RBF-99%). They found the best performance using the SVM classifier with RBF (Radial Basis Function) kernel [94]. Lee et al. (2021) tested traditional ML algorithms (e.g. RF-85%, SVM-RBF (highest SVM)-95%), but also deep learning algorithms (KNN-93%, convolutional neural network (CNN)-99%) and suggested best performance is obtained with CNN using spatiotemporal features [97]. Khare et al. extracted five brain complexity related features across all sub-bands, testing (accuracy - balanced dataset) for RF (96%), KNN (96%), SVM (95%), NN (95%), and least-squares SVM (97%). The classification with SVM yielded the highest accuracy [98]. Lee et al. (2022) used Hjorth parameters (activity, mobility, complexity), testing three algorithms with the AUC metric: XGBoost (70%), LightGBM (89%), and CatBoost (91%) with the highest discriminant ability with low  $\theta$  and high  $\delta$  [86]. Chang et al. (2022) obtained 90% AUC using bagging, superior to logitboost (89%), gentleboost (88%), decision tree (82%), SVM (86%), naïve Bayes (70%), KNN (80%). They suggested three prioritized Holo-Hilbert spectral analysis features (non-linear), at specific electrode locations [92]. Souillard-Mandar et al. (2019) suggests that the use of digital features from a cognitive test (Clock Drawing Test or CDT) can improve classification of PD. They tested multiple algorithms with the AUC metric: C4.5 (75%) decision tree, classification and regression trees (76%), SVM Gaussian (86%), RF (91%), boosted decision tree (87%), and regularized LR (91%). They obtained the best performance with regularized LR (AUC of 91%) [85].

In summary, no single algorithm consistently outperforms others across all studies. The best-performing algorithm varies depending on the specific features, data type, and classification task at hand.

## **4. DISCUSSION**

This systematic review offers a state-of-the-art extensive analysis of the ML studies that explored classifications across the spectrum of synucleinopathies, such as iRBD, PD and DLB, as well as their progression towards cognitive decline or dementia. To ensure out-of-sample generalization and facilitate potential clinical applications, our focus was on ML research using model-validation techniques (i.e., cross-validation) and accessible biomarkers, such as EEG and cognitive tests. The objective is to provide a comprehensive introduction and guide for clinicians and researchers who wish to use ML to better understand the evolution of synucleinopathies, while highlighting limitations and future research opportunities and contrasting ML findings with prior, non-ML, studies.

### **4.1 Emerging field:**

Our analysis highlights the recent emergence of ML in synucleinopathy research, with 33 studies meeting inclusion criteria, most of which were published within the last six years. While six different types of classifications have been explored, the majority of the research efforts focused on distinguishing PD from controls. There is a remarkable lack of longitudinal studies that use ML to predict cognitive trajectories in patients with iRBD or PD (toward cognitive decline or dementia). This suggests that the methodological application to these sorts of clinical and neuroscientific questions is in its infancy; or, perhaps even at an embryonic stage for certain conditions. It would be beneficial for future research to concentrate on differentiating cognitive stages within PD or in a longitudinal context in iRBD and PD. iRBD is particularly important to target with generalizable ML, since it is a prodromal state that can precede dementia diagnosis by many years.

Overall, our systematic literature review reveals that there is significant diversity in the classifications studied, but also in the algorithms, variables/features, and metrics used, which restrains the identification of stable performance trends. Therefore, our work highlights the vast



potential, but also the numerous gaps, that need to be addressed in the literature in order to facilitate the translation of these advancements into clinical applications.

Up until 2016, the main analysis approaches used to address the evolution of synucleinopathies were group comparisons (ANOVA, ANCOVA, T-test,  $\chi^2$  test) [41, 44, 110-116], correlations [41, 110, 117-122], regression [44, 111, 112, 123] or logistic regression without generalization [41, 124, 125]. Our study indicates that there is a partial overlap in variables identified as critical in Classical Statistics (CS) and ML. Some key variables in CS do not improve classification performance in a multivariable ML model, whereas some variables from CS occasionally emerge as important in multivariable prediction models. For instance, the Trail Making Test – part B, a measure of executive function which has been found to be significantly affected in PD patients who evolve to dementia [41] was also found to yield significant classification accuracy in the multivariate ML model (72% accuracy) [77]. Our literature review also suggests that CS can play an important role in preselecting input variables for the ML model, highlighting that findings from non-ML studies may provide helpful information that can be leveraged in the feature selection step of ML classification studies.

#### **4.2. Parkinson's Disease identification (or PD vs HC)**

Different type of resting state EEG features were investigated. Multiple spectral studies (n=5) identified theta power as an important feature in the ML models, across different brain regions, but predominantly in the temporal and frontal regions, bilaterally. The frontal regions also appeared important in studies investigating inter-frequency connectivity (n=1), showing that modulation of high frequencies by low frequencies, mainly for the electrodes in this region allow for significant PD detection. The brainwave patterns were also studied, in terms of their irregularity and unpredictability – through complexity measures (n=2), with promising results. The relevance of complexity measures highlights the crucial role of complex, nonlinear brain signal dynamics in PD. This concept is further reinforced by the spectral studies that have identified a spectrum of spectral power features encompassing a wide array of neural dynamics, including the combined power of various frequency bands across multiple brain regions. Thus, we find several interesting

measures to consider in the detection of PD by investigating resting-state EEG - from spectral features in different brain regions and inter-frequency connectivity to complexity measures.

Given all these interesting features, with some ML models, specifically neural networks (n=4), the model can autonomously determine which features are more important, based and the raw EEG signal, without the need for prior hand-picked features. We demonstrate that several studies in this comprehensive literature review use neural networks to significantly classify PD from control subjects. The field has seen significant advancements through the adoption of deep learning techniques, which effectively harness the rich and intricate data present in EEG studies.

The classifications of PD from healthy controls reveal that neural dynamics in PD are already significantly altered prior to the onset of dementia, presenting a considerable challenge in pinpointing the specific changes in PD that are predictive of cognitive decline or dementia's development.

### **4.3 Cognitive decline in synucleinopathies**

#### ***4.3.1 EEG slowing***

Our exhaustive ML literature review, building upon existing SC studies in synucleinopathies, reveals the interest of further exploring EEG slowing in relation to cognitive decline and dementia across the synucleinopathy spectrum using ML approaches.

The clinical condition of DLB has been shown to be characterized by EEG abnormalities in the resting state, such as increased slow-wave in posterior regions (occipital [126-128] and temporal [129, 130]). The usual dominant alpha rhythm appears to be slowed down, adopting a rapid pre-alpha/theta rhythm and becoming more variable over time [126, 128, 130]. Thus, the pre-alpha frequency dominates, either consistently or intermingled with alpha, theta, or delta activity. The critical role of EEG slowing in DLB was reinforced by Dauwan's ML study, which classified

DLB patients from Controls principally with the theta/alpha ratio, the most important feature of the ML model [109].

Similarly, the severity of cognitive impairment in PD, has been linked to EEG slowing, as evidenced by multiple SC studies [117-121, 131-135], showing correlations with spectral ratio of fast-over-slow EEG power [117, 118, 122, 132, 134] or higher dominant frequency [113, 120, 132, 133]. In our exhaustive literature review, the Chaturvedi et al. (2017) study classified PD patients without cognitive impairments from HC using the alpha1/theta ratio in the central left region, pointing to its potential in distinguishing PD before the onset of measurable cognitive decline [89].

EEG slowing during wakefulness has also been identified (in SC) as a potential biomarker of neurodegeneration in iRBD patients [42], with specific patterns of slowing in the central and occipital regions of the right hemisphere associated with the short-term onset of MCI in RBD patients [136]. Additionally, a higher baseline of EEG slowing has been correlated with subsequent development of DLB in iRBD patients [137]. Despite these known correlations and the prodromal state of DLB in iRBD, our review found limited studies investigating EEG features as predictors of dementia in iRBD. The ML study by Ruffini et al. (2019) diverged by focusing on spectrogram features (i.e., 2D time-frequency representation) through a deep convolutional neural network, identifying a single EEG channel in the parietal lobe as the model's most prominent feature [138]. This gap highlights the need for future ML research to prioritize EEG slowing measures in iRBD as predictors of cognitive decline.

#### ***4.3.2 EEG Connectivity***

EEG connectivity, which assesses communication between different brain regions, emerges as a promising method for the classification of MCI in PD, specifically using the Phase Lag Index (PLI), between all pairs of brain regions and within individual regions [79]. Indeed, the PLI was found to be more effective in distinguishing PD with MCI from PD without MCI when used alone, than when combined with spectral power measures. This shows that connectivity measures have high

importance in the ML model and that the addition of spectral power features can introduce noise to the ML model.

Among the various connectivity measures, some demonstrate limited utility in addressing specific classification challenges. Notably, EEG coherence features were less effective in differentiating DLB from controls, even when used in combination with spectral power measures and were also less crucial for identifying RBD in PD, compared to measures of complexity, spectral EEG, electrooculogram, and microsleep.

Our review highlights the need for more research that employs a variety of EEG connectivity measures as features in ML frameworks, as the properties of inter-region brain communication seem to hold additional information for understanding disease progression [139].

#### ***4.3.3 Cognitive studies and multimodality***

The previous studies that have focused on the early stages of PD show that nearly 24 to 36% of patients present cognitive impairments [140, 141]. Executive dysfunction is the earliest cognitive deficit in PD patients [142, 143], followed by disruption of verbal episodic memory [143]. CS studies have also shown that PD patients also demonstrate difficulties in visuospatial functions [144] and information processing speed [145]. When PD is accompanied by MCI, it is associated with a diverse range of deficits in attention, executive functions, visuospatial ability, and memory [146]. Varying results have been found when investigating cognitive tests as dementia markers in PD [45], with significant alterations at baseline on cognitive tests measuring attention, executive function (e.g., TMT B and semantic fluency; [45]), episodic memory [45, 147-149], visuospatial abilities [45, 149-152] or language in PD patients developing dementia [125, 147, 150-158].

Our review indicates that cognitive studies that address cognitive state or decline in PD generally employ a multimodal approach, showing that the classification accuracy is improved by the inclusion of demographic and clinical data to cognitive data. For example, the optimal features

for differentiating PD with and without dementia in the discussed ML models encompass a blend of cognitive assessments, evaluations of motor function, along with a comprehensive set of demographics, clinical, lifestyle, and functional variables. This combination of diverse features appears important for accurate classification, indicating that the model benefits significantly from a multifaceted spectrum of information.

Within those multivariable models, cognitive test features often play a more important role than clinical and demographic variables. A range of cognitive assessments are reported across studies - including tests for executive function, working memory (in the evaluation of the progression of cognitive decline or dementia in PD patients) and memory and verbal tasks (PD with and without dementia).

Our review provides insights into the specific contribution of the broader screening tests, such as the MoCA in clinical multivariable models for the classification of cognitive states in PD. We show that screening tests improve classification performance when combined with specific cognitive measures. Greater accuracy (87% AUC) was reported to distinguish PDD from PD [105] when combining verbal fluency (category) with the screening test (5-item questionnaire), instead of using the screening test alone (80% AUC). This added value is also found in discriminating MCI in PD (89% AUC) with the Korean-MMSE combined with other cognitive variables [82]. To better understand which cognitive functions are most valuable in the ML prediction models of cognitive decline in PD, we suggest a thorough assessment of all cognitive domains in future experimental designs. However, we recommend future studies to further investigate these markers and clarify the best combination of factors that allow for a clearer differentiation in the different states in PD.

In studies that compared the efficacy of various modalities (i.e., EEG, cognition and demographical) within the same cohort, the EEG modality frequently stands out as the most important for various classifications (e.g., classification of PD progression to cognitive decline or dementia), even when used with cognitive data or demographical data. However, more research are needed to further investigate multimodal approaches in order to clarify each modality's complementary contribution.

#### **4.4 Machine Learning algorithms – attempt of comparison**

Our exhaustive literature review reveals an algorithmic diversity in synucleinopathy research, with ten categories of algorithms investigated across the 33 studies. We show a greater focus on supervised learning models - the most used algorithm being RF, followed by SVM. This could also be attributed to their established proficiency in handling high-dimensional data and their robustness against overfitting respectively.

While all algorithms reviewed surpassed the threshold of chance (see Table 2 and 3), with the majority demonstrating classification performance between 80-95%, it is crucial to consider the potential influence of publication bias, commonly known as the file drawer problem, when interpreting these results. This phenomenon occurs as studies yielding significant or high accuracies are preferentially published, whereas those with non-significant, or inconclusive findings are disproportionately underrepresented in the literature. Direct comparison of algorithmic performance across studies was not feasible due to several constraints. First, when studies are categorized by our six classification types, algorithms, and performance metrics, it leads to an insufficient sample size (mean  $n=3$ ) for statistical analysis. Secondly, we do not have access to the data from the majority of the studies, which is often the case in clinical research. Finally, it is not possible to directly compare the performance of different algorithms across different studies, since factors such as problem complexity, data quality, sample size, and hyperparameter optimization efforts can vary significantly, rendering direct comparisons inappropriate. It should also be acknowledged that while studies in the "EEG" and "neuropsychological" categories have employed these modalities, the variables examined are quite diverse (e.g., EEG: connectivity, power, ratios; neuropsychological: screening tests, combinations with sociodemographic tests, or exclusively cognitive tests), as can be seen in Tables 1 and 2. This further complicates direct comparison, reinforcing our stance against oversimplified cross-study comparisons.

Through the identification of a subset of studies that performed within-study comparisons of multiple ML algorithms, we ensured uniformity in terms of the cohort and features used when comparing algorithmic performance. Our analysis of studies that performed within-study

comparisons of multiple ML algorithms reveals that no single algorithm consistently outperforms others across all studies. The choice of the best-performing algorithm varies depending on the specific features, data type, and classification task at hand, highlighting the need for tailored approaches in PD research. These findings suggest that the choice of algorithm should be driven by the specific characteristics of the dataset and the clinical question at hand. The diversity in algorithmic performance across studies underscores the importance and need for tailored approaches to data analysis in PD research. Future studies could benefit from exploring a broader range of algorithms, including those underrepresented in the current literature, to establish more comprehensive benchmarks for algorithm performance in PD classification tasks.

Numerous evaluation metrics can be used to assess the performance of ML models and some metrics can be impacted by data imbalance. The choice of the metric often depends on a range of factors, such as model complexity and data type [74, 159]. When dealing with balanced data (i.e., an equal number of subjects in each class) in a classification problem, the AUC will be slightly higher than accuracy, while accuracy will overestimate performance in cases of imbalanced data [74]. Our analysis of the metrics and sample sizes of the data classes in the binary problems revealed that 32,31% of the algorithm's performance were evaluated with the accuracy metric using imbalanced dataset (6 studies). This asymmetry in class distribution is a critical factor to consider, as it suggests that bias could be introduced in the predictive modeling, potentially compromising the generalizability of the findings of these studies. Future studies should consider class balance or imbalance before using one performance metric or another (AUC or accuracy).

## **Limits**

ML methods are not without limitations, as they present challenges related to data acquisition and preprocessing. This includes difficulties with the availability and readability of data, heterogeneity and incompleteness in data acquisition and influence of unintentional but confounding factors (e.g., age, gender, culture, statistical dependence between training and testing data) [11]. Considering the diverse features employed by the reviewed studies for various classification conditions, and the fact that the limited number of studies did not allow for a quantitative meta-analysis, we were

unable to identify a specific marker or combination of markers to target for a specific condition. Therefore, we offered a qualitative comparison of the features associated with best performance in multiple classifications, providing an overall perspective. We did not address the use of other types of imaging modalities (e.g. functional magnetic resonance imaging) and the combination of these tools together (multi-features from different modalities), which have been shown to increase classification performance.

### **Contribution and future paths**

We lay the foundation for clinicians and future researchers who wish to use ML to better understand the evolution of synucleinopathies while ensuring for out-of-sample generalization to facilitate potential clinical application (i.e., model-validation techniques and accessible biomarkers, such as EEG and cognitive tests). A constraint of this review is the limited number of studies that were conducted so far on this research question. Therefore, the observations/trends found need to be reinforced and confirmed in the future through more studies. To improve the utility of our study for future research, we suggest broadening inclusion criteria to encompass more studies, particularly in ML analyses of cognitive decline in synucleinopathies, without compromising data quality. This includes considering sleep states in EEG and other common neuroimaging and electrophysiology methods in this clinical population. However, this approach might limit clinical applications due to the higher costs and restrained accessibility of these methods. Additionally, expanding the classifications to include a comparison of DLB against other dementias, such as Alzheimer's disease, both cross-sectionally and longitudinally, could also increase the review size. Lastly, it's noteworthy that many relevant studies were excluded due to a lack of detailed model validation methods, hindering our goal of generalizing results. We recommend maintaining this stringent selection criterion in future study selections.

Future research should investigate the potential of different algorithms, modalities, and combinations to increase knowledge on the contribution of each algorithm and feature in the prediction of dementia in synucleinopathies, to refine and optimize the classifications. By addressing these aspects, our understanding of the field will continue to grow and contribute to the



development of more effective diagnostic tools and interventions for these complex conditions. Furthermore, the validation of these findings in larger, more diverse cohorts could help establish the generalizability and clinical relevance of the proposed ML-based approach for PD and related conditions. Indeed, efficient ML models could help target the best candidates for clinical trials of medication that aims to prevent cognitive decline.

The current review is timely and valuable for several reasons. First, given that cognitive deterioration begins years prior to the diagnosis of PD or dementia, predictive models such as ML are especially suited for early prognosis. Second, ML with model validation, as well as the accessibility of the variables studied (EEG and neuropsychology) favor the generalization of the results to clinical applications in the population. Early detection of individuals at risk of cognitive deterioration could allow the evaluation of the clinical treatments aimed at preventing cognitive decline. Third, since 2010, ML has experienced a strong resurgence with changes in data properties, increased computational power and cheaper memory resources, an update on recent research using ML is necessary. This review offers the opportunity to bring new theoretical knowledge, through its global overview on ML with model validation and through the comparison between classical statistics and ML with/without model validation for generalization. This overview of the key features, the mainly studied algorithms and missing blocks (to address the less studied classifications and variables and to develop better prediction algorithms) in the continuum can be used as a guide for future research using ML in this clinical population. Other strengths of this review include the application of a systematic search strategy and an exhaustive description of the methods of analysis used in these studies (features, algorithm, cross-validation, feature selection and feature importance).

### **Acknowledgments including sources of support**

The authors were funded by the Canadian Institutes of Health Research (JFG) and the National Institutes of Health/National Institute for Aging. JFG holds a Canada Research Chair in Cognitive Decline in Pathological Aging. KJ is supported by funding from the Canada Research Chairs

program and a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada and an IVADO-Apogée fundamental research project grant.

### **Conflict of Interest**

The authors have no conflict of interest to report.

**Table 1** – Example of a confusion matrix.

		<b>Predicted class</b>	
		Dementia	No dementia
<b>Actual class</b>	Dementia	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
	No dementia	<i>False positive (FP)</i>	<i>True Negative (TN)</i>

**Table 2** – Reviewed EEG studies.

		EEG							
		Population	Classifier	CV	Features	Feature selection	Best decoding performance	Most discriminant feature	
DLB vs HC	Dauwan et al. [109]	DLB=66; HC=66	RF	intern	power ( $\delta$ , $\theta$ , $\alpha 1$ , $\alpha 2$ , $\beta$ ), peak frequency; MST; PTE, $\theta/\alpha$ ratio	manual + RF	Accuracy: 92%	$\theta/\alpha$ ratio ( $\theta/(\theta + \alpha 1 + \alpha 2)$ )	
	Chaturvedi et al. [89]	PD-MCI=27 PD-nMCI=43	RF	70-30%, 5-fold	spectral power ( $\delta$ , $\theta$ , $\alpha 1$ , $\alpha 2$ , $\alpha$ , $\beta$ ), PLI	X	AUC: 74%	#1: PLI in $\delta$ FL-TR #2: PLI in $\beta$ TL-PR	
PD vs PD with PS	Betrouni et al. [78]	PD:G1=28; G2=33; G3=43; G4=5; G5=9	SVM, KNN	85-15%, 5-fold	absolute and relative power ( $\delta$ , $\theta$ , $\alpha 1$ , $\alpha 2$ , $\beta 1$ , $\beta 2$ ) global/specific, demographic	correl	Accuracy (KNN): 87%	Relative-powers: $\theta$ and $\beta 1$ from frontal, temporal, central, and parietal	
	Cesari et al. [80]	PD-RBD=26 PD-RBE=2 PD-nRBD=54	RF	50-50%, 10-fold (10x)	Micro-sleep structure (n=90), spectral (n=54), coherence (n=32) complexity (n=184), and EOG (energy) (n=36)	396 à 159	Accuracy: 81.5%	#1: micro-sleep instability #2: EEG spectral features #3: EEG complexity #4: EOG energy #5: EEG coherence	
	Chaturvedi et al. [79]	PD=50 Controls=41		10-fold	Global power (each band: $\delta$ , $\theta$ , $\alpha 1$ , $\alpha 2$ , $\beta$ ),		AUC (LR): 80%	$\theta$ power T-L, $\alpha 1/\theta$ ratio C-L	

PD vs HC			RF, SVM, J48, LR(LASSO)		power in every region (each band), $\alpha1/\theta$ ratios (all regions), median and peak frequencies in Occipital	Penalized LR + LASSO		
	Liu et al. [93]	PD=17 Controls=25	O_CCA	10-fold	Complexity (Sample Entropy values)	-----	Accuracy: 92.9%	-----
	Yuvaraj et al. [94]	PD=20 Controls=20	DT, FKNN, KNN, NB, PNN, SVM P2/P3/RBF	10-fold	Complexity (13 bispectrum): H1, Ent1, H2, H3, aWcobx, H4, Ent2, Pent, Wcobx, aWcoby, H5, Wcoby, Mavg	HOS	Accuracy (SVM- RBF): 99.6%	3 ranked bispectrum features: H1, Ent1, and H2
	Vanneste et al. [91]	PD=31 Controls=264	SVM	10-fold	$\delta, \theta, \alpha, \beta, \gamma$ for each ROI: L and R auditory, somatosensory, motor, parahippocampus, insula, dorsal ACC, subgenual ACC, and PCC	SVM	AUC: 95%	L and R parahippocampus $\theta$ and $\gamma$ ; motor $\theta, \beta, \gamma$ ; dorsal ACC $\theta, \alpha, \beta, \gamma$ ; subgenual ACC $\gamma$ ; R insula $\theta$ ; PCC $\theta, \beta, \gamma$
	Oh et al. [95]	PD=20 Controls=20	CNN	Stratified 10-fold	14 channels raw / 1588 artifact-free epochs	-----	Accuracy: 88.3%	-----
Anjum et al. [90]	PD=4 Controls=41	LEAPD	5-10-fold (10x), LOO	PSD (LPC coefficients from single channel classification)	channel >83% acc	AUC: 85.2%	#1: TP8 #2: P6 #3: FC5 #4: PO8 #5: O2 , #6: CP5	

<b>Xu et al. [96]</b>	PD=20 Controls=20	PDRNN	Stratified 10-fold	Raw signal, 1,588 epochs (non-linear extraction)	-----	Sensitivity: 84.8% Specificity: 91.8%	-----
<b>Waninger et al. [87]</b>	PD=20 Controls=20	IDFA	Not specified	Absolute and relative PSD (1–40 Hz), WC	SDA	Accuracy: 95.2%	PSD: L frontal and central ( $\theta$ , $\beta$ ), bilateral temporal ( $\theta$ , $\alpha$ , $\gamma$ ), central and R occipital ( $\beta$ , $\gamma$ ), central ( $\theta$ )
<b>Lee et al. [97]</b>	PD=21 Controls=22	KNN, SVM, RF, CNN, CRNN	5-fold nested	<i>ML</i> : mean magn, bispectr entropies/moments, PE, weighted and absolute weighted center of bispectr; <i>DL</i> : spatiotemporal (raw signal)	ANOVA, $\alpha=0.05$	AUC (CRNN): 99.2%	-----
<b>Khare et al. [98]</b>	PD=15 Controls=16	RF, KNN, SVM, ANN, LSSVM	10-fold	From each subband, extraction of: HE, HFD, HJC and HM	Kruskal Wallis test	AUC (LSSVM): 97% medication off 98.6% medication on	Four extracted features: fractal (HE, HFD) and complexity (HJC and HM)
<b>Lee et al. [86]</b>	PD=41 Controls=41	XG/CatBoost LightGBM	9-fold LOO, Iowa (external)	Hjorth parameters (activity, mobility, complexity) from 62 Ag/AgCl channels	-----	AUC: 91% (CatBoost)	Highest discriminant ability: low $\theta$ (4–6 Hz) and high $\delta$ (2–4 Hz)

	<b>Chang et al. [92]</b>	PD=99 Controls=59	L/GBoost, Bag, DT, SVM, NB, KNN	10-fold + external (n=64)	HHSA features (non- linear)	X	AUC: 90% (Bag)	1) FZ, AM 1–2 Hz, Cf 128–256 2) F8, AM frequency 2–4 Hz, Cf 4–8 Hz; 3) C3, AM 8–16 Hz, Cf 32–64 Hz
<b>PD → CD / dementia</b>	<b>Cozac et al. [99]</b>	PD=37 (3 yrs) Severe CD	RF	Internal	Motor, cognitive, psychiatric, EEG (GRMP in $\delta$ , $\theta$ , 1, $\alpha$ 2, $\beta$ )	-----	AUC: 75%	GRMP $\theta$
	<b>Arnaldi et al. [77]</b>	PD=54 (4.8 yrs) CogSta=33 CogWorse=21	LDA	LOO	Clinical, cognitive, rs- EEG, DAT, SPECT	Step-wise LR	Accuracy: 82%	Baseline posterior mean frequency
<b>iRBD → PD or DLB</b>	<b>Ruffini et al. [138]</b>	iRBD=121 (1- 10 yrs) PD=14 DLB=13 Still iRBD=94 Controls=91	DCNN: 5- layer RNN	LPO	A spectrogram for each of the 14 EEG channels	-----	AUC: 78% HC+iRBD vs. PD+DLB DCNN: (p4)	P4 channel spectrogram

ACC, anterior cingulate cortex; Ag/AgCl, Ag-AgCl reference electrode; ANN, artificial neural network; ANOVA, Analysis of Variance; AUC, Area Under the Curve; Bag, Bagging algorithm; CD, cognitive decline; CNN, convolutional neural network; CogSta, cognitively stable; CogWorse, cognitively worsened; CRNN, Convolutional Recurrent Neural Network; DAT, dopamine transporter imaging; DL, Deep Learning; DLB, dementia with Lewy bodies; DT, Decision Tree; EEG, electroencephalography; EOG, electrooculography; FKNN,

fuzzy K-nearest neighbor; FL-TR, frontal left and temporal right; FZ, frontal cephalic channel; G= Group; GRMP, global relative median power; HE, Hurst exponent; HFD, Higuchi fractal dimension; HHSA, Holo-Hilbert Spectral Analysis; HJC, Hjorth complexity; HM, mobility; HOS, Higher-order spectra; iRBD, idiopathic rapid eye movement sleep behavior disorder; J48, decision tree classification algorithm based on Iterative Dichotomiser 3; K-means, K-means clustering algorithm; KNN, K-nearest neighbor; L, left; LDA, linear discriminant analysis; IDFA, linear discriminant function analysis; LEAPD, Linear-predictive-coding EEG Algorithm for PD; LightGBM, Light Gradient Boosting Machine; LOO, leave-one-out; LPC, linear predictive coding; LR, Logistic Regression; LSSVM, least squares support vector machine; MCI, mild cognitive impairment; ML, Machine Learning; MST, minimum spanning tree; NB, naïve bayes; O\_CCA, optimal center constructive covering algorithm; PCC, posterior cingulate cortex; PD, Parkinson's Disease; PD with PS, PD with a prodromal state; PD-MCI, PD with MCI; PD-nMCI, PD without MCI; PD-nRBD, PD without RBD; PD-RBD, PD with RBD; PD-RBE, PD with REM behavioral events; PDRNN, pooling-based deep recurrent neural network; PLI, phase lag index; PNN, probabilistic neural network; PSD, power spectral density; PTE, phase transfer entropy; R, right; RF, Random Forest; ROI, region of interest; rs-EEG, resting-state EEG; SDA, Stepwise discriminant analysis; SPECT, single-photon emission computed tomography; SVM-RBF, SVM with radial basis function; SVM, Support Vector Machine; TL-PR, between the temporal left and the parietal right regions; WC, wavelet coefficients; XG/CatBoost, variants of gradient boosting algorithms; Y, years;  $\alpha$ , alpha;  $\alpha_1$ , alpha1;  $\alpha_2$ , alpha2;  $\beta$ , beta;  $\gamma$ , gamma;  $\delta$ , delta;  $\theta$ , theta.



**Table 3** – Reviewed Cognitive studies.

		Cognitive						
		Population	Classifier	CV	Features	Feature selection	Best decoding performance	Most discriminant feature
PD vs Control	Souillard-Mandar et al. [85]	PD=126 Controls=406	C4.5, CART, SVM Gaussian, RF, Boosted DT, LR	Stratified 5-fold	CDT ( <i>pen stroke</i> ) features : five sets from digital-pen, clinician, operationalized clinician, simplest	-----	AUC: -RF: 91%; -Regularized LR: 91%	-----
	Byeon et al. [82]	PD-MCI=45 PD-NC=51	RF	Internal	Age, gender, education, handedness, family history (dementia/PD), cigarette pack/coffee/pesticide exposure, disease history, PD related motor signs, RBDs, cognitive (K-MMSE, K-MoCA, GDS, CDR, K-IADL, UPDRS, H&Y S&E ADL)	-----	Accuracy: 65.6%	#1: CDR sum of boxes #2: UPDRS motor #3: K-MMSE #4: K-MoCA
PD vs PD with PS	Jeon et al. [83]	PD-NC=101 PD-CI=101	RF, LR, SVM	80 :20	MoCA domain scores, cognitive complaints, SGDS	-----	Accuracy (RF) : 87%	MoCA score with cognitive complaints (0.87)

PD vs PDD/DLB	<b>Chong-Wen et al. [81]</b>	PD-RBD=69 PD- nRBD=100	RF, DT, LR	70% train, 30% test, 10-fold	PQSI, SCOPA-AUT, MoCA, MMSE, age, LEDD, PDCourse, ESS, NMSQ, diseasetype, RLSRS, HAMD, PDOnstage, UPDRS III, UPDRS mvt score, UPDRS I spirit, HAMA, H&Y, Educationlevel, mvtdisorder, Gender	-----	AUC (RF): 80.7%	PQSI (#1), SCOPA-AUT, MoCA, MMSE, AGE, LEDD, PD- course, UPDRS total score, ESS, NMSQ, disease type, RLSRS, HAMD, UPDRS III and PDOnstage
	<b>Ortelli et al. [84]</b>	PD=400	15 algorithmic methods	k-fold, test/hold out set	7 pre-screened candidate predictors: CoMDA score, gender, age, disease, disease duration, education, and L1 score.	value of Informati on Gain	AUC (QDA CoMDA-ML): 81%	#1) CoMDA #2) age*CoMDA and Education*CoMDA
	<b>Filoteo et al. [104]</b>	PDD=24 DLB=24	DFA	42 PDD new sample	CVLT raw scores: recall level, learning, recall error, recognition	effect size, redundan ce, or DI	Accuracy (with best features): 78.6%	CVLT: short-delay cued recall, percentage of perseveration errors, and List B recall
	<b>Dujardin et al. [106]</b>	PD=147 PDD=41	Multivariabl e LR	Bootsrap resampli ng	MMSE≤26, word generation≤7, 5-words test free recalled words, pill questionnaire score, age≥70	stepwise selection	AUC: 93%	-----

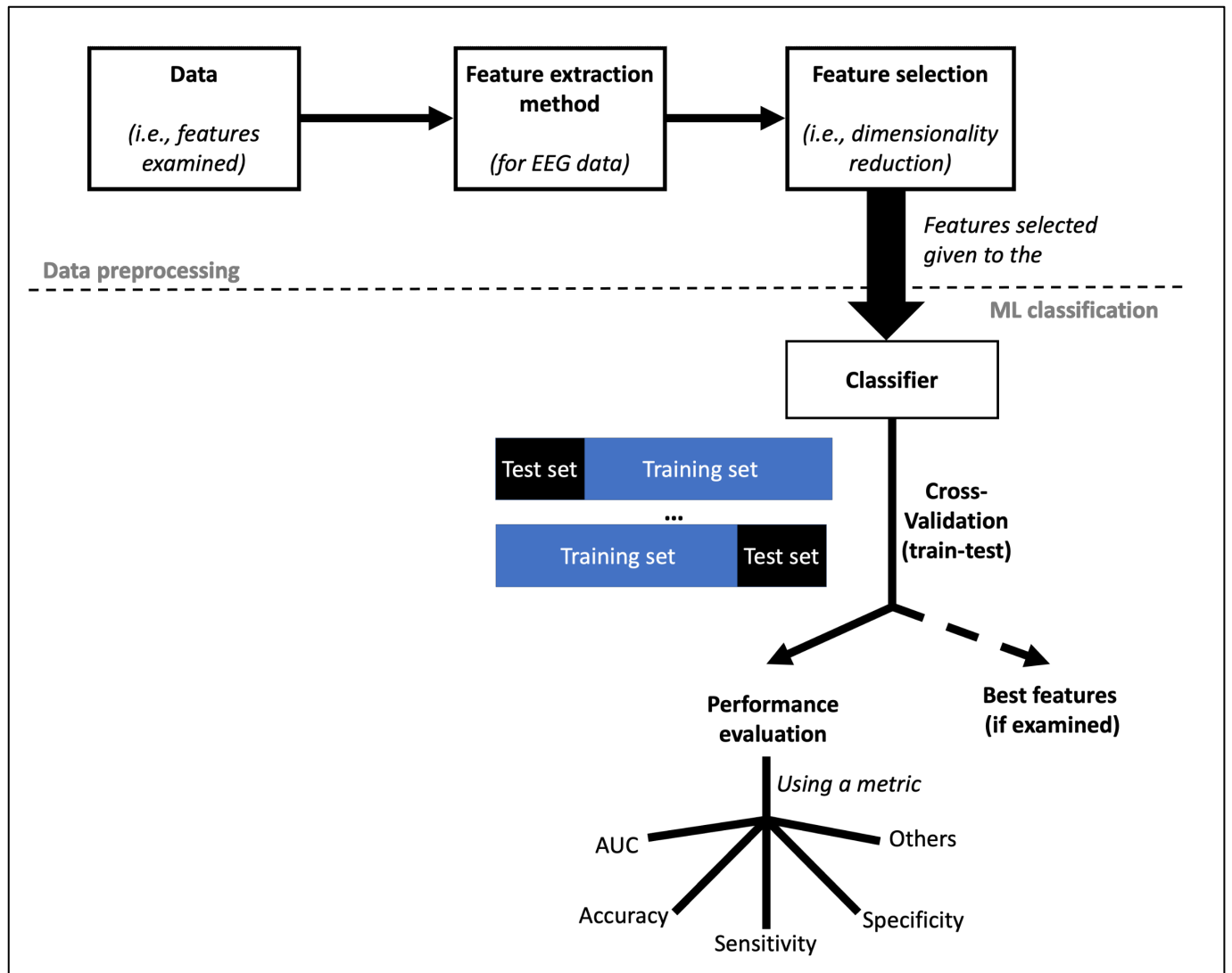
<b>Lee et al. [105]</b>	PDD=102 PD-ND=163	Binary LR	New sample (PDD=53; PD-ND=84)	PDD-5S items, category verbal fluency, 3-word recall of the MMSE (recall, serial 7's, pentagon drawing)	binary LR, forward selection	AUC: 87%	5-item questionnaire (score $\geq 3$ ), five screening questions (score= 1–2) and Category verbal fluency (score $\leq 10$ items)
<b>Byeon et al. [107]</b>	EOPD=66 PD-NC=276	RF, DA, Naive Bayes	Internal	K-MMSE, K-MoCA, Sum of boxes in CDR score, Global CDR score, Motor UPDRS, K-IADL, Total UPDRS, S&E ADL, H&Y staging, Late motor complications, coffee, Education, cigarettes pack year, BDI, tremor, postural instability, rigidity, gender, RBDs, hypertension, handness, diabetes, hyperlipidemia, Carbon monoxide poisoning, Family PD/dementia history, bradykinesia, manganese poisoning, TBI, atrial fibrillation	-----	AUC (RF): 89.5%	RF: #1: K-MMSE score #2: K-MoCA #3: CDR sum of boxes #4: CDR global score #5: UPDRS (motor) #6: KIADL
<b>Byeon et al. [108]</b>	PDD=328	8 SVR	5-fold	REM, SBDs, PD-related motor signs, demographic variables, disease history, PD family history, ADL score, K-MoCA	-----	MSE (SVR-RBF with nu parameter; all variables): 7.8%	K-IADL, UPDRS (total, motor), tremor, postural instability, age, age

					score, K-MMSE score, K-IADL score, UPDRS (total and motor scores), H&Y stage			at PD diagnosis, education (high school graduation or +), PD family history, cigarettes pack year (21–40), coffee, TBI, atrial fibrillation, RBD, depression
<b>PD → CD / dementia</b>	<b>Arnaldi et al. [77]</b>	PD=54 CogSta=33 CogWorse=21	LDA	LOO	Clinical, cognitive (language, attention and working memory, executive functions, visuospatial, memory, and MCI), DAT, SPECT and rs-EEG	step-wise LR	Accuracy: -TMT A: 63% -Phonological verbal fluency: 65% -Age: 68% -TMT B: 72% -MAH caudate SBR: 80% -Posterior qEEG MF: 82%	Posterior qEEG MF
	<b>Hogue et al. [101]</b>	PD=351 ECD=38	Binary LR	Bootstrap resampling	Demographic, disease history, motor, non-motor (HVL, JOLO, semantic fluency (animals), WAIS-III L-N sequencing, SDMT)	bivariate LR	Optimism-adjusted c-statistic (5 best features): 75%	Verbal memory retention, right-sided bradykinesia, education, subjective report of cognitive impairment, RBD

<b>Salmanpour et al. [102]</b>	184	LOLIMOT,	65% train, 5%, validation, 30% test	93 features at yrs 0 and 1: UPDRS, MoCA, age/sex, disease duration, RBD, ESS, UPSIT, SDMT, LNS, HVLT, BJLOT, STAIA, QUIP, SCOPA, MSEADL, neurological exam, vital signs, socio-economics, and family history	NSGAI	Error: 168% (LOLIMOT, 6 best features)	MoCA y 0. and y1, RBD Questionnaire y1, LNS Number 4 y0, LNS Number 3 y1, STAIA y0
	PD→MoCA score prediction at year 4	RBF, MLP-BP, LASSOLAR, RF, RNN, BRR, DT, PAR, TS					
<b>Horne et al. [103]</b>	202 PD (PD-NC=121; PD-MCI=91) → 51 PDD (43 PD-MCI, 8 PD-NC)	Bayesian regression	LOO-IC	Neuropsychiatric (NPI total score, hallucinations, anxiety), global cognitive (Z) score*, age, motor score	ROC	Delta ELPD: -global cognitive score): 36 - age: 11	Global cognitive score Age

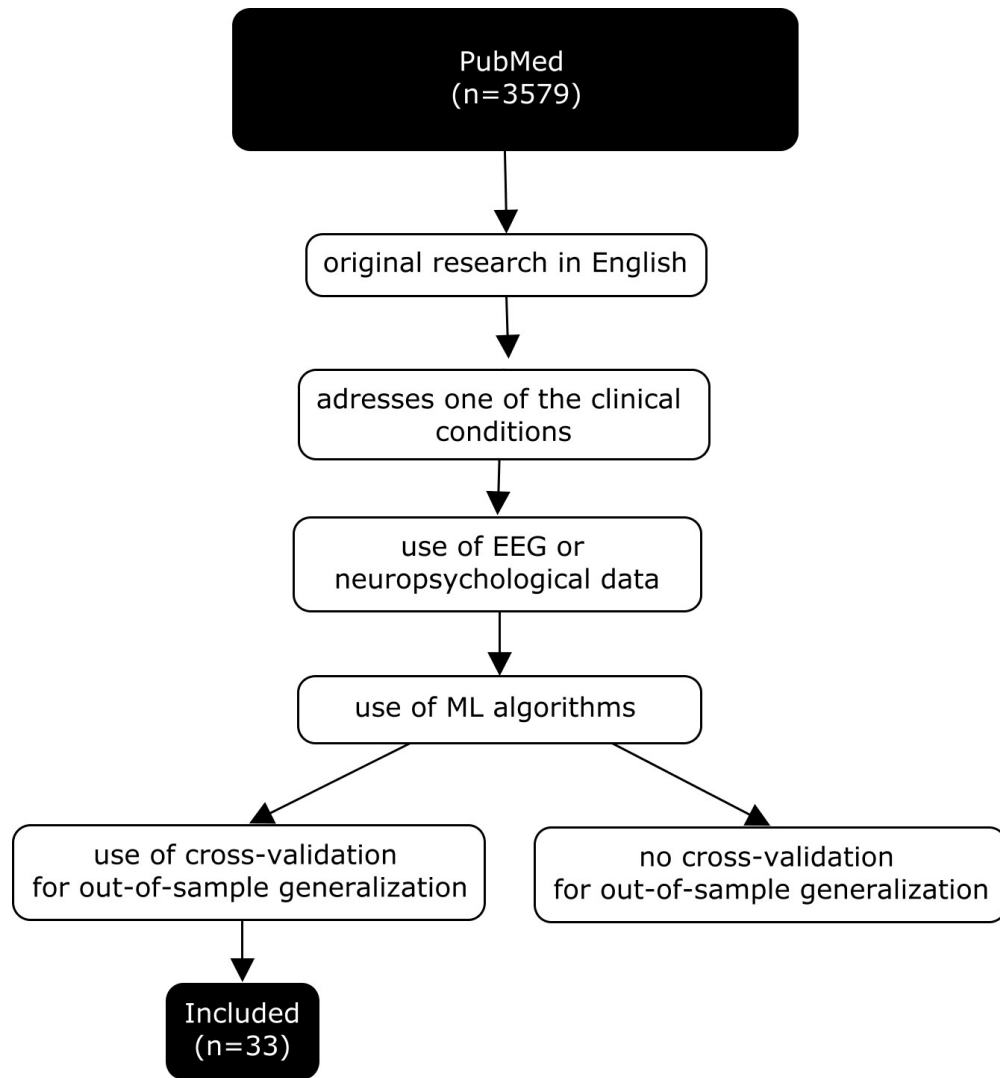
ADL, Schwab and England Activities of Daily Living; AUC, Area Under the Curve; BDI, Beck depression inventory; BJLOT, Benton Judgment of Line Orientation; BRR, bayesian ridge regression; C4.5, CART, SVM Gaussian; CDR, Clinical Dementia Rating; CDT, Clock Drawing Test; CogSta, cognitively stable; CogWorse, cognitively worsened; CoMDA, Cognition in Movement Disorders Assessment; CVLT, California Verbal Learning Test; DAT, dopamine transporter imaging; DFA, Discriminant function analyses; DI, distinct impairment; DLB, dementia with Lewy bodies; DT, Decision Tree; ECD, early cognitive decline; ELPD, the expected log pointwise predictive density; ESS, Epworth Sleeping Scale score; GDS, Geriatric Depression Scale; H&Y, Hoehn and Yahr stage;

HAMA, the 14-item Version of the Hamilton Anxiety Scale; HAMD, Hamilton Depression Scale; HVLT, Hopkins Verbal Learning Test Revised; IC, impaired cognition; JOLO, Benton Judgment of Line Orientation; K-IADL, Korean Instrumental Activities of Daily Living; K-MMSE, Korean Mini-Mental State Examination; K-MoCA, Korean Montreal Cognitive Assessment; L1, level 1; LASSOLAR, least absolute shrinkage and selection operator – least angle regression; LDA, linear discriminant analysis; LEDD, levodopa equivalent daily dose; LNS, Letter Number Sequencing; LOLIMOT, local linear model trees; LOO-IC, leave-one-out information criterion; LOO, leave-one-out; LR, Logistic Regression; MAH, more affected hemisphere; MCI, mild cognitive impairment; MF, mean frequency; MLP-BP, multilayer perceptron-back propagation; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; MSEADL, modified Schwab & England Activities of Daily Living; NC, normal cognition; NMSQ, Nonmotor Symptom Quest; PAR, passive aggressive regression; PD, Parkinson’s Disease; PD-CI, PD with cognitive impairment; PD-MCI, PD with mild cognitive impairment; PD-NC, PD with normal cognition; PD-ND, nondemented PD; PD-nRBD, PD without RBD; PD-RBD, PD with RBD; PDD-5S, brief screening tool for PDD; PDD, PD with dementia; PQSI, Pittsburgh Sleep Quality Index; QDA, Quadratic Discriminant Analysis; qEEG, quantitative electroencephalography; QUIP, Questionnaire for Impulsive-Compulsive Disorders; RBD, REM Sleep Behavior Disorder; RBF, radial basis function; REM, rapid eye movement; RF, Random Forest; RLSRS, Restless Leg Syndrome Rating Scale; RNN, recurrent neural network; rs-EEG, resting-state EEG; S&E ADL, Schwab & England Activities of Daily Living; SBD, Sleep behavior disorders; SBR, specific to nondisplaceable binding ratios; SCOPA, Scales for Outcome in Parkinson’s Disease; SCOPA-AUT, Scales for Outcomes in Parkinson's disease; SD, standard deviation; SDMT, Symbol Digit Modalities Test; SGDS, short version of the Geriatric Depression Scale; SPECT, single-photon emission computed tomography; STAIA, State-Trait Anxiety Inventory for Adults; SVR, support vector machine regression; TBI, traumatic brain injury; TMT B, Trail Making Test part B; UPDRS, Unified Parkinson’s Disease Rating Scale; WAIS-III L-N sequencing, WAIS-III, Wechsler Adult Intelligence Scale-III letter-number sequencing; y, year.



**Figure 1 - Overview of the typical machine learning pipeline examined in this review.**

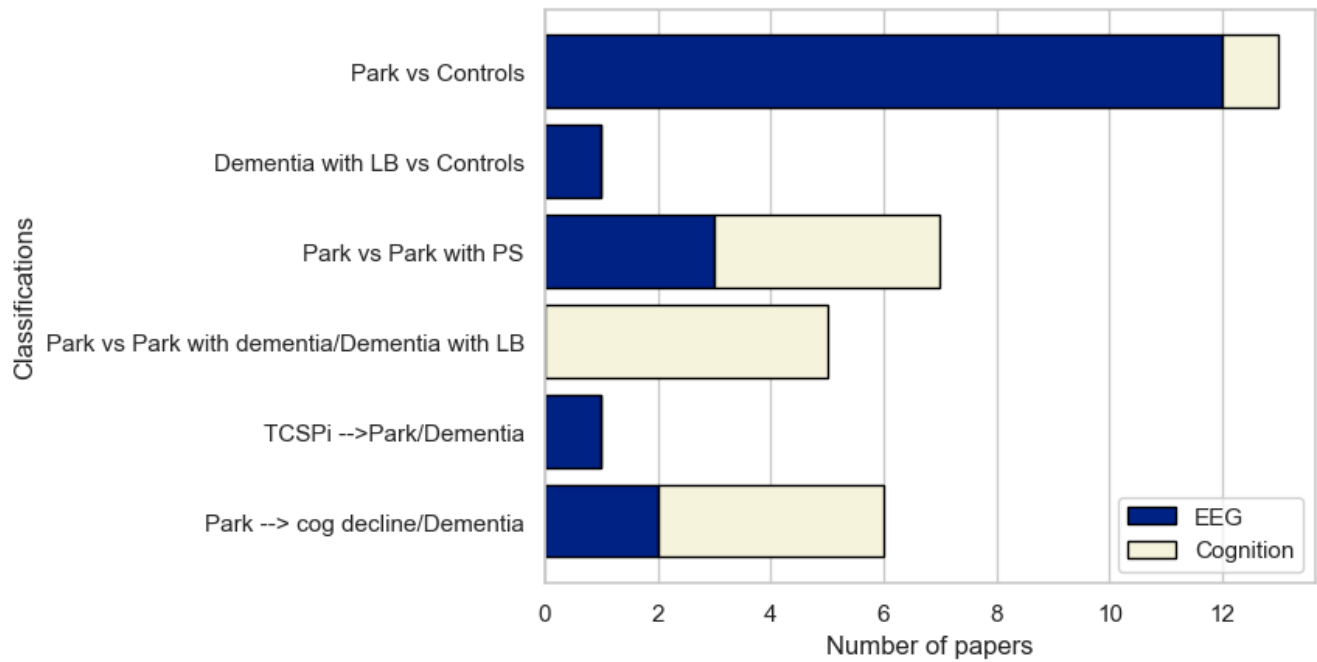
AUC, Area Under the Curve; EEG, Electroencephalography; ML, Machine Learning.



**Figure 2 – Systematic Review flow diagram of selected studies.**

EEG, electroencephalography; ML, Machine Learning.



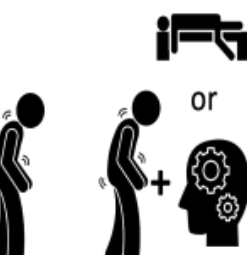







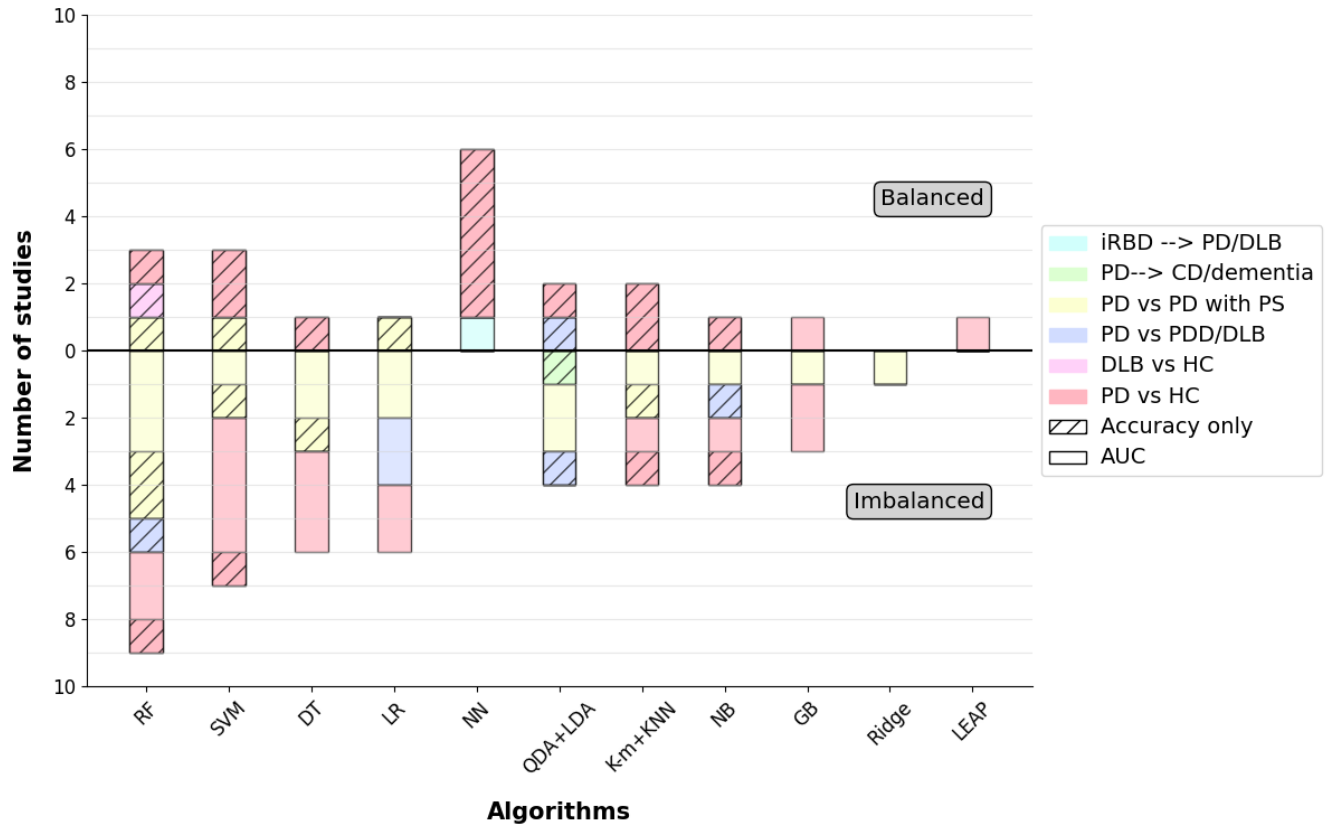
**Figure 3 – Number of EEG and cognitive papers for each classification.** CD, cognitive decline; DLB, dementia with Lewy bodies; EEG, electroencephalography; HC, healthy controls; iRBD, idiopathic rapid eye movement sleep behavior disorder; PD, PD, Parkinson’s Disease; PD with PS, PD with a prodromal state; PDD, PD with dementia.

**Classification**

**Notable features ( in EEG ● and Cognition ● )**

<p><b>DLB vs HC</b></p> 	<p><math>\theta / \alpha</math> ratio</p> <p>(Dauwan et al., 2016)</p>						
<p><b>PD vs HC</b></p> 	<p>low <math>\theta</math> and high <math>\delta</math></p> <p>(Lee et al., 2022)</p>	<p>PSD :</p> <p>1) TP8, 2) P6, 3) FC5</p> <p>(Anjum et al., 2020)</p>	<p>L and R parahipp <math>\theta</math> and <math>\gamma</math>; motor <math>\theta</math>, <math>\beta</math>, <math>\gamma</math>; dorsal ACC <math>\theta</math>, <math>\alpha</math>, <math>\beta</math>, <math>\gamma</math>; subgenual ACC <math>\gamma</math>; R insula <math>\theta</math>; PCC <math>\theta</math>, <math>\beta</math>, <math>\gamma</math></p> <p>(Vanneste et al., 2018)</p>	<p>Complexity (n=3)</p> <p>(Liu et al., 2017; Yuvaraj et al., 2018; Khare et al., 2021)</p>			
	<p><math>\theta</math> power temporal-L, <math>\alpha / \theta</math> central-L</p> <p>(Chaturvedi et al., 2017)</p>	<p>PSD : <math>\theta</math> and <math>\beta</math> in the L and C frontal; <math>\theta</math>, <math>\alpha</math>, and <math>\gamma</math> in bilateral temporal; <math>\beta</math> and <math>\gamma</math> in C and R occipital</p> <p>(Waninger et al., 2020)</p>	<p>Connectivity FZ, 1-2 -- 128/256 Hz F8, 2-4 --4-8 Hz C3, 8-16--32-64 Hz</p> <p>(Chang et al., 2022)</p>	<p>raw EEG signal (neural networks) (n=3)</p> <p>(Oh et al., 2018; Xu et al., 2020; Lee et al., 2021)</p>			
<p><b>PD vs PD with PS (RBD or MCI)</b></p> 	<p>relative <math>\theta</math> and relative <math>\beta</math> for all regions (except occipital)</p> <p>(Betrouni et al., 2019) MCI</p>	<p>1) PLI <math>\delta</math> frontal-L, temporal-R 2) PLI <math>\beta</math> temporal-L, parietal-R</p> <p>(Chaturvedi et al., 2019) MCI</p>	<p>1) Micro-sleep 2) REM <math>\delta</math>1/spectrum + max peak <math>\delta</math> 3) Complexity</p> <p>(Cesari et al., 2021) RBD</p>	<p>PSQI</p> <p>(Chong-Wen et al., 2022) RBD</p>	<p>MoCA + cognitive complaints</p> <p>(Jeon et al., 2022) MCI</p>	<p>1) CDR 2) UPDRS motor 3) K-MMSE</p> <p>(Byeon et al., 2020) MCI</p>	<p>1) CoMDA score 2) age*CoMDA Education* CoMDA</p> <p>(Ortelli et al., 2022) MCI</p>
<p><b>PD vs PDD/DLB</b></p> 	<p>CVLT</p> <p>(Filoteo et al., 2009)</p>	<p>five screening questions + 5 item questionnaire + verbal fluency</p> <p>(Lee et al., 2016)</p>	<p>1) MMSE 2) MoCA 3) CDR</p> <p>(Byeon et al., 2020)</p>	<p>MMSE + word generation + Five word test + pill questionnaire + age</p> <p>(Dujardin et al., 2010)</p>	<p>IADL, UPDRS, tremor, postural instability, age, education, family PD, cigarettes, coffee, TBI, atrial fibrillation, RBD and depression</p> <p>(Byeon et al., 2021)</p>		
<p><b>PD → CD/dementia</b></p> 	<p>GRMP <math>\theta</math></p> <p>(Cozac et al., 2016)</p>	<p>posterior mean frequency</p> <p>(Arnaldi et al., 2017)</p>	<p>TMT part B</p> <p>(Arnaldi et al., 2017)</p>	<p>MoCA, RBD Questionnaire, Letter Number Sequencing and STAI</p> <p>(Salmanpour et al., 2019)</p>	<p>verbal memory, bradykinesia-R, years of education, subjective report of CI and RBD</p> <p>(Hogue et al., 2018)</p>	<p>Global cognitive score</p> <p>(Horne et al., 2021)</p>	
<p><b>iRBD → PD or DLB</b></p> 	<p>Time frequency representation on parietal (P4)</p> <p>(Ruffini et al., 2019)</p>						

**Figure 4 – Notable features for each classification.** ACC, anterior cingulate cortex; C3, central on the left side; CD, cognitive decline; CDR, Clinical Dementia Rating; CI, cognitive impairments; CoMDA, Cognition in Movement Disorders Assessment; CVLT, California Verbal Learning Test; DLB, dementia with Lewy bodies; EEG, electroencephalography; F8, frontal on the right side; FC5, fronto-central on the left side; FZ, frontal cephalic channel; GRMP, global relative median power; HC, healthy controls; IADL, Instrumental Activities of Daily Living; iRBD, idiopathic rapid eye movement sleep behavior disorder; K-MMSE, Korean Mini-Mental State Examination; L, left; MCI, Mild Cognitive Impairments; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; P6, parietal on the right side; Parahipp, parahippocampal; PCC, posterior cingulate cortex; PD with PS, PD with a prodromal state PD, Parkinson’s Disease; PDD, PD with dementia; PLI, phase lag index; PSD, Power Spectral Density; PSQI, Pittsburgh Sleep Quality Index; R, right; RBD, rapid eye movement sleep behavior disorder; STAI, State-Trait Anxiety Inventory for Adults; TBI, traumatic brain injury; TMT, Trail Making Test; TP8, temporal-parietal on the right side; UPDRS, Unified Parkinson’s Disease Rating Scale;  $\alpha$ , alpha;  $\beta$ , beta;  $\gamma$ , gamma;  $\delta$ , delta;  $\theta$ , theta.



**Figure 5 – Frequency of algorithms tested across studies for various classification problems using AUC and/or Accuracy metrics.** AUC, Area Under the Curve; CD, cognitive decline; DLB, dementia with Lewy bodies; DT, Decision Trees; EEG, electroencephalography; GB, Gradient Boosting; HC, healthy controls; iRBD, idiopathic rapid eye movement sleep behavior disorder; K-m+KNN, K-means and K-Nearest Neighbors; LEAP, Linear-predictive-coding EEG Algorithm for PD; LR, Logistic Regression; NB, Naïve Bayes; NN, Neural Networks; PD with PS, PD with a prodromal state; PD, PD, Parkinson’s Disease; PDD, PD with dementia; QDA+LDA, Quadratic Linear Discriminant Analysis and Linear Discriminant Analysis; RF, Random Forest; SVM, Support Vector Machine. We excluded certain studies from our analysis for the following reasons: 1) Studies lacking both AUC and accuracy metrics, including Horne 2021 (NB), Salmanpour 2019 (LOLIMOT, NN, LDA, RF, NB), Hogue 2019 (LR), and Byeon 2021 (SVM); 2) Studies with no information on dataset balance, such as Cozac 2016 (RF). However, the study by Xu 2020 was incorporated into the accuracy-based studies because we could derive the accuracy from the reported sensitivity and specificity. The algorithms from Arnaldi et al., 2017 (appears in EEG and Cognition categories) were counted once. Note that more than 38.46% of algorithms tested on AUC were also evaluated on the accuracy metric (i.e., 7 studies; see supplementary materials).

## References

1. Alzheimer's Disease International, M.P., Anders Wimo, Maëlen Guerchet, Gemma-Claire Ali, Yu-Tzu Wu, Matthew Prina, World Alzheimer Report 2015: The Global Impact of Dementia - an analysis of prevalence, incidence, cost and trends. 2015, Alzheimer's Disease International (ADI): London.
2. Zaccai, J., C. McCracken, and C. Brayne, A systematic review of prevalence and incidence studies of dementia with Lewy bodies. *Age Ageing*, 2005. 34(6): p. 561-6.
3. Mueller, C., et al., The prognosis of dementia with Lewy bodies. *Lancet Neurol*, 2017. 16(5): p. 390-398.
4. Alves, G., et al., Epidemiology of Parkinson's disease. *J Neurol*, 2008. 255 Suppl 5: p. 18-32.
5. Hely, M.A., et al., The Sydney multicenter study of Parkinson's disease: the inevitability of dementia at 20 years. *Mov Disord*, 2008. 23(6): p. 837-44.
6. Miglis, M.G., et al., Biomarkers of conversion to  $\alpha$ -synucleinopathy in isolated rapid-eye-movement sleep behaviour disorder. *The Lancet Neurology*, 2021. 20(8): p. 671-684.
7. Fereshtehnejad, S.M., et al., Evolution of prodromal Parkinson's disease and dementia with Lewy bodies: a prospective study. *Brain*, 2019. 142(7): p. 2051-2067.
8. Bzdok, D. and J.P. Ioannidis, Exploration, inference, and prediction in neuroscience and biomedicine. *Trends in neurosciences*, 2019. 42(4): p. 251-262.
9. Abu-Mostafa, Y.S., M. Magdon-Ismail, and H.-T. Lin, *Learning from data*. Vol. 4. 2012: AMLBook New York.
10. Bzdok, D., *Classical Statistics and Statistical Learning in Imaging Neuroscience*. *Front Neurosci*, 2017. 11: p. 543.
11. Bzdok, D. and A. Meyer-Lindenberg, *Machine Learning for Precision Psychiatry: Opportunities and Challenges*. *Biol Psychiatry Cogn Neurosci Neuroimaging*, 2018. 3(3): p. 223-230.
12. Myszczyńska, M.A., et al., Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat Rev Neurol*, 2020. 16(8): p. 440-456.
13. Tautan, A.M., B. Ionescu, and E. Santarnecchi, Artificial intelligence in neurodegenerative diseases: A review of available tools with a focus on machine learning techniques. *Artif Intell Med*, 2021. 117: p. 102081.

14. Belic, M., et al., Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease-A review. *Clin Neurol Neurosurg*, 2019. 184: p. 105442.
15. Ij, H., Statistics versus machine learning. *Nat Methods*, 2018. 15(4): p. 233.
16. Bzdok, D. and B.T.T. Yeo, Inference in the age of big data: Future perspectives on neuroscience. *Neuroimage*, 2017. 155: p. 549-564.
17. Goodfellow, I., Bengio, Y., & Courville, A, *Deep learning*. 2016: MIT press.
18. Breiman, L., *Statistical Modeling: The Two Cultures* (with comments and a rejoinder by the author). *Statistical Science*, 2001. 16(3).
19. Domingos, P., A few useful things to know about machine learning. *Communications of the ACM*, 2012. 55(10): p. 78-87.
20. Trambaiolli, L.R., et al., Improving Alzheimer's disease diagnosis with machine learning techniques. *Clin EEG Neurosci*, 2011. 42(3): p. 160-5.
21. Khatri, U. and G.R. Kwon, An Efficient Combination among sMRI, CSF, Cognitive Score, and APOE epsilon4 Biomarkers for Classification of AD and MCI Using Extreme Learning Machine. *Comput Intell Neurosci*, 2020. 2020: p. 8015156.
22. Geraedts, V.J., et al., Clinical correlates of quantitative EEG in Parkinson disease: A systematic review. *Neurology*, 2018. 91(19): p. 871-883.
23. Postuma, R.B., et al., MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord*, 2015. 30(12): p. 1591-601.
24. Todorova, A., P. Jenner, and K. Ray Chaudhuri, Non-motor Parkinson's: integral to motor Parkinson's, yet often neglected. *Pract Neurol*, 2014. 14(5): p. 310-22.
25. Svenningsson, P., et al., Cognitive impairment in patients with Parkinson's disease: diagnosis, biomarkers, and treatment. *The Lancet Neurology*, 2012. 11(8): p. 697-707.
26. Edition, F., *Diagnostic and statistical manual of mental disorders*. Am Psychiatric Assoc, 2013. 21(21): p. 591-643.
27. Dalrymple-Alford, J.C., et al., Characterizing mild cognitive impairment in Parkinson's disease. *Mov Disord*, 2011. 26(4): p. 629-36.
28. Litvan, I., et al., MDS Task Force on mild cognitive impairment in Parkinson's disease: critical review of PD-MCI. *Mov Disord*, 2011. 26(10): p. 1814-24.
29. Litvan, I., et al., Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Mov Disord*, 2012. 27(3): p. 349-56.

30. Aarsland, D., Cognitive impairment in Parkinson's disease and dementia with Lewy bodies. *Parkinsonism Relat Disord*, 2016. 22 Suppl 1: p. S144-8.
31. Aarsland, D. and M.W. Kurz, The epidemiology of dementia associated with Parkinson disease. *J Neurol Sci*, 2010. 289(1-2): p. 18-22.
32. Buter, T.C., et al., Dementia and survival in Parkinson disease: a 12-year population study. *Neurology*, 2008. 70(13): p. 1017-22.
33. Hobson, P. and J. Meara, Mild cognitive impairment in Parkinson's disease and its progression onto dementia: a 16-year outcome evaluation of the Denbighshire cohort. *Int J Geriatr Psychiatry*, 2015. 30(10): p. 1048-55.
34. McKeith, I.G., et al., Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium. *Neurology*, 2017. 89(1): p. 88-100.
35. Emre, M., et al., Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Mov Disord*, 2007. 22(12): p. 1689-707; quiz 1837.
36. McKeith, I.G., et al., Research criteria for the diagnosis of prodromal dementia with Lewy bodies. *Neurology*, 2020. 94(17): p. 743-755.
37. Cagnin, A., et al., Clinical and Cognitive Phenotype of Mild Cognitive Impairment Evolving to Dementia with Lewy Bodies. *Dement Geriatr Cogn Dis Extra*, 2015. 5(3): p. 442-9.
38. Sadiq, D., et al., Prodromal Dementia with Lewy Bodies and Prodromal Alzheimer's Disease: A Comparison of the Cognitive and Clinical Profiles. *J Alzheimers Dis*, 2017. 58(2): p. 463-470.
39. Sateia, M.J., International classification of sleep disorders. *Chest*, 2014. 146(5): p. 1387-1394.
40. Postuma, R.B., et al., Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study. *Brain*, 2019. 142(3): p. 744-759.
41. Genier Marchand, D., et al., Detecting the Cognitive Prodrome of Dementia with Lewy Bodies: A Prospective Study of REM Sleep Behavior Disorder. *Sleep*, 2017. 40(1).
42. Ferini-Strambi, L., et al., Neuropsychological, electrophysiological, and neuroimaging biomarkers for REM behavior disorder. *Expert Rev Neurother*, 2019. 19(11): p. 1069-1087.
43. Perez, F., et al., Risk of dementia in an elderly population of Parkinson's disease patients: a 15-year population-based study. *Alzheimers Dement*, 2012. 8(6): p. 463-9.

44. Youn, S., et al., Progression of cognitive impairments in idiopathic REM sleep behaviour disorder. *J Neurol Neurosurg Psychiatry*, 2016. 87(8): p. 890-6.
45. De Roy, J., et al., Detecting the Cognitive Prodrome of Dementia in Parkinson's Disease. *J Parkinsons Dis*, 2020. 10(3): p. 1033-1046.
46. Silva, F.L.D., EEG: origin and measurement. In *EEG-fMRI*. Berlin, Heidelberg: Springer.
47. Jasper, H.H., The ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.*, 1958. 10: p. 370-375.
48. Klem, G.H., The ten-twenty electrode system of the international federation. the internanional federation of clinical nenrophysiology. *Electroencephalogr. Clin. Neurophysiol. Suppl.*, 1999. 52: p. 3-6.
49. Mörchen, F., Time series feature extraction for data mining using DWT and DFT. 2003.
50. Hari, R., & Puce, A., *MEG-EEG Primer*. 2017: Oxford University Press.
51. Sanei, S., & Chambers, J. A. , *EEG signal processing*. 2013: John Wiley & Sons.
52. Grass, A.M. and F.A. Gibbs, A Fourier Transform of the Electroencephalogram. *Journal of Neurophysiology*, 1938. 1(6): p. 521-526.
53. Mathewson, K.J., et al., Regional electroencephalogram (EEG) alpha power and asymmetry in older adults: a study of short-term test-retest reliability. *Front Aging Neurosci*, 2015. 7: p. 177.
54. Yamada, T., & Meng, E., *Practical guide for clinical neurophysiologic testing: EEG*. 2012: Lippincott Williams & Wilkins.
55. Nunez, P.L., et al., EEG coherency. *Electroencephalography and Clinical Neurophysiology*, 1997. 103(5): p. 499-515.
56. Kaminski, M.J. and K.J. Blinowska, A new method of the description of the information flow in the brain structures. *Biol Cybern*, 1991. 65(3): p. 203-10.
57. Stam, C.J. and B.W. van Dijk, Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets. *Physica D: Nonlinear Phenomena*, 2002. 163(3-4): p. 236-251.
58. Stam, C.J., G. Nolte, and A. Daffertshofer, Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum Brain Mapp*, 2007. 28(11): p. 1178-93.



59. Lezak, M.D., Howieson, D. B., Loring, D. W., & Fischer, J. S., *Neuropsychological assessment*. 2004, USA: Oxford University Press.
60. Roebuck-Spencer, T.M., et al., *Cognitive Screening Tests Versus Comprehensive Neuropsychological Test Batteries: A National Academy of Neuropsychology Education Paperdagger*. *Arch Clin Neuropsychol*, 2017. 32(4): p. 491-498.
61. Folstein, M.F., Folstein, S. E., & McHugh, P. R., "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. . *Journal of psychiatric research*, 1975. 12(3): p. 189-198.
62. Nasreddine, Z.S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H., " The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment": Corrigendum. *J Am Geriatr Soc*, 2019. 67(9): p. 1991.
63. Elkhani, N. and R.C. Muniyandi, *Membrane computing inspired feature selection model for microarray cancer data*. *Intelligent Data Analysis*, 2017. 21: p. S137-S157.
64. Cai, J., et al., *Feature selection in machine learning: A new perspective*. *Neurocomputing*, 2018. 300: p. 70-79.
65. Mahesh, B., *Machine learning algorithms-a review*. *International Journal of Science and Research (IJSR)*. [Internet], 2020. 9: p. 381-386.
66. Muhammad, I. and Z. Yan, *Supervised Machine Learning Approaches: A Survey*. *ICTACT Journal on Soft Computing*, 2015. 05(03): p. 946-952.
67. James, G., et al., *Unsupervised learning*, in *An Introduction to Statistical Learning: with Applications in Python*. 2023, Springer. p. 503-556.
68. Singh, A., Thakur, N., & Sharma, A. , *A review of supervised machine learning algorithms*. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. Ieee. , 2016, March: p. 1310-1315.
69. Fisher, R.A., *The Use of Multiple Measurements in Taxonomic Problems*. *Annals of Eugenics*, 1936. 7(2): p. 179-188.
70. Burges, C.J.C., *Data Mining and Knowledge Discovery*, 1998. 2(2): p. 121-167.
71. Bhardwaj, A., et al., *A genetically optimized neural network model for multi-class classification*. *Expert Systems with Applications*, 2016. 60: p. 211-221.

72. Xanthopoulos, P., Pardalos, P. M., & Trafalis, T. B, Linear discriminant analysis. *Robust data mining 2013*, New York, NY.: Springer. 27-33.
73. Arisoy, E., Sainath, T. N., Kingsbury, B., & Ramabhadran, B. Deep neural network language models. in *Proceedings of the NAACL-HLT 2012 Workshop: Will We Ever Really Replace the N-gram Model? On the Future of Language Modeling for HLT*. 2012, June.
74. Thölkea, P., et al., Class imbalance should not throw you off balance: Choosing classifiers and performance metrics for brain decoding with imbalanced data. 2022.
75. Lever, J., Classification evaluation: It is important to understand both what a classification metric expresses and what it hides. *Nature methods*, 2016. 13(8): p. 603-605.
76. Tholke, P., et al., Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *Neuroimage*, 2023. 277: p. 120253.
77. Arnaldi, D., et al., Prediction of cognitive worsening in de novo Parkinson's disease: Clinical use of biomarkers. *Mov Disord*, 2017. 32(12): p. 1738-1747.
78. Betrouni, N., et al., Electroencephalography-based machine learning for cognitive profiling in Parkinson's disease: Preliminary results. *Mov Disord*, 2019. 34(2): p. 210-217.
79. Chaturvedi, M., et al., Phase lag index and spectral power as QEEG features for identification of patients with mild cognitive impairment in Parkinson's disease. *Clin Neurophysiol*, 2019. 130(10): p. 1937-1944.
80. Cesari, M., et al., A data-driven system to identify REM sleep behavior disorder and to predict its progression from the prodromal stage in Parkinson's disease. *Sleep Med*, 2021. 77: p. 238-248.
81. Chong-Wen, W., L. Sha-Sha, and E. Xu, Predictors of rapid eye movement sleep behavior disorder in patients with Parkinson's disease based on random forest and decision tree. *PLoS One*, 2022. 17(6): p. e0269392.
82. Byeon, H., Is the Random Forest Algorithm Suitable for Predicting Parkinson's Disease with Mild Cognitive Impairment out of Parkinson's Disease with Normal Cognition? *Int J Environ Res Public Health*, 2020. 17(7).

83. Jeon, J., et al., Accuracy of Machine Learning Using the Montreal Cognitive Assessment for the Diagnosis of Cognitive Impairment in Parkinson's Disease. *J Mov Disord*, 2022. 15(2): p. 132-139.
84. Ortelli, P., et al., Optimization of cognitive assessment in Parkinsonisms by applying artificial intelligence to a comprehensive screening test. *NPJ Parkinsons Dis*, 2022. 8(1): p. 42.
85. Souillard-Mandar, W., et al., Learning Classification Models of Cognitive Conditions from Subtle Behaviors in the Digital Clock Drawing Test. *Mach Learn*, 2016. 102(3): p. 393-441.
86. Lee, S.B., et al., Predicting Parkinson's disease using gradient boosting decision tree models with electroencephalography signals. *Parkinsonism Relat Disord*, 2022. 95: p. 77-85.
87. Waninger, S., et al., Neurophysiological Biomarkers of Parkinson's Disease. *J Parkinsons Dis*, 2020. 10(2): p. 471-480.
88. Angelakis, E., et al., Peak alpha frequency: an electroencephalographic measure of cognitive preparedness. *Clin Neurophysiol*, 2004. 115(4): p. 887-97.
89. Chaturvedi, M., et al., Quantitative EEG (QEEG) Measures Differentiate Parkinson's Disease (PD) Patients from Healthy Controls (HC). *Front Aging Neurosci*, 2017. 9: p. 3.
90. Anjum, M.F., et al., Linear predictive coding distinguishes spectral EEG features of Parkinson's disease. *Parkinsonism Relat Disord*, 2020. 79: p. 79-85.
91. Vanneste, S., J.J. Song, and D. De Ridder, Thalamocortical dysrhythmia detected by machine learning. *Nat Commun*, 2018. 9(1): p. 1103.
92. Chang, K.H., et al., Evaluating the Different Stages of Parkinson's Disease Using Electroencephalography With Holo-Hilbert Spectral Analysis. *Front Aging Neurosci*, 2022. 14: p. 832637.
93. Liu, G., et al., Complexity Analysis of Electroencephalogram Dynamics in Patients with Parkinson's Disease. *Parkinsons Dis*, 2017. 2017: p. 8701061.
94. Yuvaraj, R., U. Rajendra Acharya, and Y. Hagiwara, A novel Parkinson's Disease Diagnosis Index using higher-order spectra features in EEG signals. *Neural Computing and Applications*, 2016. 30(4): p. 1225-1235.
95. Oh, S.L., et al., A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Computing and Applications*, 2018. 32(15): p. 10927-10933.

96. Xu, S., et al., Using a deep recurrent neural network with EEG signal to detect Parkinson's disease. *Ann Transl Med*, 2020. 8(14): p. 874.
97. Lee, S., et al., A convolutional-recurrent neural network approach to resting-state EEG classification in Parkinson's disease. *J Neurosci Methods*, 2021. 361: p. 109282.
98. Khare, S.K., V. Bajaj, and U.R. Acharya, Detection of Parkinson's disease using automated tunable Q wavelet transform technique with EEG signals. *Biocybernetics and Biomedical Engineering*, 2021. 41(2): p. 679-689.
99. Cozac, V.V., et al., Increase of EEG Spectral Theta Power Indicates Higher Risk of the Development of Severe Cognitive Decline in Parkinson's Disease after 3 Years. *Front Aging Neurosci*, 2016. 8: p. 284.
100. Fereshtehnejad, S.M., et al., New Clinical Subtypes of Parkinson Disease and Their Longitudinal Progression: A Prospective Cohort Comparison With Other Phenotypes. *JAMA Neurol*, 2015. 72(8): p. 863-73.
101. Hogue, O., H.H. Fernandez, and D.P. Floden, Predicting early cognitive decline in newly-diagnosed Parkinson's patients: A practical model. *Parkinsonism Relat Disord*, 2018. 56: p. 70-75.
102. Salmanpour, M.R., et al., Optimized machine learning methods for prediction of cognitive outcome in Parkinson's disease. *Comput Biol Med*, 2019. 111: p. 103347.
103. Horne, K.L., et al., Neuropsychiatric Symptoms Are Associated with Dementia in Parkinson's Disease but Not Predictive of it. *Mov Disord Clin Pract*, 2021. 8(3): p. 390-399.
104. Filoteo, J.V., et al., Verbal learning and memory in patients with dementia with Lewy bodies or Parkinson's disease with dementia. *J Clin Exp Neuropsychol*, 2009. 31(7): p. 823-34.
105. Lee, W.J., et al., PDD-5S: A useful screening tool for Parkinson's disease dementia. *Parkinsonism Relat Disord*, 2016. 25: p. 85-90.
106. Dujardin, K., et al., Parkinson's disease dementia can be easily detected in routine clinical practice. *Mov Disord*, 2010. 25(16): p. 2769-76.
107. Byeon, H., Best early-onset Parkinson dementia predictor using ensemble learning among Parkinson's symptoms, rapid eye movement sleep disorder, and neuropsychological profile. *World J Psychiatry*, 2020. 10(11): p. 245-259.

108. Byeon, H., Predicting the Severity of Parkinson's Disease Dementia by Assessing the Neuropsychiatric Symptoms with an SVM Regression Model. *Int J Environ Res Public Health*, 2021. 18(5).
109. Dauwan, M., et al., Random forest to differentiate dementia with Lewy bodies from Alzheimer's disease. *Alzheimers Dement (Amst)*, 2016. 4: p. 99-106.
110. Fantini, M.L., Farini, E., Ortelli, P., Zucconi, M., Manconi, M., Cappa, S., & Ferini-Strambi, L., Longitudinal Study of Cognitive Function in Idiopathic REM Sleep Behavior Disorder. *Sleep*, 2011. 34(5): p. 619-625.
111. Terzaghi, M., et al., Cognitive performances and mild cognitive impairment in idiopathic rapid eye movement sleep behavior disorder: results of a longitudinal follow-up study. *Sleep*, 2013. 36(10): p. 1527-32.
112. Genier Marchand, D., et al., How does dementia with Lewy bodies start? prodromal cognitive changes in REM sleep behavior disorder. *Ann Neurol*, 2018. 83(5): p. 1016-1026.
113. Soikkeli, R., et al., Slowing of EEG in Parkinson's disease. *Electroencephalography and Clinical Neurophysiology*, 1991. 79(3): p. 159-165.
114. Fonseca, L.C., et al., Dementia, mild cognitive impairment and quantitative EEG in patients with Parkinson's disease. *Clin EEG Neurosci*, 2009. 40(3): p. 168-72.
115. Neufeld, M.Y., et al., EEG frequency analysis in demented and nondemented parkinsonian patients. *Dementia*, 1994. 5(1): p. 23-8.
116. Pozzi, D., et al., Quantified electroencephalographic changes in Parkinson's disease with and without dementia. *Eur J Neurol*, 1994. 1(2): p. 147-52.
117. Kamei, S., et al., Quantitative EEG analysis of executive dysfunction in Parkinson disease. *J Clin Neurophysiol*, 2010. 27(3): p. 193-7.
118. Morita, A., S. Kamei, and T. Mizutani, Relationship between slowing of the EEG and cognitive impairment in Parkinson disease. *J Clin Neurophysiol*, 2011. 28(4): p. 384-7.
119. Mostile, G., et al., Electroencephalographic lateralization, clinical correlates and pharmacological response in untreated Parkinson's disease. *Parkinsonism Relat Disord*, 2015. 21(8): p. 948-53.
120. Stanzione, P., et al., Quantitative EEG changes in non-demented Parkinson's disease patients before and during L-dopa therapy. *European Journal of Neurology*, 1996. 3(4): p. 354-362.

121. He, X., et al., Changes in theta activities in the left posterior temporal region, left occipital region and right frontal region related to mild cognitive impairment in Parkinson's disease patients. *Int J Neurosci*, 2017. 127(1): p. 66-72.
122. Eichelberger, D., et al., Correlation of Visuospatial Ability and EEG Slowing in Patients with Parkinson's Disease. *Parkinsons Dis*, 2017. 2017: p. 3659784.
123. Tanaka, H., et al., Event-related potential and EEG measures in Parkinson's disease without and with dementia. *Dement Geriatr Cogn Disord*, 2000. 11(1): p. 39-45.
124. Caviness, J.N., et al., Differential spectral quantitative electroencephalography patterns between control and Parkinson's disease cohorts. *Eur J Neurol*, 2016. 23(2): p. 387-92.
125. Janvin, C.C., D. Aarsland, and J.P. Larsen, Cognitive predictors of dementia in Parkinson's disease: a community-based, 4-year longitudinal study. *J Geriatr Psychiatry Neurol*, 2005. 18(3): p. 149-54.
126. Bonanni, L., et al., EEG comparisons in early Alzheimer's disease, dementia with Lewy bodies and Parkinson's disease with dementia patients with a 2-year follow-up. *Brain*, 2008. 131(Pt 3): p. 690-705.
127. Bonanni, L., et al., Quantitative electroencephalogram utility in predicting conversion of mild cognitive impairment to dementia with Lewy bodies. *Neurobiol Aging*, 2015. 36(1): p. 434-45.
128. Bonanni, L., et al., EEG Markers of Dementia with Lewy Bodies: A Multicenter Cohort Study. *J Alzheimers Dis*, 2016. 54(4): p. 1649-1657.
129. Briel, R.C., et al., EEG findings in dementia with Lewy bodies and Alzheimer's disease. *J Neurol Neurosurg Psychiatry*, 1999. 66(3): p. 401-3.
130. Stylianou, M., et al., Quantitative electroencephalography as a marker of cognitive fluctuations in dementia with Lewy bodies and an aid to differential diagnosis. *Clin Neurophysiol*, 2018. 129(6): p. 1209-1220.
131. Caviness, J.N., et al., Longitudinal EEG changes correlate with cognitive measure deterioration in Parkinson's disease. *J Parkinsons Dis*, 2015. 5(1): p. 117-24.
132. Latreille, V., et al., Electroencephalographic prodromal markers of dementia across conscious states in Parkinson's disease. *Brain*, 2016. 139(Pt 4): p. 1189-99.

133. Caviness, J.N., et al., Both early and late cognitive dysfunction affects the electroencephalogram in Parkinson's disease. *Parkinsonism Relat Disord*, 2007. 13(6): p. 348-54.
134. Guner, D., et al., Contribution of Quantitative EEG to the Diagnosis of Early Cognitive Impairment in Patients With Idiopathic Parkinson's Disease. *Clin EEG Neurosci*, 2017. 48(5): p. 348-354.
135. Helkala, E.L., et al., Slow-wave activity in the spectral analysis of the electroencephalogram and volumes of hippocampus in subgroups of Alzheimer's disease patients. *Behavioral Neuroscience*, 1996. 110(6): p. 1235-1243.
136. Iranzo, A., et al., Electroencephalographic slowing heralds mild cognitive impairment in idiopathic REM sleep behavior disorder. *Sleep Med*, 2010. 11(6): p. 534-9.
137. Rodrigues Brazete, J., et al., Electroencephalogram slowing predicts neurodegeneration in rapid eye movement sleep behavior disorder. *Neurobiol Aging*, 2016. 37: p. 74-81.
138. Ruffini, G., et al., Deep Learning With EEG Spectrograms in Rapid Eye Movement Behavior Disorder. *Front Neurol*, 2019. 10: p. 806.
139. Sorrentino, P., et al., Flexible brain dynamics underpins complex behaviours as observed in Parkinson's disease. *Sci Rep*, 2021. 11(1): p. 4051.
140. Foltynie, T., et al., The cognitive ability of an incident cohort of Parkinson's patients in the UK. The CamPaIGN study. *Brain*, 2004. 127(Pt 3): p. 550-60.
141. Muslimovic, D., et al., Cognitive profile of patients with newly diagnosed Parkinson disease. *Neurology*, 2005. 65(8): p. 1239-45.
142. Zgaljardic, D.J., et al., A review of the cognitive and behavioral sequelae of Parkinson's disease: relationship to frontostriatal circuitry. *Cognitive and behavioral neurology*, 2003. 16(4): p. 193-210.
143. Troster, A.I., Neuropsychological characteristics of dementia with Lewy bodies and Parkinson's disease with dementia: differentiation, early detection, and implications for "mild cognitive impairment" and biomarkers. *Neuropsychol Rev*, 2008. 18(1): p. 103-19.
144. Kerai, J.H., et al., Visuospatial transformation impairments in Parkinson's disease. *J Clin Exp Neuropsychol*, 2012. 34(10): p. 1053-64.
145. Muslimovic, D., et al., Cognitive decline in Parkinson's disease: a prospective longitudinal study. *J Int Neuropsychol Soc*, 2009. 15(3): p. 426-37.

146. Aarsland, D., et al., Mild cognitive impairment in Parkinson disease: a multicenter pooled analysis. *Neurology*, 2010. 75(12): p. 1062-1069.
147. Levy, G., et al., Memory and executive function impairment predict dementia in Parkinson's disease. *Mov Disord*, 2002. 17(6): p. 1221-6.
148. Pedersen, K.F., et al., Prognosis of mild cognitive impairment in early Parkinson disease: the Norwegian ParkWest study. *JAMA Neurol*, 2013. 70(5): p. 580-6.
149. Latreille, V., et al., Sleep spindles in Parkinson's disease may predict the development of dementia. *Neurobiology of aging*, 2015. 36(2): p. 1083-1090.
150. Williams-Gray, C.H., et al., The distinct cognitive syndromes of Parkinson's disease: 5 year follow-up of the CamPaIGN cohort. *Brain*, 2009. 132(Pt 11): p. 2958-69.
151. Williams-Gray, C.H., et al., Evolution of cognitive dysfunction in an incident Parkinson's disease cohort. *Brain*, 2007. 130(Pt 7): p. 1787-98.
152. Williams-Gray, C.H., et al., The CamPaIGN study of Parkinson's disease: 10-year outlook in an incident population-based cohort. *J Neurol Neurosurg Psychiatry*, 2013. 84(11): p. 1258-64.
153. Jacobs, D.M., et al., Neuropsychological characteristics of preclinical dementia in Parkinson's disease. *Neurology*, 1995. 45(9): p. 1691-6.
154. Klassen, B.T., et al., Quantitative EEG as a predictive biomarker for Parkinson disease dementia. *Neurology*, 2011. 77(2): p. 118-24.
155. Mahieux, F., et al., Neuropsychological prediction of dementia in Parkinson's disease. *J Neurol Neurosurg Psychiatry*, 1998. 64(2): p. 178-83.
156. Olde Dubbelink, K.T., et al., Predicting dementia in Parkinson disease by combining neurophysiologic and cognitive markers. *Neurology*, 2014. 82(3): p. 263-70.
157. Woods, S.P. and A.I. Troster, Prodromal frontal/executive dysfunction predicts incident dementia in Parkinson's disease. *J Int Neuropsychol Soc*, 2003. 9(1): p. 17-24.
158. Ye, B.S., et al., Dementia-Predicting Cognitive Risk Score and Its Correlation with Cortical Thickness in Parkinson Disease. *Dement Geriatr Cogn Disord*, 2017. 44(3-4): p. 203-212.
159. Japkowicz, N., & Shah, M., Performance evaluation in machine learning. *Machine Learning in Radiation Oncology: Theory and Applications*. 2015, Springer, Cham. p. 41-56.



## Supplementary Informations

**S1 Table** – Performance of all algorithms tested and presentation of all metrics used in review studies.

Paper	Classif	data	ML model	AUC	Accuracy	Sensitivity	Specificity	Precision	F-score	ER	Other
Ruffini 2019	iRBD --> PD/DLB/HC (HC+RBD) vs. (PD+DLB)	EEG	DCNN	78%	73%	x	x	x	x	x	x
			RNN	77%	72%	x	x	x	x	x	x
Horne 2021	PD--> PDD	Cog	BR	x	x	x	x	x	x	x	delta ELPD : 36 (SD 8)
Salmanpour 2019	184 PD--> MoCA score prediction at year 4	Cog	LOLIMOT,	x	x	x	x	x	x	1.65	x
			RBF	x	x	x	x	x	x	6.54 ± 0.48	x
			MLP-BP	x	x	x	x	x	x		x
			LASSOLAR	x	x	x	x	x	x	1.83 ± 0.13	x
			RF	x	x	x	x	x	x	x	x
			RNN	x	x	x	x	x	x	x	x
			BRR	x	x	x	x	x	x	x	x
			DT	x	x	x	x	x	x	x	x
			PAR	x	x	x	x	x	x	x	x
TS	x	x	x	x	x	x	x	x	x		
Arnaldi 2017 -n	PD CogSta vs CogWorse	Cog	LDA	x	72%	x	x	x	x	x	x
Hogue 2018	PD --> ECD	Cog	Binary LR	x	75%	x	x	x	x	x	x

Arnaldi 2017 -e	PD CogSta vs CogWorse	EEG	LDA	0.85	82%	84%	80%	x	x	x	OR : 21.33 PPV:0.76; NPV:0.87
Cozac 2016	PD --> severe CD	EEG	RF	75%	x	63%	77%	x	x	x	x
Byeon 2020	PD-MCI vs PD-NC	Cog	RF	x	65.6%	70.6%	60.0%	x	x	x	x
			Dtree	x	67.7%	51.1%	82.4%	x	x	x	x
Cesari 2021	PD-RBD vs PD-noRBD vs PD-REM-be	EEG	RF	x	81.48%	80.65%	82.61%	x	x	x	x
Jeon 2022	PD-NC vs PD-CI (MCI or dementia)	Cog	RF	x	0.87 ± 0.05	0.91	0.82	x	x	x	PPV:0.84; NPV:0.89
			LR	x	0.93 ± 0.01	0.46	0.98	x	x	x	PPV:0.77; NPV:0.94
			SVM	x	0.93 ± 0.01	0.39	0.99	x	x	x	PPV:0.83; NPV:0.93
Betrouni 2019	5 groups of PD (severities of CI)	EEG	k-NN	x	87% ± 2.8	x	x	x	x	x	x
			SVM	x	86% ± 3.5	x	x	x	x	x	x
Chaturvedi 2019	PD-MCI vs PD-noMCI	EEG	RF	0.74 ± 0.17	x	x	x	x	x	x	x
Chong-Wen 2022	PD-RBD vs PD-noRBD	Cog	RF	80.73%	83.05%	67.39%	93.06%	x	x	x	PPR:86.11; NPR: 81.71
			LR	71.62%	66.67%	60.87%	71.43%	x	x	x	PPR:63.64; NPR:68.97
			regression trees	75.10%	70.59%	56.52%	82.14%	x	x	x	PPR:72.22; NPR:69.70

Ortelli 2022	PD-MCI vs PD-NC vs PD-impaired cognition (IC)	Cog	QDA	0.7647	0.6071	x	x	0.6212	0.5976	x	recall: 0.5695
			Naive Bayes	0.7738	0.6102	x	x	0.6064	0.5907	x	recall: 0.5603
			LR	0.7861	0.6018	x	x	0.5171	0.5373	x	recall: 0.5305
			Ridge classifier	0.0000	0.5961	x	x	0.4620	0.5138	x	recall: 0.5194
			LDA	0.7731	0.5960	x	x	0.5404	0.5422	x	recall: 0.5277
			Extra Trees Classifier	0.7465	0.5816	x	x	0.5507	0.5592	x	recall: 0.5346
			RF	0.7544	0.5702	x	x	0.5384	0.5400	x	recall: 0.5140
			Ada Boost	0.7074	0.5646	x	x	0.5395	0.5454	x	recall: 0.5145
			CatBoost	0.7435	0.5560	x	x	0.5364	0.5393	x	recall: 0.5119
			DT	0.6414	0.5304	x	x	0.5343	0.5249	x	recall: 0.4993
			Light gradient boosting machine	0.7268	0.5303	x	x	0.5163	0.5174	x	recall: 0.4872
			Gradient B	0.7299	0.5302	x	x	0.5099	0.5132	x	recall: 0.4831
			extreme gradient boosting	0.7147	0.5245	x	x	0.5037	0.5071	x	recall: 0.4777
			K Neighbors	0.5992	0.4154	x	x	0.4223	0.4096	x	recall: 0.3940
SVM-Linear Kernel	0	0.3610	x	x	0.2116	0.2363	x	recall: 0.3509			
Byeon 2021	PDD of different severity	Cog	Nu-SVR-RBF	x	x	x	x	x	x	conv rate : 0.078	

Filoteo 2009	PDD vs DLB	Cog	DFA	x	81.3% on same sample	75%	87.5%	x	x	x	TPR: 78.6% on new PDD sample	
Lee 2016	PDD vs PD-nodementia	Cog	LR	0.869	x	77.4%	96.4%	x	x	x	x	
Byeon 2020	PD-MCI vs PD-normal cognition	Cog	RF	x	89.5%	x	x	x	x	10.5%	x	
			n-BM	x	85.8%	x	x	x	x	x	x	
			DA	x	78.3%	x	x	x	x	x	x	
Dujardin 2010	iPD vs PDD	Cog	LR	0.93	x	85.37%	83.67%	x	x	x	x	
Dauwan 2016	DLB vs HC	EEG	RF	x	92%	95%	89%	x	x	x	x	
Xu 2020,	PD vs HC	EEG		x	x	84.84%	91.81%	88.31%	x	11.28% false HC; 11.49% false PD	x	
			DPRNN									
Koch 2013		EEG	K-means	x	x	95%	80%	x	x	x	x	
Oh 2020		EEG	CNN	x		88.25%	84.71%	91.77%	x	x	x	x
Liu 2017		EEG	O_CCA	x		92.86%	x	x	x	x	x	x
			SVM	x		97.62%	x	x	x	x	2.38%	x
			KNN	x		90.48%	x	x	x	x	9.52%	x
			NB	x		88.10%	x	x	x	x	11.90%	x
			RF	x		83.33%	x	x	x	x	19.67%	x
Waninger 2020		EEG	IDFA	x		95.24%	94.74%	95.65%	x	x	NPV: 95.65% PPV: 94.74%	
Yuvaraj 2018	EEG	SVM-RBF	x		99.62 ± 0.57	100.00 ± 0.00	99.25 ± 0.53	99.38 ± 0.47	0.98 ± 0.05	x	x	
		DT	x		90.65 ± 2.76	94.17 ± 2.03	85.63 ± 1.05	94.17 ± 1.03	0.90 ± 0.06	x	x	

			fuzzy KNN (m:1.24;k:3)	x	93.01 ± 2.89	95.00 ± 1.83	90.21 ± 2.61	95.00 ± 1.83	0.93 ± 0.05	x	x
			KNN (k-2)	x	90.83 ± 3.97	96.67 ± 1.30	82.50 ± 2.72	96.67 ± 0.88	0.90 ± 0.05	x	x
			NB	x	88.72 ± 2.94	94.17 ± 1.62	80.83 ± 3.82	94.17 ± 1.62	0.88 ± 0.06	x	x
			PNN (σ = 0.284)	x	93.98 ± 1.39	96.21 ± 0.83	90.61 ± 1.24	96.67 ± 1.83	0.94 ± 0.05	x	x
			SVM linear	x	91.98 ± 2.95	95.48 ± 1.83	87.58 ± 3.82	95.00 ± 1.83	0.92 ± 0.05	x	x
			SVM poly 2	x	91.98 ± 2.95	96.74 ± 1.83	95.36 ± 1.83	96.67 ± 1.83	0.96 ± 0.04	x	x
			SVM poly 3	x	95.12 ± 1.28	95.83 ± 1.39	95.83 ± 1.39	95.83 ± 1.39	0.95 ± 0.04	x	x
Chaturvedi 2017	EEG		RF	0.8	0.78	x	x	x	0.774	x	OR: 10.8
			LR	0.63	0.56	x	x	x	x	x	x
			SVM	0.73	0.747	x	x	x	x	x	x
			DT -J48	0.67	0.68	x	x	x	x	x	x
Anjum 2020	EEG	LEAP	85.2 (69.5–100)	85.7 (67.3–96)	85.7 (57.2–98.2)	85.7 (57.2–98.2)	x	0.839	x	OR : 36 (4.3–299) PPV: 85.7 (62–95.7); NPV: 85.7 (62–95.7)	
Chang 2022	EEG		Bag	0.90	0.81	0.85	0.75	0.75	0.80	x	x
			LogitBoost,	0.89	0.79	0.85	0.70	0.70	0.74	x	x
			Gentle Boost	0.88	0.79	0.90	0.60	0.70	0.77	x	x
			Tree	0.82	0.80	0.80	0.75	0.80	0.80	x	x
			SVM	0.86	0.74	0.60	0.87	0.80	0.69	x	x
			NB	0.7	0.60	0.20	0.95	0.80	0.32	x	x
			k-NN	0.8	0.70	0.65	0.80	0.79	0.70	x	x
	Cog	RF	0.91 (0.11)	x	x	x	x	x	x	x	

Souillard- Mandar 2016			SVM Gaussian kernels	0.86 (0.08)	x	x	x	x	x	x	x
			CART	0.76 (0.10)	x	x	x	x	x	x	x
			C4.5	0.75 (0.06)	x	x	x	x	x	x	x
			boosted DT	0.87 (0.08)	x	x	x	x	x	x	x
			regularized LR	0.91 (0.08)	x	x	x	x	x	x	x
Lee 2022	EEG	CatBoost DT	0.912	0.893	x	x	x	0.903	x	OR : 115.5	
		XGBoost,	0.701	0.679	x	x	x	0.64	x	OR : 3.778	
		LightGBM	0.886	0.857	x	x	x	3.778	x	OR : 107.25	
Vanneste 2018	EEG	SVM	0,95 (+/- 0.02)	x	x	x	x	x	TPR:0.93 (+/- 0.02); 0.11 (+/- 0.01)		
Khare 2021	EEG	LSSVM (RBF-SB7)	x	97,65%	96,67%	98,70%	98,76%	x	x	x	
		RF	x	96.46	x	x	x	x	x	x	
		KNN	x	96.16	x	x	x	x	x	x	
		SVM	x	95.20	x	x	x	x	x	x	
		ANN	x	95.48%	x	x	x	x	x	x	
		LSSVM Sinc	x	59.34%	x	x	x	x	x	x	
		LSSVM Poly	x	86.95%	x	x	x	x	x	x	
		LSSVM Morlet	x	97.42%	x	x	x	x	x	x	
		LSSVM MH	x	91.66%	x	x	x	x	x	x	
Lee 2021	EEG	CRNN	x	99,20%	x	x	98,90%	x	x	recall : 99.4%	

AdaBoost, Adaptive Boosting; ANN, Artificial Neural Networks; AUC, Area Under the Curve; Bag, Bagging algorithm; BR, Bayesian regression; BRR, Bayesian ridge regression; C4.5, decision tree algorithm developed by Quinlan; CART, classification and regression trees; CatBoost, Category Boosting; CI, cognitive impairment; CNN, convolutional neural network; Cog, Cognition; CRNN, convolutional recurrent neural network; DA, Discriminant Analysis; DCNN, deep neural network; DFA, linear discriminant function analysis; DLB, Dementia with Lewy Bodies; DPRNN, dual-path recurrent neural network; DT, Decision Tree; e-GB, gradient boosting; EEG, electroencephalography; ELPD, expected log pointwise predictive density; ER, error rate; FMI, Fowlkes-Mallows index; FPR, false positive rate; HC, Healthy controls; iRBD, idiopathic REM Sleep Behavior Disorder; KNN, K-Nearest Neighbours; l-GBM, Light gradient boosting machine; LASSOLAR, lasso model implemented using the least-angle regression algorithm ; LDA, linear discriminant analysis; IDFA, linear discriminant function analysis; LEAP, Linear-predictive-coding EEG Algorithm for PD; LightGBM, light gradient-boosting machine; LOLIMOT, local linear model trees; LR, logistic regression; LSSVM, least squares support vector machine; MCC, Mathew's correlation coefficient; ML, Machine Learning; MLP-BP, Multilayer Perceptron trained with backpropagation; MoCA, Montreal Cognitive Assessment; n-BM, naive Bayesian model; NB, Naïve Bayes; NPV/R, negative predictive value/rate; Nu-SVR-RBF, Radial Basis Function kernel with nu parameter SVM; O\_CCA, optimal center constructive covering algorithm; OR, Odds ratio; PAR, Passive Aggressive Regression; PD, Parkinson's Disease; PDD, PD'dementia; PD-REM-be; PD-REM behavioral events; PNN, probabilistic neural network; PPV/R, positive predictive value/rate; QDA, quadratic discriminant analysis; RBF, SVM with radial basis function; RF, Random Forest; RNN, Recurrent Neural Network; SVM, Support Vector Machine; TPR, True positive rate; TS, Thiel-Sen R ; XG Boost, extreme Gradient Boosting.

**Chapitre 3 – Article 2 : Predicting cognitive decline by  
clinical markers in isolated REM sleep behaviour disorder: a  
machine learning approach**



**Prediction of cognitive decline by clinical markers in idiopathic REM sleep behavior disorder: a machine learning approach**

Loubna Mekki Berrada<sup>1</sup>, Arthur Dehgan<sup>1</sup>, Ronald B. Postuma<sup>2,3</sup>, Yann Harel<sup>1</sup>, Hamza Abdelhedi<sup>1</sup>, Jacques Montplaisir<sup>3,4</sup>, Karim Jerbi\*<sup>1,5</sup>, Jean-François Gagnon\*<sup>3,6</sup>

<sup>1</sup>Cognitive and Computational Neuroscience Lab, Psychology Department, University of Montreal, QC, Canada

<sup>2</sup>Department of Neurology, Montreal General Hospital, Montreal, QC, Canada

<sup>3</sup>Centre for Advanced Research in Sleep Medicine, Centre intégré universitaire de santé et de service sociaux de Nord-de-l'Île-de-Montréal – Hôpital du Sacré-Coeur de Montréal, Montreal, QC, Canada

<sup>4</sup>Department of Psychiatry, Université de Montréal, Montreal, QC, Canada

<sup>5</sup>MILA (Quebec Artificial Intelligence Institute), Montreal, QC, Canada

<sup>6</sup>Department of Psychology, Université du Québec à Montréal, Montreal, QC, Canada

\*These authors share senior authorship.

**Abstract:**

Synucleinopathies, such as dementia with Lewy bodies (DLB) and Parkinson's disease (PD), are characterized by several motor and non-motor symptoms that limit the autonomy and impact the lives of patients and their relatives. The development of neuroprotective interventions for synucleinopathies is still at an early stage, and one of the populations to be targeted in future clinical trials will be individuals in the prodromal stages. Isolated/idiopathic rapid eye movement sleep behaviour disorder (iRBD), a parasomnia characterized by abnormal behaviors during sleep, is a major prodromal stage of synucleinopathies. However, there is a large heterogeneity in the presentation and progression of clinical symptoms in this population. Consequently, longitudinal studies of iRBD using a machine learning (ML) approach and easily available clinical markers offer an unprecedented opportunity to determine the predictive factors of the development of dementia or parkinsonism. In this study, we used a ML algorithm (Random Forest; RF) to determine the contribution of several features (sociodemographic, clinical, cognitive, and other clinical markers) that were measured at baseline, to identify iRBD patients who developed dementia first. Patients with polysomnography-confirmed iRBD (n=179) were followed for a mean of 5.6 years. At follow-up, 96 were still disease-free, including 30 with MCI; 34 developed PD, including 18 with MCI; and 26 developed DLB. The model effectively distinguished at baseline patients who developed dementia from those who progressed to PD first (0.80 +/- 0.03, p=2.22 e-16). Cognitive markers of executive functions (Trail Making Test), verbal learning (Rey auditory-verbal learning test), and mild cognitive impairment (MCI) played a major role in distinguishing the different pathological trajectories in ML models. In single-feature RF models, the performance of the MCI variable alone was equivalent to its use within a multivariate RF model for classifying dementia versus PD-first or idiopathic cases. Similarly, the RAVLT (but not the TMT part A nor part B) exhibited comparable effectiveness when used singly as opposed to within a multivariate model for distinguishing dementia from healthy controls, thereby reinforcing their clinical importance. Our findings demonstrate the efficacy of ML coupled with cognitive variables in paving the way for future clinical trials and deeper investigations in iRBD. To robustly validate the utility of ML in clinical settings for iRBD, further multicentric studies are needed.

**Key words:** Isolated/idiopathic REM Sleep Behaviour Disorder, Machine Learning, Cognition, Dementia, Parkinson's disease, Mild Cognitive Impairment.

## 1. Introduction

Synucleinopathies are a category of neurodegenerative diseases that includes Parkinson's disease (PD), dementia with Lewy bodies (DLB), and multiple system atrophy (MSA). No curative treatment is yet available for these conditions and the development of neuroprotective interventions remain embryonic. Therefore, targeting individuals in the prodromal stages of synucleinopathies must increasingly be the focus of clinical trials to test the efficacy of these interventions. Isolated/idiopathic rapid eye movement sleep behaviour disorder (iRBD) is a parasomnia recognized as a major prodromal stage of synucleinopathies [1, 2]. Indeed, a multicenter longitudinal study reported that the estimated conversion rate increases steadily with time after RBD diagnosis, reaching 10.6% at 2 years, 17.9% at 3 years, 31.3% at 5 years, 51.4% at 8 years, and 60.2% at 10 years [2]. Among the patients who phenoconverted, 43.5% developed dementia first, while 56.5% developed parkinsonism first [2]. In the latter subtype, a significant proportion of patients also had mild cognitive impairment (MCI). The presence of MCI in iRBD or PD increases the risk of conversion to DLB or PD with dementia [1-4]. There is a large heterogeneity in presentation and progression of the clinical profiles in iRBD patients, and this makes early detection of prodromal synucleinopathies and disease management challenging. Identifying iRBD patients at an early stage is critical for recruiting participants for clinical trials focusing on interventions that aim to prevent cognitive decline or parkinsonism in this population.

Several clinical markers have been found to predict phenoconversion in individuals with iRBD [1, 2]. However, few make it possible to identify those who developed dementia first, versus parkinsonism first phenotype. Longitudinal studies showed that the most significant early cognitive predictive feature of conversion to DLB in iRBD was impaired executive functions [5-7]. Lower performance on tests measuring verbal episodic memory also appeared to be a good predictive feature of dementia in this population, particularly in the two years before conversion [5-7]. Moreover, approximately one third of iRBD patients have MCI [1], which is associated with a high likelihood of phenoconversion to a dementia-first phenotype [2, 5-7]. However, none of these studies used a machine-learning (ML) models with model-validation methods, which test the ability of the model to generalize out of the sample used to train it. ML with model validation

identifies generalizable patterns by using learning algorithms on rich data, and then testing the resulting model on new data. It allows the prediction of what will happen next (ex: predicting the next stage of disease progression), which helps in clinical decision-making about the future (e.g., treatment choice) [8]. Unlike traditional classical statistical methods (i.e., explanatory analyses, and other inferential statistical analyses), which primarily focus on individual variable importance based on p values and predefined modeling hypotheses, ML thrives on minimal assumptions, often via "black box" algorithms. These "black box" algorithms are particularly skilled at discerning intricate relationships within data, especially when observations vary across variables. The goal is to efficiently model and predict outcomes, such as identifying patients who will develop dementia using input variables like cognitive measures, and then assess the model's predictive accuracy for future results [9-11]. Our approach in the present study builds upon previous research by examining variables that have been studied in the literature and instead considering them in the context of their interactions with one another, in order to optimize prediction. To date, the use of ML methods to predict the development of dementia in iRBD patients based on cognitive data has been limited. The few studies that have done so did not use model-validation methods [12]. Since iRBD is a prodromal state that can precede dementia diagnosis by many years, it is an important pathology to target with both ML algorithms and with the use of model-validation techniques.

The aim of this study was to assess the generalization capacity and accuracy of ML to predict later conversion of iRBD patients into the dementia first phenotype (DLB). We also identify which combination of sociodemographic, clinical, cognitive, and other clinical features (psychiatric, sleep, sensory, motor, and autonomic) allow for the highest prediction accuracy, based on previously studied markers – also evaluating the respective contribution of each selected variable. We examined the longitudinal classification of iRBD patients using their baseline data to predict their conversion to DLB versus other clinical outcomes.

## **2. Materials and methods**

### **2.1 Participants**

iRBD patients were recruited and followed from 2002 to 2022 at the Centre for Advanced Research in Sleep Medicine of the Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Île-de-Montréal – Hôpital du Sacré-Coeur de Montréal (CIUSSS-NÎM-HSCM), Canada. They were part of a longitudinal study on iRBD. Potential participants were identified based on meeting video-polysomnographic (vPSG) criteria for RBD [13, 14], and the first complete neuropsychological assessment was selected for data collection. Other inclusion criteria were: 1) a minimum of 6 years of education (primary school completed), 2) aged between 45-89, and 3) French or English as a native language. Exclusion criteria were: 1) dementia or another neurodegenerative disease [15], 2) parkinsonism [16], 3) major psychiatric disorders [15], 4) narcolepsy, 5) epilepsy, 6) unstable diabetes or hypertension, 7) history of stroke, brain tumour, head injury, encephalitis or chronic obstructive pulmonary disease, and 8) other medical conditions (e.g., limb amputation or active alcoholism). Healthy controls (free of MCI or RBD) were recruited, for comparisons with normal value, through newspaper advertisements and by word of mouth. Ethical approval was obtained from the hospital ethics committee (CIUSSS-NÎM-HSCM) and all participants signed a consent form.

### **2.2 Procedures**

#### ***2.2.1 Baseline Assessment***

Data was collected on a total of 58 variables through medical history, vPSG, and neurological and neuropsychological assessments [2, 17, 18]. The complete list of variables is presented in Table 1. Diagnosis of the cognitive condition, made through consensus between the neurologist (RBP) and neuropsychologist (PhD students in neuropsychology under the supervision of JFG) based on their evaluation of patient assessments, were: 1) without MCI (or normal cognition), 2) with MCI, or 3) dementia. Cognitive tests assessing attention and executive functions, verbal episodic memory, and visuospatial abilities were administered, scored and normed according to standard procedure [19-21]. MCI was defined based on (1) the presence of subjective complaint of cognitive changes by

the patient or spouse/caregiver on the interview or through the Cognitive Failures Questionnaire [20]; (2) the presence of objective measures of cognitive impairment (i.e.,  $\geq 1.5$  standard deviation below the standardized mean) on at least two tests assessing the same cognitive domain (i.e., attention and executive functions, learning and memory or visuospatial); (3) the absence of significant decline in daily living functioning; (4) the absence of dementia; and (5) no other medical condition explaining the cognitive condition [5, 21, 22].

### ***2.2.2 Longitudinal follow-up examination***

An annual follow-up examination of iRBD patients was conducted by a neurologist and neuropsychologist to investigate the conversion into a neurodegenerative disease. Parkinsonism was defined as bradykinesia plus at least one of rest tremor or rigidity [16, 23]. A diagnosis of dementia was given by consensus between the neuropsychologist and neurologist that screening tests showed cognitive impairment (Mini-Mental State Examination  $< 28$  or Montreal Cognitive Assessment  $< 25$ ) with significant functional impact related to impaired cognition [6, 7]. For patients who could not attend in person to the evaluation, the diagnosis was given based on phone contact with them or their caregivers, completed by a clinical chart review. This procedure was used in multiple previous studies that used this database [5, 17, 24, 25].

## **2.3 Conditions for classification**

Three main categories of longitudinal classifications (total of seven) were determined:

- 1) DLB versus iRBD:
  - a. DLB versus iRBD without MCI
  - b. DLB versus iRBD with MCI
  - c. DLB versus all iRBD
- 2) DLB vs PD:
  - a. DLB versus PD without MCI
  - b. DLB versus PD with MCI
  - c. DLB versus all PD
- 3) DLB versus Controls.

To prevent a circular reasoning bias, the classifications considering the MCI condition (i.e., a and b under both DLB versus iRBD and DLB versus PD) were conducted without including the MCI categorical variable.

## **2.4 Data cleaning and variable/subject management**

Since our database includes both categorical and continuous data types and considering that ML models can handle both types well, we examined the eight classifications using both data types. The initial data matrix contained missing scores where certain tests had not been completed by all participants. To create a usable matrix for each classification condition, a data cleaning procedure was implemented after selecting the subjects for the respective binary classifications - we removed any variables with missing values to preserve all subjects. Subsequently, to mitigate multicollinearity, we excluded variables exhibiting a correlation coefficient greater than 0.85 within each binary classification condition.

## **2.5 Machine learning (ML) analysis and statistics – multifeatured and single feature**

To determine the contribution of several features taken at baseline, and to identify patients who developed dementia first, we used supervised learning, “the ML task of inferring a function from labeled training data” [26]. Decision trees are one of the most intuitive and popular algorithms in data mining, due to their capacity to effectively manage heterogeneous data and non-linear effects between variables. The algorithm we used to implement the classifications is Random Forest (RF). RF develops several decision trees to simultaneously consider all input variables and assign an importance score to each variable. This score quantifies its contribution to the average improvement of the classification [27]. The starting node of a decision tree is called the "root" and the terminal nodes are the "leaves". The tree is built recursively, meaning it partitions the elements of a node into two descendant nodes, aiming to make these as homogeneous as possible. We measured the heterogeneity of a node with the Gini index. With each split of a node, the explanatory variable that maximizes the impurity variation is selected. The RF mechanism involves introducing

randomness in the construction of multiple CART trees (in the choice of the sample and/or explanatory variables) which form the forest, and then aggregating these trees by the average of the predicted values or by the majority of the predicted class (i.e., going with the majority vote of the trees) [28].

The metric used to assess the performance of a classifier was the accuracy. The accuracy quantifies the overall proportion of successful predictions  $acc = \frac{TP+TN}{TP+FP+TN+FN}$  [29]. The ML pipeline for the multifeatured analysis for each binary classification problem is found below (pseudo-code representation): **10 seeds** (the 2 Stages process below is executed for 10 different seeds, which corresponds to the initial values for random number generation, ensuring that the results are not due to chance):

### **Stage 1 – Exhaustive variable selection with hyperparameter optimization:**

#### ***1 random bootstrap subsampling (balancing groups)***

**1 - A) Exhaustive selection of variables with internal optimization of hyperparameters in a nested cross-validation (5-folds):** all possible combinations of variables are considered to find the best subset. Simultaneously, hyperparameters are optimized internally (GridSearch), within the inner loop of the nested cross-validation. The outer loop is used for model evaluation. Outcome of this step: the best-performing combination of variables.

**1 - B) Hyperparameter optimization for the best performing combination of variable:** Using the best combination of variables from the first stage (1-A), we perform another round of 5-fold cross-validation for hyperparameters optimization (GridSearch). This step is done to ensure the best hyperparameters are generalizable. Outcome of this step: the optimal hyperparameters for the best-performing combination of variables.



## **Stage 2 – Model evaluation through bootstrap subsampling with cross-validation:**

### ***50 random bootstrap subsampling (balancing groups)***

Using the best combination of variables (1-A) and hyperparameters (1-B) from Stage 1, each of the 50 subsamples is subjected to 5-fold cross-validation to evaluate model performance. For each subsample, precision scores and variable importance scores are calculated, leading to 50 sets of scores. An average precision score is derived across all 50 subsamples, and each variable is assigned an average importance score.

Output from the 10 seeds: we obtain 10 optimal sets of variable combinations, each associated with its own accuracy score. Additionally, for each variable that was selected as part of the multivariate model, we calculate its importance score across all 10 seeds when it is included in the model (i.e., its frequency of selection across the seeds, referred to as its 'seed appearance'). Therefore, for each classification, the results presented correspond to an average across 10 seeds.

In summary, this pipeline emphasizes rigorous evaluation and optimization to ensure the model's performance is not a result of overfitting and can generalize well to new data. It incorporates balancing of groups, exhaustive search for the best feature set, nested cross-validation for reliable performance estimation, and extensive hyperparameter tuning.

Computations were conducted using the Compute Canada (<https://osf.io/k89fh/wiki/Compute%20Canada/>) high performance computing system. Since the results are an average created through multiple seeds, statistical significance of the mean accuracy obtained is assessed using the binomial law.

In the pipeline for single feature analysis within a binary classification context, each feature undergoes 50 rounds of balanced bootstrap subsampling. A RF classifier is trained and optimized for hyperparameters (GridSearch) through a 5-folds cross-validation. The performance of each feature is quantified by its mean accuracy score, derived from these bootstraps. This process is repeated independently for each feature to identify its impact on the specific binary classification task.

In the single feature analysis method, each feature is evaluated independently to determine its accuracy in a binary classification problem using a RF model. Specifically, the process involves 50 iterations of bootstrap subsampling to ensure balanced groups, where each iteration includes a 5-fold cross-validation to train the RF model, fine-tune hyperparameters through GridSearch, and test the model's performance. The best resulting score from each bootstrap iterations are aggregated to calculate the average accuracy and standard deviation for the feature, which reflects its performance consistency in the classification task.

### **3. Results**

#### **3.1 Participants**

A total of 278 participants were recruited at baseline (see Figure 1). Among the 204 iRBD patients, 25 were excluded: 11 for possible or probable neurodegenerative diseases, 5 for history of severe head injury or chronic obstructive pulmonary disease, 4 for major psychiatric disorders, 1 for narcolepsy and 4 for other medical conditions. Hence, a total of 179 iRBD patients, including 67 with MCI, and 74 controls were included in the study. At follow-up (mean = 5.6 years), 21 patients, including 8 with MCI, did not have a second visit and 2 others developed MSA. These patients were excluded from the longitudinal analysis. Among the 156 patients with follow-up, 96 were still disease-free, including 30 with MCI; 34 developed PD, including 18 with MCI; and 26 developed DLB. Data for each group are presented in Supplemental Material. For continuous variables, baseline mean performance of all patients and controls are in Supplemental Material Table 1.

#### **3.2 After data cleaning - variables and subjects**

The exact features tested for each classification after data cleaning are listed in Appendix 1. The correlation analysis reveals strong correlations ( $>0.85$ ) within all seven classification categories, notably between WAIS Digit Span Total (z) and both WAIS Digit Span Forward (z) and WAIS Digit Span Backward (z), as well as between RAVLT Immediate Recall (z) and RAVLT Trials 1-

5 (z). To reduce multicollinearity, we removed the specific variables from the dataset and kept the global variables (i.e., WAIS Digit Span total (z) and RAVLT 1-5 (z)).

The number of subjects and variables that were retained for each classification condition after data cleaning is detailed in Table 2 (refer to the Supplemental Material Table 2 for a complete list of variables).

### **3.3 After exhaustive feature selection – variables**

Following the exhaustive feature selection process, which tested all possible combinations of the variables provided, a dimensionality reduction is obtained for all classifications. For instance, for the DLB vs PD without MCI classification, the number of variables decreased from 26 to an average of  $12.4 \pm 8.92$  across 10 seeds (see Table 2). This indicates that multiple variables were introducing noise, rather than contributing to optimize the multimodal ML model. The best features of each multivariable model (i.e., for each classification) can be found in Table 2 and are illustrated, for 4 classifications in Figures 2-4. The 'best features' were identified as those with the greatest importance, and often selected across multiple seeds for inclusion in the optimal multivariable model, supporting their validity (i.e., they were unlikely to have been chosen due to chance).

### **3.4 Binary classification – Distinguishing the progression to dementia at follow-up (~5.6 years) from other clinical trajectories**

#### **3.4.1 Classification of dementia at follow-up from PD or idiopathic state**

We examined the potential of various sociodemographic, cognitive and clinical features to predict 1) DLB vs iRBD with or without MCI, and 2) DLB vs PD with or without MCI. The predictive performance for these classifications was  $0.78 \pm 0.05$  ( $p=2.22 \text{ e-}16$ ) and  $0.80 \pm 0.03$  ( $p=2.22 \text{ e-}16$ ), respectively.

In Figure 2A, we present a scatter plot identifying the best features for the first classification—iRBD maintenance with or without MCI versus dementia progression. The MCI categorical variable emerged as a highly important feature, consistently selected in the optimal model across the 10 seeds, following exhaustive feature selection, with an average importance score of  $0.91 \pm 0.006$ . A similar pattern emerges when differentiating dementia from all PD phenotypes, with MCI being the best variable ( $0.84 \pm 0.004$  mean importance and selected in the optimal multifeatured model in 10/10 seeds) (see Supplementary Figure 1).

In the single-feature analysis presented in Figure 2B, the MCI alone performed comparably to the multivariate model that included MCI, for dementia versus PD ( $0.81 \pm 0.002$ ) and iRBD stability ( $0.80 \pm 0.003$ ), both statistically significant with  $p < 0.05$ . Moreover, when excluding MCI from the model significantly reduced its predictive accuracy for DLB versus all iRBD ( $0.667 \pm 0.01$ ) and DLB versus all PD ( $0.641 \pm 0.03$ ), confirming the critical role of the MCI variable. Consequently, the most predictive features for DLB versus all iRBD shifted to a cluster that included memory and executive function variables, along with age (Supplementary Figure 2A), and for all PD versus dementia, age and a motor function variable (Alternating Tapping Test), in interaction with cognitive variables of memory, attention and executive functions (with less importance). (Supplementary Figure 2B).

MCI's importance was further reinforced by its prevalence at baseline; 88% of iRBD patients who later developed dementia had a MCI, in contrast to 26% of those progressing to PD first. In summary, the presence of MCI serves as a pivotal distinction between the evolutionary trajectories toward PD or stable iRBD and dementia.

### 3.4.2 Classification of DLB versus Controls

The ML model demonstrated an accuracy of  $0.79 \pm 0.02$  ( $p = 2.22 \times 10^{-16}$ ) in differentiating iRBD patients who developed dementia at follow-up from Controls.

Figure 3A presents the results of the multifeature analysis. The model significantly distinguished iRBD patients progressing to dementia from Controls. Notably, no clinical measures were included, due to the absence of such data for Controls. The best feature was the RAVLT 1-5, a test of verbal episodic learning (memory), followed by executive functions (TMT part B) and an attention-related task (TMT part A), with sociodemographic variables being of lesser importance.

Figure 3B highlights a critical insight: the best features from the multivariate analysis also exhibited superior performance when evaluated independently against other single variables from the dataset. This is especially true for the RAVLT 1-5, which alone matched the performance of the comprehensive multimodal RF. However, the TMT part B and TMT part A did not perform as well alone, indicating that they are more effective to optimize prediction when used in combination with other cognitive variables.

In summary the model's multivariate approach efficiently identifies potential dementia development in iRBD patients when compared to Controls, with specific cognitive tests assessing memory (RAVLT 15) proving to be particularly effective both independently and as part of a composite predictive model.

### 3.4.3 DLB versus no cognitive decline conditions (without MCI) at follow-up

The multifeature RF models successfully differentiated between patients who developed dementia from those that will not cognitively decline (i.e., PD without MCI and iRBD without MCI). A mean accuracy of  $0.71 \pm 0.05$  ( $p=2.22 \times 10^{-16}$ ) was obtained for the iRBD without MCI versus DLB classification and  $0.76 \pm 0.04$  ( $p=1.46 \times 10^{-11}$ ) for the PD without MCI versus DLB classification. Figure 4A illustrates the importance and selection frequency of features in the RF classification models between DLB and the conditions without MCI at follow-up.

The evolutionary trajectory towards PD-first in the absence of cognitive decline, is mainly distinguished by executive and memory variables. These variables were also important in distinguishing iRBD without MCI from dementia but were accompanied by a broader range of variables: age, motor and psychiatric.

For both classifications without cognitive decline, the univariate RF model illustrated in Figure 4B (i.e., plot of mean accuracy scores of individual features), highlights that despite some features reaching statistical significance, the integrated approach of the multifeature model holds superior predictive value, systematically outperforming the single-feature model.

#### **4. Discussion**

The aim of this study was to determine the accuracy of a thoroughly validated ML model in classifying subsequent conversion to DLB in iRBD patients, identifying the optimal combination of three modalities: 1) sociodemographic and clinical, 2) cognitive, and 3) other clinical (psychiatric, sleep, sensory, motor, and autonomic) markers. Figure 5 summarizes the findings for all classifications and highlights the key features to distinguish evolution to dementia from the other trajectory pathways.

The model effectively distinguished baseline patients who developed dementia from those who either progressed first to PD or remained idiopathic, as well as, at a lesser accuracy, from those without cognitive impairment (i.e., PD without MCI and iRBD without MCI) at follow-up. This study demonstrates the crucial role of cognitive markers of executive functions (TMT), memory (RAVLT), and MCI, established from univariate statistical studies, in ML models. Additionally, our analysis revealed that multivariate RF models surpassed univariate RF models in accuracy when discerning specific cognitive outcomes, notably the absence of MCI at follow-up. Our findings highlight the potential of ML and cognitive variables for future clinical trials and investigations in iRBD and offer a new perspective on previous research in the field.

When comparing our results to previous studies using inferential statistics, it is essential to recognize the distinct interpretation of findings. Inference focuses on isolating individually significant variables based on p-values, while prediction identifies variable sets that accurately forecast outcomes using new data [10, 30, 31]. Therefore, variables important in null hypothesis testing may not align with those maximizing predictive performance in new individuals or contexts [32].

iRBD is a major prodromal stage for synucleinopathies [33-36]. In the present study, 60 out of the 156 patients (38%) followed over a mean period of 5.6 years developed a synucleinopathy. Of these, 43% developed dementia first (DLB) and 57% developed parkinsonism first (PD), which is similar to the proportion reported in a recent multicentric study [2]. Given the heterogeneity in clinical presentation and progression within PD, iRBD, and DLB, selecting participants for neuroprotective intervention trials in prodromal synucleinopathies poses challenges. Our study contributes to precise the description of the clinical phenotype and specific predictors of conversion subtypes in this population, which could help refine selection criteria for future clinical trials.

Although several markers exist for iRBD patients at risk of synucleinopathies, few specifically identify those at risk for DLB conversion [5, 6, 37]. We found that MCI serves as a pivotal variable within a multivariable ML model to distinguish between the progression to DLB versus other trajectories, such as PD or remaining idiopathic. Intriguingly, MCI's predictive power (i.e., accuracy) stands strong even as a singular feature within an RF model. This finding highlights the clinical relevance and potential of the MCI variable for phenoconversion prediction and also corroborates previous findings in classical statistics showing that MCI is associated with a high likelihood of phenoconversion to a dementia-first phenotype [2, 5-7]. Also, even if a substantial subset of iRBD patients presents with MCI - over a third [21, 38] (37% at baseline in our study), which is a known risk factor for dementia, not all cases progress to DLB [39-41] and some patients with MCI may return to normal cognition at follow-up. In our study, a stark contrast was seen in the prevalence of MCI among patients who developed DLB (88.5%) compared to those who developed PD (26%) or remained idiopathic (31%), reinforcing the predictive value of MCI et consolidating the previous findings that iRBD patients who developed PD first are similar to

patients who remained disease-free on cognitive tests performance or MCI diagnosis frequency [5]. Moreover, excluding MCI from our multivariate models significantly diminished their predictive power for DLB against all iRBD and all PD (16% and 17% less accuracy), reiterating the key role of the MCI variable. The early stages of synucleinopathies exhibit significant overlap, making it challenging to distinguish between different types of neuropathological developments. However, our study confirms the value of MCI in differentiating these synucleinopathy trajectories using a ML model 5 years before the conversion occurs.

Prior research has shown that deficits in attention, executive function, and visuospatial abilities are distinguishing characteristics among iRBD patients who progress to PD (with and without MCI) as opposed to those who develop DLB [5]. However, our findings reveal a nuanced perspective when employing a ML model. Specifically, we found that these measures do not contribute importantly to the discrimination of DLB from all PD when applying the model to new patient data (when including and also when excluding the MCI variable from the ML model). This suggests that, within our ML framework, these measures's abilities may not be as critical as previously thought for classifying DLB versus all PD. Indeed, the exclusion of MCI from the ML model resulted in a shift of the most predictive features for DLB versus all PD towards age and motor functions, which, while still interacting with cognitive variables of memory, attention, and executive functions, displayed a diminished weight in the multivariate ML analysis. However, when MCI was not factored into the classification of DLB versus all iRBD, executive function variables did emerge as the most significant predictor, along with age. These shifts point to an interesting insight: cognitive variables, while pivotal, may not act in isolation. The interplay between cognitive, motor, and age-related factors appears more intricate within the ML context than in traditional statistical analyses. Notably, age remains a consistent marker and its association with an increased probability of transitioning to neurodegenerative conditions aligns with previous findings [2]. This divergence between statistical and ML outcomes suggests that ML models may capture complex, multidimensional interactions, offering a more dynamic understanding of the factors that influence the progression from iRBD to neurodegenerative states.



In our study, we also trained the ML model to differentiate iRBD patients on the path to DLB from those without cognitive impairments at follow-up—specifically, controls, PD without MCI, and iRBD without MCI.

As depicted in Figure 5, the RF algorithm identified memory and executive function variables as the key ones in classifying these three conditions from dementia a mean 5.6 years before conversion. Previous longitudinal studies using classical statistics on cognitive changes that precede dementia development in iRBD found decline in executive functions (TMT part B) as the earliest manifestations, 6 years before conversion [6, 7], while significant decline in learning (RAVLT) and semantic verbal fluency was found closest to conversion (i.e., two to four years prior to conversion) [6]. We found that for the classification of DLB at follow-up from controls using a RF model, memory (verbal episodic learning, RAVLT 1-5) was the best feature, followed by executive functions (TMT part B). The fact that the main predictors of DLB are cognition is consistent with predictions of the Unified Staging System and Braak Model. Interestingly, contrarily to what previous studies suggest, the TMT part B did not perform as well alone, indicating that it is more effective to optimize prediction 5.6 years before conversion when used in combination with other cognitive variables. Thus, our study reveals that the TMT B, even if found as altered the earliest in patients that will develop DLB in previous studies, demonstrates an ability to optimize predictive accuracy mostly in interaction with other cognitive variables, mainly with the RAVLT 1-5 in multivariable RF. However, the RAVLT 1-5 used alone in a RF shows similar accuracy to the multivariate RF model, suggesting its potential utility in clinical settings.

However, when discerning more specific cognitive outcomes from dementia, precisely the absence of MCI at follow-up, multivariate RF models systematically surpassed univariate RF models in accuracy. To differentiate at follow-up patients with dementia from those with either iRBD or PD who did not develop MCI, executive and memory variables were again, found as the best features to optimize accuracy for both classifications (DLB vs. iRBD without MCI and DLB vs. PD without MCI) in a multivariate RF model, but never surpassed the multivariate RF performance when used alone in a univariate RF model. Our study also reveals that, to classify absence of MCI in iRBD from DLB at follow-up, a more complex ML model is needed, than for

absence of MCI in PD from DLB. Indeed, specific to the DLB versus iRBD without MCI, memory and executive functions are not the only important variables of the multivariate model. Indeed, the model is characterized by a range of high-impact variables drawn from a synergistic interaction of sociodemographic, clinical, and cognitive modalities. Indeed, the model is more complex, including as top variables a cluster with age, motor, olfaction along with cognitive (memory and executive functions) features. This suggests that to classify absence of MCI in iRBD from DLB at follow-up, a combination of variables from the three modalities is needed. These findings show the complexity of dementia prognosis and the strength of a multimodal approach in capturing the nuanced interplay of cognitive, sociodemographical and clinical variables in differentiating DLB from the most distant cognitive condition - non-cognitive impaired iRBD.

The distinction between iRBD patients who developed PD with normal cognition from PD with MCI has rarely been made in studies that longitudinally followed the conversion of iRBD – we addressed this gap by also studying the presence of the MCI condition in iRBD and PD at follow-up, pioneering the application of ML to such data. We found that the ML models exhibit diminished efficacy in distinguishing iRBD patients who are likely to develop MCI, either in isolation (i.e., iRBD with MCI; 0.62 +/- 0.07,  $p=2.22 \times 10^{-16}$ ) or in conjunction with PD (i.e., PD with MCI; not significant), from those who will progress to dementia. These findings indicate lower performances in classifying the presence of MCI at follow-up (in either iRBD or PD) from DLB, which could imply that iRBD and PD with MCI share neuropathological similarities with DLB. It is already well recognized that PD patients with MCI are at higher risk of developing dementia [42] and present more severe cerebral abnormalities [43-45].

Some pathophysiological models have been put forth to clarify the progression of iRBD towards PD and dementia. Of these, the Braak model comprises six distinct stages [46]. It suggests that the propagation of alpha-synucleins begins in the dorsal motor nucleus of the vagus nerve and olfactory bulb (stage 1), intensifies in the brainstem, substantia nigra, basal prosencephalon, and mesocortex (transentorhinal region – stages 2-4) with the onset of significant PD symptoms, and ultimately impacts limbic structures and neocortex, resulting in cognitive symptoms (stages 5-6). This influential staging system has inspired alternative models, such as the Unified Staging System

for Lewy Body Disorders [47, 48], which have been proposed to more accurately represent the heterogeneity of Lewy body diseases. While these models have improved our comprehension of synucleinopathy progression, they cannot fully explain the pathophysiological differences underlying the PD-first versus DLB-first presentations. The precise nature of this distinction remains uncertain, with significant variability evident even in the prodromal stages of PD (pathological spread, imaging markers, onset and progression of symptoms) due to a variety of pathophysiological mechanisms [49]. We hypothesize that the pathophysiological evolutionary patterns during the prodromal phase of synucleinopathies differ between patients initially at risk for cognitive decline and those initially at risk for parkinsonism. We further propose that ML can capture the complex underlying factors contributing to the various prodromal presentations of PD and DLB through its complex "black box" capabilities.

Recognizing the complexity of synucleinopathies' prodromal stages, we employed ML methods to tackle this challenge. While some studies have investigated DLB and iRBD prodromal stages using ML and neuroimaging [50], our study is the first to use ML with model-validation to predict dementia in iRBD patients through easily obtainable measures (cognitive, sociodemographic, and clinical data). The significant out-of-sample classification accuracies we achieved highlights the potential of ML in clinical practice and for applications to other pathologies. We also show the importance of cognitive markers do distinguish the prodromal phenotypes and that combining different types of biomarkers (i.e., sociodemographic and clinical to cognitive) results in more accurate predictions of the absence of MCI at follow-up, from dementia. Our approach builds upon previous research by examining variables that have been previously studied in the literature but considering them in the context of their interactions with one another in the purpose of optimizing prediction, providing a more comprehensive understanding of the relationships between them. We show that to predict the evolution of neuropathology, the ML model requires different tests depending on the clinical question asked: evolution towards PD (with and without MCI), DLB, or MCI. We offer a continuum of the most important variables to address depending on the synucleinopathy condition. Other strengths of this study include dealing with data imbalance, model optimization, and rigorous validation techniques that clinicians new to ML can follow. To further improve the sensitivity and specificity of the ML approach, it should be tested on larger and more diverse samples (e.g., creating an international

uniformized database for iRBD), as well as with different algorithms in multimodal context with neuroimaging data (e.g., electroencephalography).

The present study has some limitations. Among all the ML algorithms available, we used the RF algorithm since it was best suited for our data and scientific question. Different results could have been found with other algorithms. We did not address MCI subtypes due to limitations in classification conditions and reduced group sizes when divided, which is not recommended in ML classification. The MCI should be further investigated in its subtypes, in the context of the development of ML algorithms. Our large longitudinal iRBD dataset had substantial missing data, forcing us to remove variables from our classifications, which means that not all variables in Table 1 were considered for each classification.

**Table 1** – Inventory of 58 database variables categorized into three modalities.

<b>Sociodemographic and clinical</b>	<b>Cognition</b>	<b>Other clinical markers</b>
<i>Sociodemographic</i>	MCI*	<i>Psychiatric</i>
Age (yrs)		BDI-II
Gender*	<i>Screening</i>	BDI-PC
Education (yrs)	MMSE	BAI
	MoCA	BAI-PC
		UPDRS I – Hallucinations
<i>Clinical</i>	<i>Attention and executive functions</i>	UPDRS I – Depression
RBD duration Sx	WAIS Digit Span forward (z)	UPDRS I – Apathy
RBD duration Dx	WAIS Digit Span backward (z)	Psychiatric History*
% REM tonic EMG	WAIS Digit Span total (z)	Antidepressant*
% REM phasic EMG	TMT A (z)	Anxiolytic*
RBD familial history*	TMT B (z)	
Clonazepam*	TMT B-A (z)	<i>Sleep</i>
Melatonin*	TMT infractions	ISI
Pramipexole*	<sup>1</sup> Semantic verbal fluency (z)	ESS
	Animals verbal fluency (z)	
	Phonetic verbal fluency (z)	<i>Sensory</i>
		UPSIT
	<i>Episodic verbal memory</i>	FM-100
	RAVLT 1-5 (z)	
	RAVLT Trial 1 (z)	<i>Motor</i>
	RAVLT B (z)	UPDRS-III
	RAVLT IR (z)	Purdue Pegboard
	RAVLT DR (z)	ATT
	RAVLT Recognition (z)	Timed Up and Go
	RAVLT False Recognition	
	RAVLT Intrusions	<i><sup>2</sup>Autonomic</i>
		Orthostatic hypotension
	<i>Visuospatial abilities</i>	SBPD
	ROCF copy (z)	Urinary
	Block Design (z) or VOSP	Sexuality (Erectile Sx)
	Bells test (z)	Constipation

ATT, Alternate Tap Test; BAI, Beck Anxiety Inventory; BAI-PC; BAI Primary Care; BDI-II, Beck Depression Inventory, Second Edition; BDI-PC, BDI Primary Care; EMG, Electromyography; ESS, Epworth Sleepiness Scale; FM-100, Farnsworth-Munsell 100-hue test; ISI, Insomnia Severity Index; MCI, Mild Cognitive Impairment; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; Phonetic Fluency, letters P, F, L in French or F, A, S in English, one minute each; Q, question; RAVLT 1-5, Rey Auditory-Verbal Learning Test sum of trials 1 to

5; RAVLT B, RAVLT List B; RAVLT IR, RAVLT immediate recall; RAVLT DR, RAVLT delayed recall; REM, rapid eye movement; RBD, REM sleep behaviour disorder; ROCF, Rey-Osterrieth Complex Figure copy; Sx, Symptoms; SBPD, Systolic blood pressure drop; TMT A, Trail Making Test part A; TMT B, TMT part B; TMT B-A, TMT part B minus part A; UPDRS, Unified Parkinson's Disease Rating Scale; UPSIT, University of Pennsylvania Smell Identification Test; VOSP, Visual Object and Space Perception Battery (number location and object decision); WAIS, Wechsler Adult Intelligence Scale Third Edition; (z), z score; \*, categorical variable; #Used for MCI diagnosis at baseline. <sup>1</sup>Semantic Fluency, animals or fruits/vegetables, one minute each; <sup>2</sup>Assessed using a structured interview modeled on the Multiple System Atrophy Rating Scale (Wenning et al., 2004).

**Table 2** – Performances for all classifications - multifeatures (mean through the 10 seed).

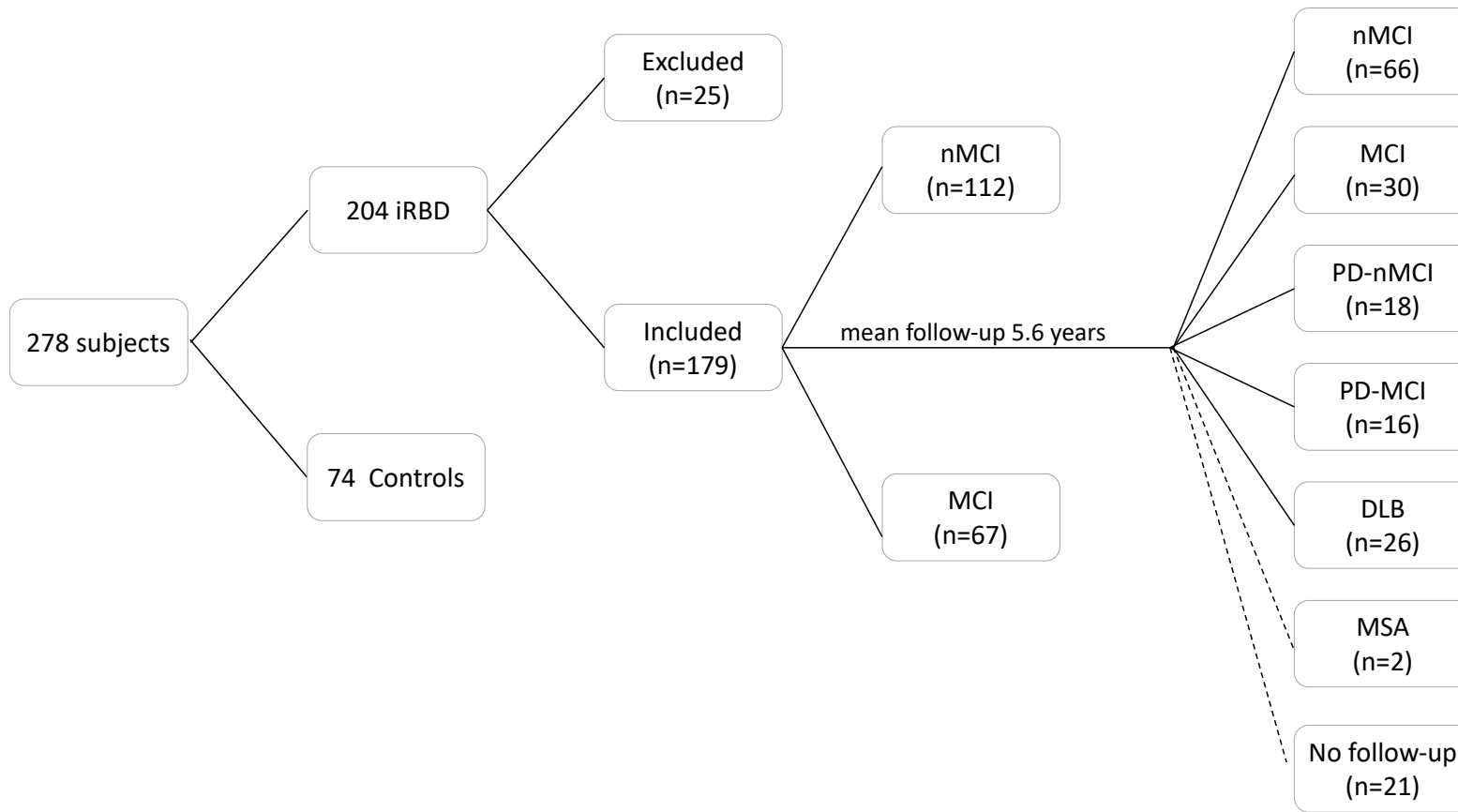
	<b>Classifications</b>						
	<b>[0]</b> DLB vs PD-nMCI	<b>[1]</b> DLB vs iRBD-nMCI	<b>[2]</b> DLB vs Controls	<b>[3]</b> DLB vs all PD	<b>[4]</b> DLB vs all iRBD	<b>[5]</b> DLB vs iRBD-MCI	<b>[6]</b> DLB vs PD-MCI
<b>Accuracy</b>	<b>0.76 ± 0.04</b>	<b>0.71 ± 0.05</b>	<b>0.79 ± 0.02</b>	<b>0.80 ± 0.03</b>	<b>0.78 ± 0.05</b>	<b>0.62 ± 0.07</b>	0.55 ± 0.06
<b>p value</b>	1.46e-11	2.22 e-16	2.22 e-16	2.22 e-16	2.22 e-16	2.22 e-16	2.32 e-10
<b>Subjects</b>	DLB=26 PD-nMCI= 18	DLB= 26 iRBD-nMCI=66	DLB= 26 Controls=74	DLB=26 all PD=34	DLB=26 all iRBD=96	DLB=26 iRBD- MCI=30	DLB=26 PD-MCI=16
<b>Nb variables before FS</b>	26	20	13	26	19	21	25
<b>after FS</b>	12.4 ± 8.92	10.4 ± 5.80	4.0 ± 3.68	3.6 ± 5.5	2.8 ± 4.05	10.1 ± 6.87	12.1 ± 8.517
<b>Best features<sup>1</sup> in multi feature analysis</b>	TMT B (z) RAVLT 1-5 (z)	Age (yrs) TMT B (z) UPSIT RAVLT B (z) RAVLT 1-5 (z) RAVLT Recognition (z) RAVLT DR (z) ATT	RAVLT 1-5 (z) TMT B (z) TMT A (z)	MCI*	MCI*	Age (yrs) Education (yrs) RAVLT B (z) TMT A (z)	RBD Duration Dx WAIS Digit Span total (z) Age (yrs) RAVLT DR (z)
<b>Best features in single</b>	RAVLT 1-5 (z) 0.73 ± 0.02 TMT B (z) 0.71 ± 0.04	RAVLT Recognition (z) 0.69 ± 0.01 ATT 0.66 ± 0.02	RAVLT 1-5 (z) <b>0.80 ± 0.02</b> TMT B (z) 0.74 ± 0.03	MCI* <b>0.81 ± 0.002</b> Age (yrs) 0.67 ± 0.01	MCI* <b>0.80 ± 0.003</b>	RAVLT Recognition (z) 0.63 ± 0.02 Age (yrs)	Age (yrs) 0.67 ± 0.04 Clonazepam* 0.66 ± 0.009

<b>feature analysis</b>	Timed Up and Go 0.66 ± 0.02	RAVLT B (z) 0.66 ± 0.01	TMT A (z) 0.70 ± 0.03	RAVLT 1-5 (z) 0.67 ± 0.03	RAVLT Recognition (z) 0.68 ± 0.02	TMT B (z) 0.65 ± 0.02	0.62 ± 0.02	Clonazepam* 0.62 ± 0.003	WAIS Digit Span total (z) 0.65 ± 0.3
-------------------------	--------------------------------	----------------------------	--------------------------	------------------------------	--------------------------------------	--------------------------	-------------	-----------------------------	---

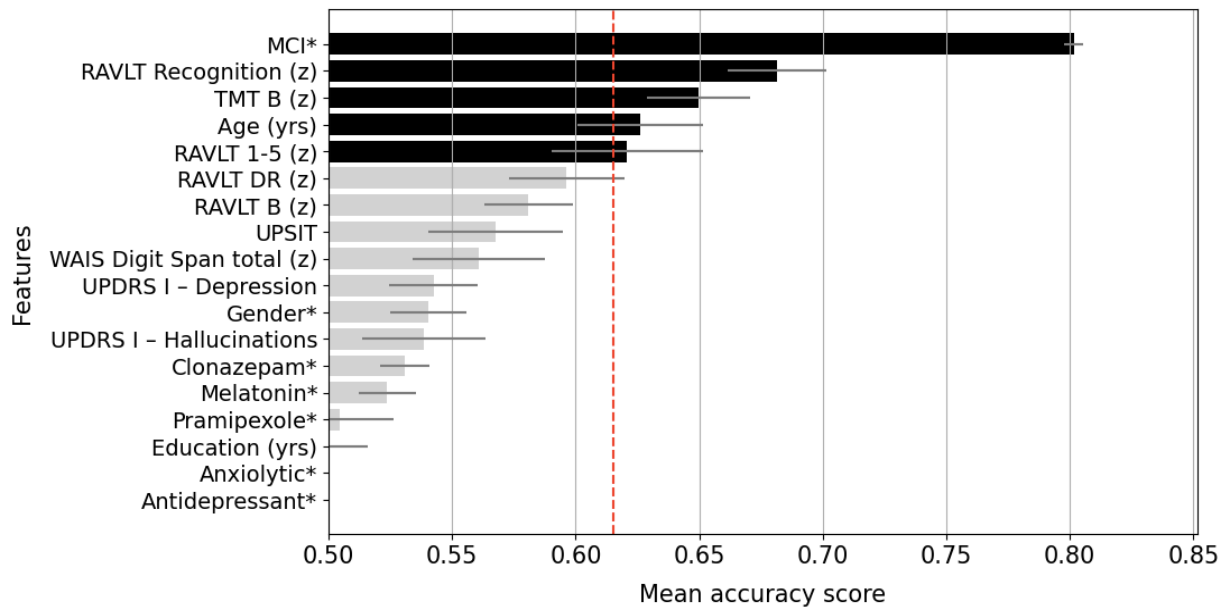
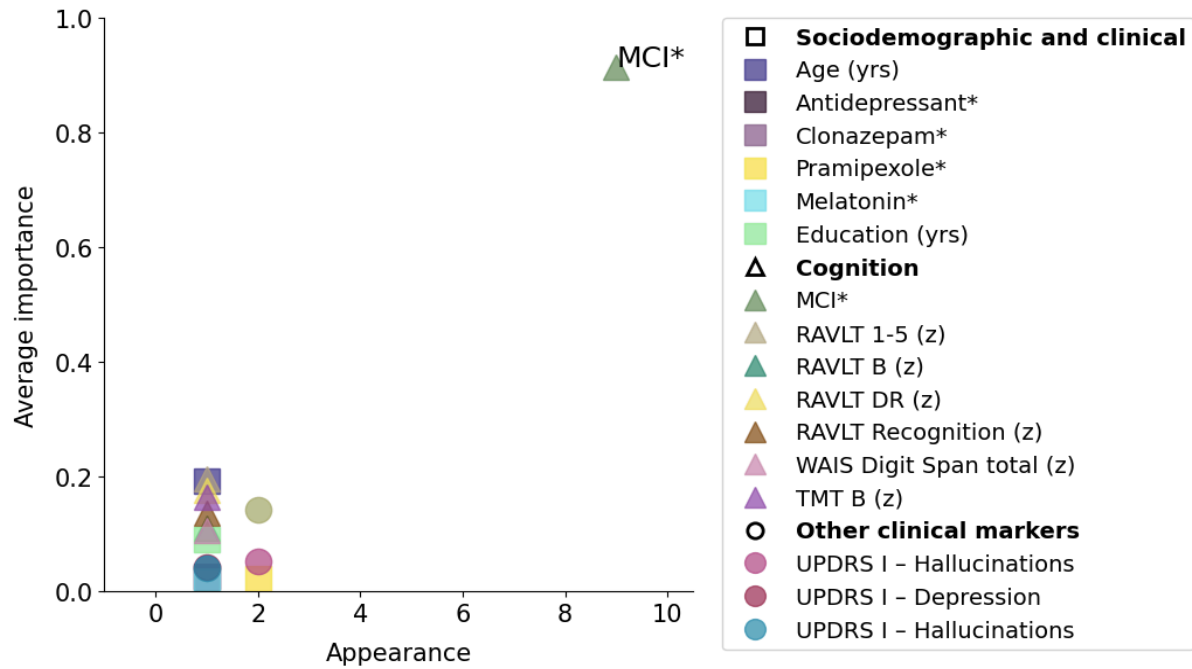
---

ATT, Alternate Tap Test; MCI, Mild Cognitive Impairment; RAVLT 1-5, Rey Auditory-Verbal Learning Test sum of trials 1 to 5; RAVLT B, RAVLT List B; RAVLT DR, RAVLT delayed recall; REM, rapid eye movement; RBD, REM sleep behaviour disorder; TMT A, Trail Making Test part A; TMT B, TMT part B; UPSIT, University of Pennsylvania Smell Identification Test; WAIS, Wechsler Adult Intelligence Scale Third Edition; (z), z score; \*, categorical variable; <sup>1</sup>Best features were defined as the features with the highest importance and that appeared in multiple seed in the multifeatured model.

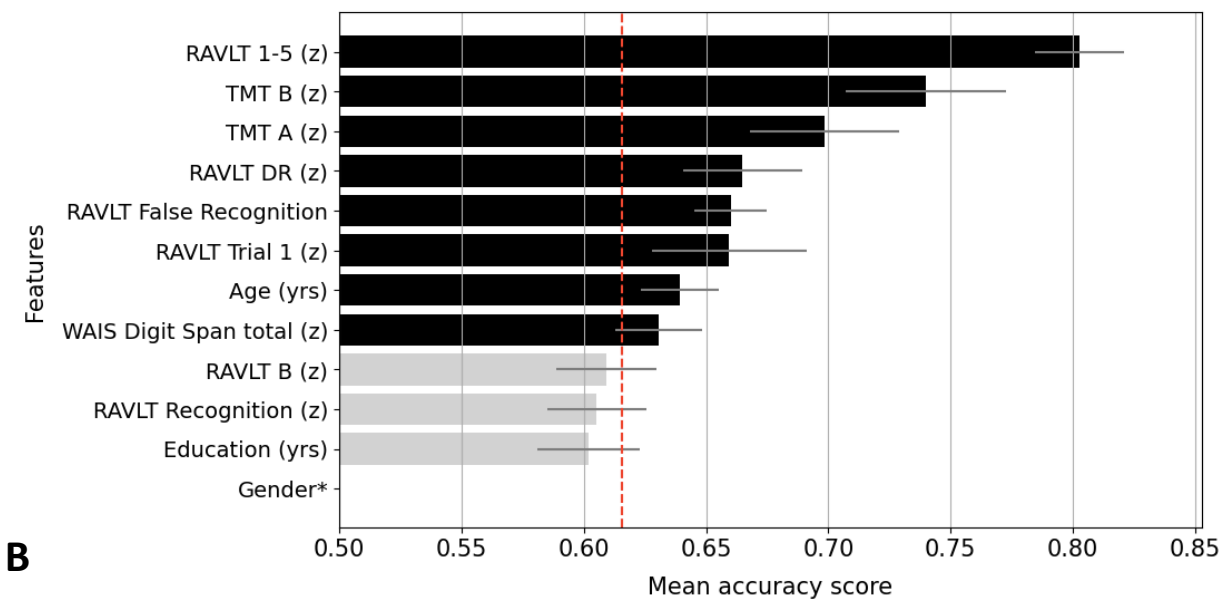
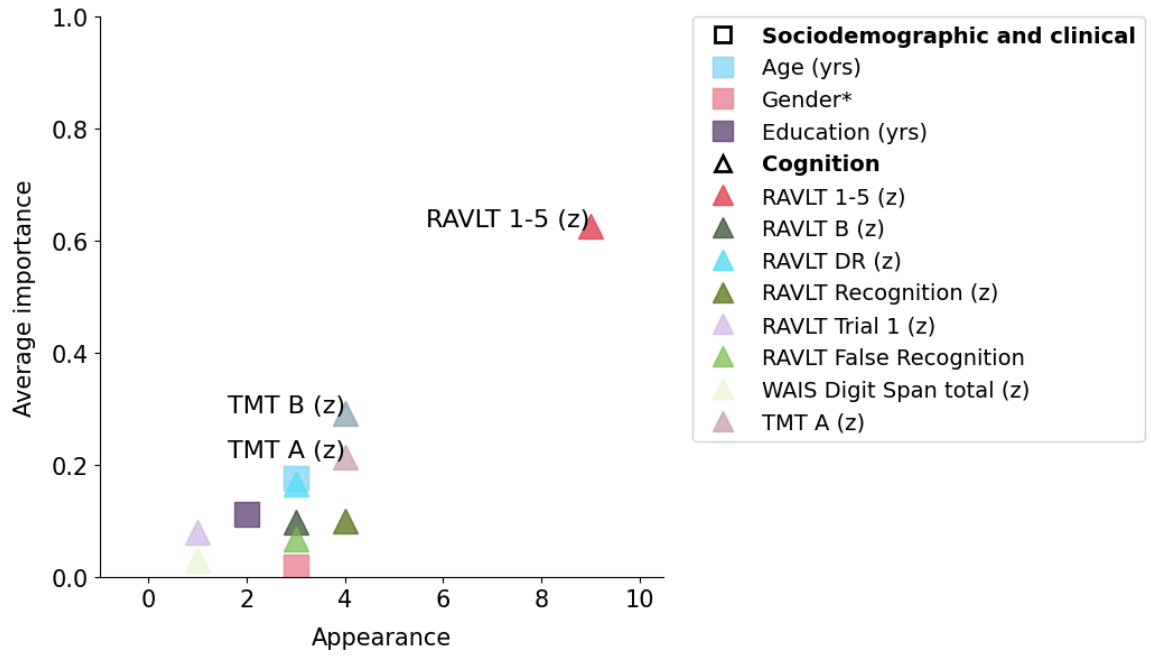




**Figure 1 – Participant flowchart.** DLB, dementia with Lewy bodies; iRBD, isolated/idiopathic rapid eye movement sleep behavior disorder; MCI, patients with mild cognitive impairment; MSA, multiple system atrophy; nMCI, patients without MCI; PD, Parkinson’s disease; ----- = not included in the longitudinal analysis.

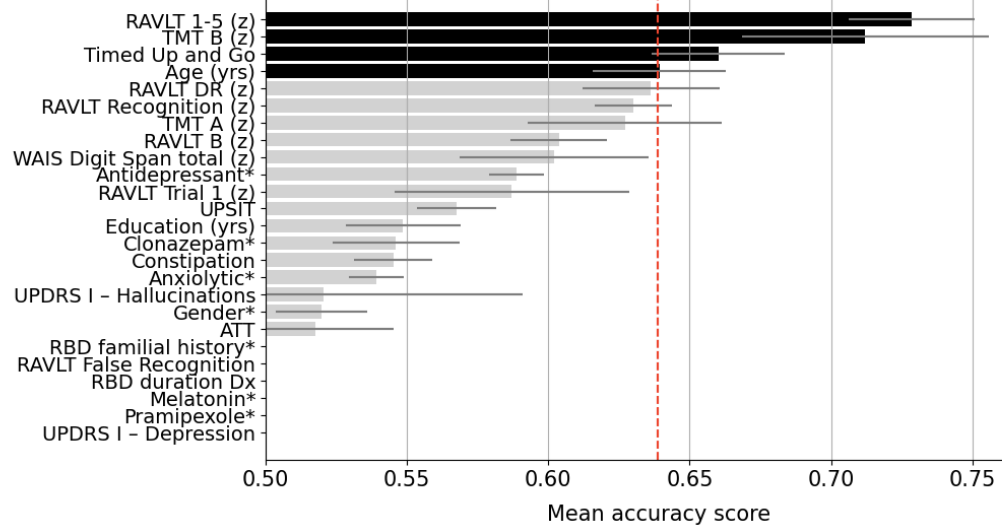


**Figure 2 – DLB vs all iRBD classification. A) Multifeature Random Forest Model:** Feature Appearance vs Importance through 10 seeds. The scatterplot presents the selected variables after exhaustive feature selection. Each point represents a distinct feature in this multivariate model, with its importance to the model indicated by its vertical position and its frequency of selection represented by its horizontal position. Features with an importance score greater than 0.5 are labeled within the plot to emphasize their relevance. The performance of the model also appears inside the scatter plot. **B) Univariate Random Forest Model:** Here, the mean accuracy score of individual features is plotted when each is used alone within the model (i.e., mean of 50 bootstrapping for subject balancing, with 5 fold cross-validation each). The red line across the graph denotes the threshold for statistical significance ( $p < 0.05$ ). Features that surpass this threshold appear in black and are statistically significant. MCI, Mild Cognitive Impairment; RAVLT 1-5, Rey Auditory-Verbal Learning Test sum of trials 1 to 5; RAVLT B, RAVLT List B; RAVLT DR, RAVLT delayed recall; TMT B, Trail Making Test part B; UPDRS, Unified Parkinson's Disease Rating Scale; UPSIT, University of Pennsylvania Smell Identification Test; WAIS, Wechsler Adult Intelligence Scale Third Edition; (z), z score; \*, categorical variable.

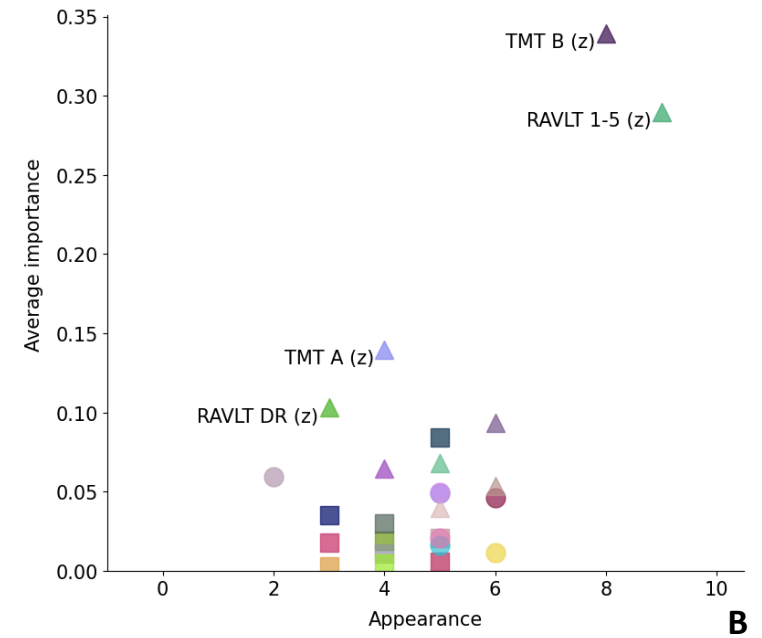


**Figure 3 – DLB vs Controls classification. A) Multifeature Random Forest Model:** Feature Appearance vs Importance through 10 seeds. The scatterplot presents the selected variables after exhaustive feature selection. Each point represents a distinct feature in this multivariate model, with its importance to the model indicated by its vertical position and its frequency of selection represented by its horizontal position. Features with an importance score greater than 0.2 are labeled within the plot to emphasize their relevance. The performance of the model also appears inside the scatter plot. Note that in the data cleaning phase, 'Other clinical markers' were removed due to missing data from control subjects. **B) Univariate Random Forest Model:** Here, the mean accuracy score of individual features is plotted when each is used alone within the RF model (i.e., mean of 50 bootstrapping for subject balancing, with 5 folds cross-validation each). The red line across the graph denotes the threshold for statistical significance ( $p < 0.05$ ). Features that surpass this threshold appear in black and are statistically significant. RAVLT 1-5, Rey Auditory-Verbal Learning Test sum of trials 1 to 5; RAVLT B, RAVLT List B; RAVLT DR, RAVLT delayed recall; TMT A, Trail Making Test part A; TMT B, TMT part B; WAIS, Wechsler Adult Intelligence Scale Third Edition; (z), z score; \*, categorical variable.

DLB vs PD-noMCI

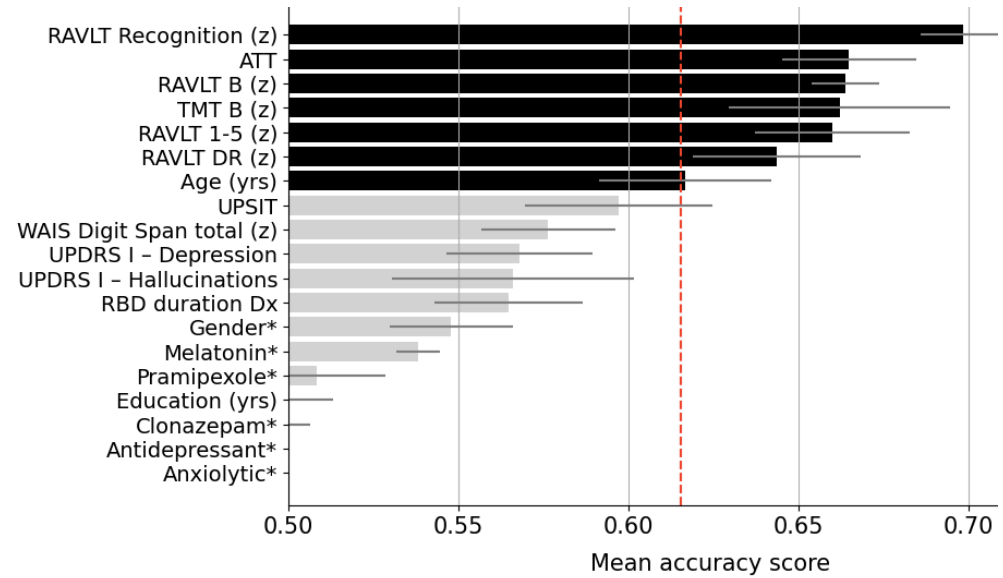


A

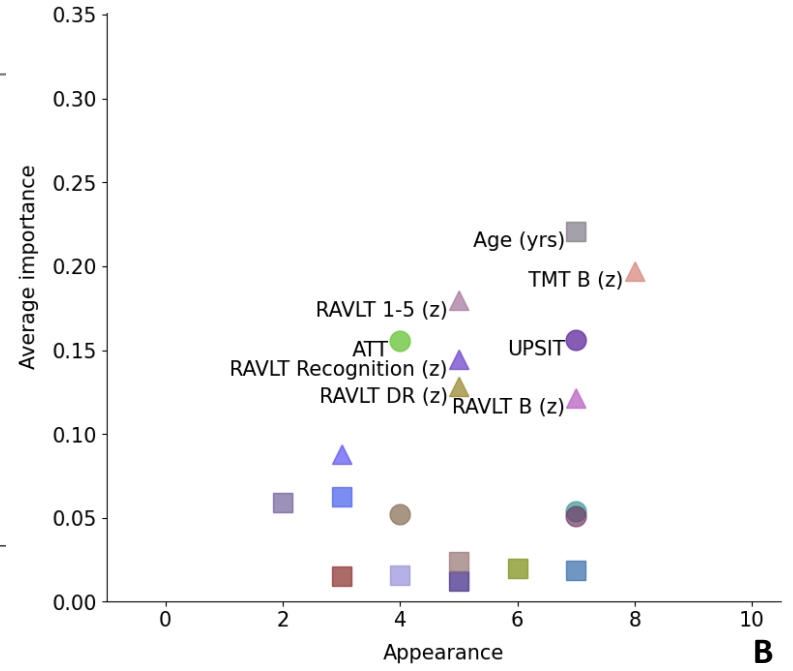


B

DLB vs iRBD-noMCI

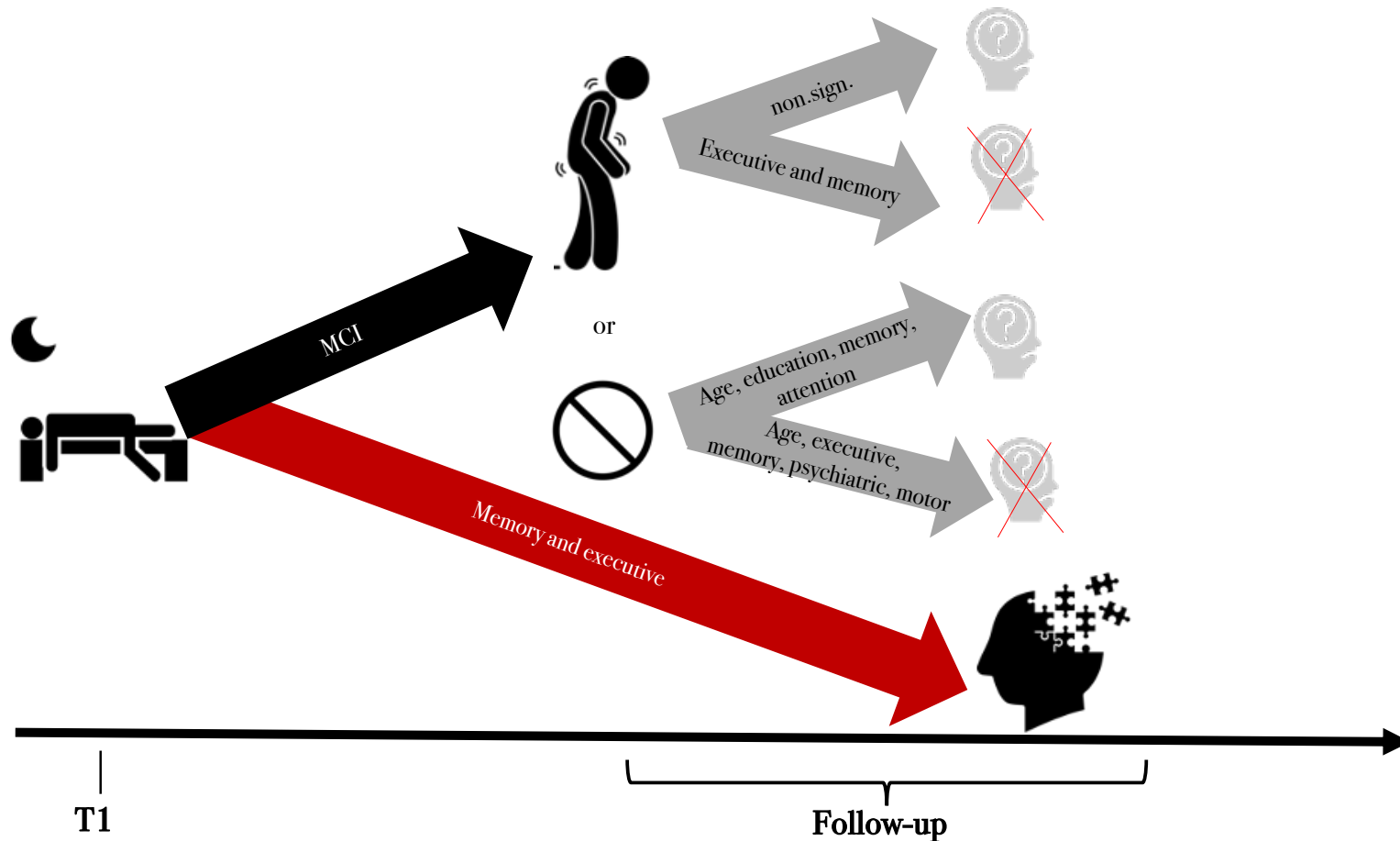


A



B

**Figure 4 – Classifications for the conditions without cognitive impairment at follow-up (i.e., MCI) – DLB vs PD-noMCI and DLB vs iRBD-noMCI. A) Univariate Random Forest Model:** The mean accuracy score of individual features is plotted when each is used alone within the RF model (i.e., mean of 50 bootstrapping for subject balancing, with 5 fold cross-validation each). The red line across the graph denotes the threshold for statistical significance ( $p < 0.05$ ). Features that surpass this threshold appear in black and are statistically significant. **B) Multifeature Random Forest Model:** The scatterplot presents the selected variables after exhaustive feature selection - feature appearance vs importance through the 10 seeds. Each point represents a distinct feature in this multivariate model, with its importance to the model indicated by its vertical position and its frequency of selection represented by its horizontal position. Features with an importance score greater than 0.1 are labeled within the plot to emphasize their relevance. The performance of the model also appears inside the scatter plot. ATT, Alternate Tap Test; RAVLT 1-5, Rey Auditory-Verbal Learning Test sum of trials 1 to 5; RAVLT B, RAVLT List B; RAVLT IR, RAVLT immediate recall; RAVLT DR, RAVLT delayed recall; RBD, REM sleep behaviour disorder; TMT A, Trail Making Test part A; TMT B, TMT part B; UPDRS, Unified Parkinson’s Disease Rating Scale; UPSIT, University of Pennsylvania Smell Identification Test; WAIS, Wechsler Adult Intelligence Scale Third Edition; (z), z score; \*, categorical variable.



**Figure 5 – Comparative overview of optimal features in the RF model for classifying dementia-first progression from other trajectories at follow-up.** Black arrow – Dementia with Lewy Bodies (DLB) vs Parkinson's Disease (PD) or stable idiopathic state (iRBD); grey arrows – DLB vs absence/presence of Mild Cognitive Impairment (MCI) in PD or iRBD. The red arrow corresponds to the DLB vs Controls classification.



## References

1. Miglis, M.G., et al., Biomarkers of conversion to  $\alpha$ -synucleinopathy in isolated rapid-eye-movement sleep behaviour disorder. *The Lancet Neurology*, 2021. 20(8): p. 671-684.
2. Postuma, R.B., et al., Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study. *Brain*, 2019. 142(3): p. 744-759.
3. Anang, J.B., et al., Predictors of dementia in Parkinson disease: a prospective cohort study. *Neurology*, 2014. 83(14): p. 1253-60.
4. Aarsland, D., et al., Cognitive decline in Parkinson disease. *Nat Rev Neurol*, 2017. 13(4): p. 217-231.
5. Genier Marchand, D., et al., Detecting the Cognitive Prodrome of Dementia with Lewy Bodies: A Prospective Study of REM Sleep Behavior Disorder. *Sleep*, 2017. 40(1).
6. Genier Marchand, D., et al., How does dementia with Lewy bodies start? prodromal cognitive changes in REM sleep behavior disorder. *Ann Neurol*, 2018. 83(5): p. 1016-1026.
7. Joza, S.H., MT; Jung, KY; Kunz, D; Arnaldi, D; Lee, JY; Ferini-Strambi, L; Antelmi, E; Sixel-Döring, F; Cochen De Cock, V; Montplaisir, JY; Welch, J; Kim, HJ; Bes, F; Mattioli, P; Woo, KA; Marelli, S; Plazzi, G; Mollenhauer, B; Pelletier, A; Razzaque, J; Sunwoo, JS; Girtler, N; Trenkwalder, C; Gagnon, JF; Postuma, RB; International REM Sleep Behavior Disorder Study Group., Prodromal dementia with Lewy bodies in REM sleep behavior disorder: A multicenter study. . *Alzheimer's & Dementia* 2023.
8. Ij, H., Statistics versus machine learning. *Nat Methods*, 2018. 15(4): p. 233.
9. Abu-Mostafa, Y.S., M. Magdon-Ismail, and H.-T. Lin, Learning from data. Vol. 4. 2012: AMLBook New York.
10. Bzdok, D. and A. Meyer-Lindenberg, Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging*, 2018. 3(3): p. 223-230.
11. Bzdok, D., Classical Statistics and Statistical Learning in Imaging Neuroscience. *Front Neurosci*, 2017. 11: p. 543.
12. Loubna Mekki Berrada, G.A., Jean-François Gagnon, Karim Jerbi, Prediction of cognitive decline in synucleinopathies from EEG and neuropsychological data: A review of machine learning approaches. en prep, 2023.

13. Medicine, A.A.o.S., International Classification of Sleep Disorders. 3 ed. 2014, Darien, IL: American Academy of Sleep Medicine.
14. Montplaisir, J., et al., Polysomnographic diagnosis of idiopathic REM sleep behavior disorder. *Movement disorders*, 2010. 25(13): p. 2044-2051.
15. Association, A.P., Diagnostic and statistical manual of mental disorders (DSM-5®). 2013: American Psychiatric Pub.
16. Postuma, R.B., et al., MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord*, 2015. 30(12): p. 1591-601.
17. Postuma, R.B., et al., Parkinson risk in idiopathic REM sleep behavior disorder: preparing for neuroprotective trials. *Neurology*, 2015. 84(11): p. 1104-1113.
18. Rahayel, S., et al., A prodromal brain-clinical pattern of cognition in Synucleinopathies. *Annals of Neurology*, 2021. 89(2): p. 341-357.
19. Strauss, E., Sherman, E. M., & Spreen, O., A compendium of neuropsychological tests: Administration, norms, and commentary. . 2006: American chemical society.
20. Broadbent, D.E., et al., The cognitive failures questionnaire (CFQ) and its correlates. *British journal of clinical psychology*, 1982. 21(1): p. 1-16.
21. Gagnon, J.F., et al., Mild cognitive impairment in rapid eye movement sleep behavior disorder and Parkinson's disease. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 2009. 66(1): p. 39-47.
22. Rahayel, S., et al., Cortical and subcortical gray matter bases of cognitive deficits in REM sleep behavior disorder. *Neurology*, 2018. 90(20): p. e1759-e1770.
23. Gilman, S., et al., Second consensus statement on the diagnosis of multiple system atrophy. *Neurology*, 2008. 71(9): p. 670-676.
24. Postuma, R.B., et al., Olfaction and color vision identify impending neurodegeneration in rapid eye movement sleep behavior disorder. *Annals of neurology*, 2011. 69(5): p. 811-818.
25. Postuma, R.B., et al., How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder. *Brain*, 2012. 135(6): p. 1860-1870.
26. Mohri, M., A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*. [SI]. 2012, The MIT Press.
27. Vemuri, V.K., *The Hundred-Page Machine Learning Book: by Andriy Burkov*. 2020: Quebec City, Canada. p. 160.

28. Singh, A., Thakur, N., & Sharma, A. , A review of supervised machine learning algorithms. In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom).Ieee. , 2016, March: p. 1310-1315.
29. Thölkea, P., et al., Class imbalance should not throw you off balance: Choosing classifiers and performance metrics for brain decoding with imbalanced data. 2022.
30. Breiman, L., Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). Statistical Science, 2001. 16(3).
31. Donoho, D., 50 years of data science. Journal of Computational and Graphical Statistics, 2017. 26(4): p. 745-766.
32. Bzdok, D. and J.P. Ioannidis, Exploration, inference, and prediction in neuroscience and biomedicine. Trends in neurosciences, 2019. 42(4): p. 251-262.
33. Schenck, C.H., et al., A case of REM sleep behavior disorder with autopsy-confirmed Alzheimer's disease: postmortem brain stem histochemical analyses. Biol Psychiatry, 1996. 40(5): p. 422-5.
34. Iranzo, A., et al., Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: a descriptive study. The Lancet Neurology, 2006. 5(7): p. 572-577.
35. Postuma, R., et al., Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder. Neurology, 2009. 72(15): p. 1296-1300.
36. Gagnon, J.-F., J.-A. Bertrand, and D. Génier Marchand, Cognition in rapid eye movement sleep behavior disorder. Frontiers in neurology, 2012. 3: p. 82.
37. Fereshtehnejad, S.M., et al., Evolution of prodromal Parkinson's disease and dementia with Lewy bodies: a prospective study. Brain, 2019. 142(7): p. 2051-2067.
38. Zhang, J.-R., et al., Rapid eye movement sleep behavior disorder symptoms correlate with domains of cognitive impairment in Parkinson's disease. Chinese Medical Journal, 2016. 129(04): p. 379-385, C1.
39. Gauthier, S., et al., Mild cognitive impairment. Lancet, 2006. 367(9518): p. 1262-70.
40. Fischer, P., et al., Conversion from subtypes of mild cognitive impairment to Alzheimer dementia. Neurology, 2007. 68(4): p. 288-291.
41. Ganguli, M., et al., Mild cognitive impairment, amnesic type: an epidemiologic study. Neurology, 2004. 63(1): p. 115-121.

42. Janvin, C.C., et al., Subtypes of mild cognitive impairment in Parkinson's disease: progression to dementia. *Movement disorders: official journal of the Movement Disorder Society*, 2006. 21(9): p. 1343-1349.
43. Garcia-Garcia, D., et al., Posterior parietooccipital hypometabolism may differentiate mild cognitive impairment from dementia in Parkinson's disease. *European journal of nuclear medicine and molecular imaging*, 2012. 39: p. 1767-1777.
44. Huang, C., et al., Metabolic abnormalities associated with mild cognitive impairment in Parkinson disease. *Neurology*, 2008. 70(16 Pt 2): p. 1470-7.
45. Agosta, F., et al., Mild cognitive impairment in Parkinson's disease is associated with a distributed pattern of brain white matter damage. *Hum Brain Mapp*, 2014. 35(5): p. 1921-9.
46. Braak, H., et al., Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of aging*, 2003. 24(2): p. 197-211.
47. Beach, T.G., et al., Unified staging system for Lewy body disorders: correlation with nigrostriatal degeneration, cognitive impairment and motor dysfunction. *Acta neuropathologica*, 2009. 117: p. 613-634.
48. Adler, C.H. and T.G. Beach, Neuropathological basis of nonmotor manifestations of Parkinson's disease. *Mov Disord*, 2016. 31(8): p. 1114-9.
49. Berg, D., et al., Prodromal Parkinson disease subtypes—key to understanding heterogeneity. *Nature Reviews Neurology*, 2021. 17(6): p. 349-361.
50. Cesari, M., et al., A data-driven system to identify REM sleep behavior disorder and to predict its progression from the prodromal stage in Parkinson's disease. *Sleep Med*, 2021. 77: p. 238-248.

## Supplementary Information

**S1 Table** – Baseline performance and prevalence for all categories of patients that received a follow -up diagnosis.

	<b>all RBD (n=156)</b>	<b>iRBD- nMCI (0)</b>	<b>iRBD- MCI (1)</b>	<b>PD-NC (2)</b>	<b>PD- MCI (3)</b>	<b>DLB (4)</b>	<b>Controls (n=74)</b>
<b>Sociodemographic and clinical</b>							
<i><b>Sociodemographic</b></i>							
Age (yrs)	66,46 +/- 7,20	64,53 +/- 6,87	66,17 +/- 6,18	64,90 +/- 7,98	67,62 +/- 6,72	72,03 +/- 6,24	67,03 +/- 8,30
Gender* (% male)	76,28% (119/156)	78,78% (52/66)	76,67% (23/30)	77,78% (14/18)	81,25% (13/16)	65,38% (17/26)	70,27%(52 /74)
Education (yrs)	13,86+/- 3,75	14,8 +/- 3,62	12,87+/- 3,72	13,89 +/- 3,38	12,19 +/- 3,66	13,54 +/- 3,94	14,77+/- 3,23
Follow-up duration (yrs)	4,37 +/- 3,06	4,12 +/- 2,88	5,56 +/- 3,34	3,85+/- 2,63	4,89 +/- 3,18	3,65 +/- 3,17	----
<i><b>Clinical</b></i>							
RBD duration Sx	7,72 +/- 7,55	7,81 +/- 7,62	6,96 +/- 7,23	8,94 +/- 7,2	10,5 +/- 11,65	5,68 +/- 3,46	----
RBD duration Dx	1,29 +/- 2,56	0,86 +/- 1,74	0,86 +/- 2,36	1,83 +/- 3,22	2,87 +/- 4,42	1,5 +/- 2,20	----
% REM tonic EMG	51,60 +/- 34,23	43,46 +/- 31,42	59,89 +/- 39,35	58,85 +/- 32,84	60,45 +/- 37,41	52,64 +/- 34,50	----
% REM phasic EMG	37,99 +/- 22,83	35,88 +/- 20,96	47,48 +/- 27,05	40,53 +/- 24,95	31,63 +/- 18,19	33,52 +/- 22,55	----
Familial History* (yes)	45,51% (71/156)	50% (33/66)	43,33% (13/30)	33,33% (6/18)	50% (8/16)	42,31% (11/26)	----
RBD Medication* (yes)	40,38% (63/156)	34,84% (23/66)	50% (15/30)	44,44% (8/18)	56,25% (9/16)	30,77% (8/26)	----
Clonazepam* (yes)	35,26% (55/156)	28,79% (19/66)	46,47% (14/30)	38,89% (7/18)	56,25% (9/16)	23,08% (6/26)	----
Melatonin* (yes)	12,18% (19/156)	18,18% (12/66)	10% (3/30)	11,11% (2/18)	0% (0/16)	7,69% (2/26)	----
Pramipexole* (yes)	3,21% (5/156)	1,52% (1/66)	3,33% (1/30)	5,56% (1/18)	0% (0/16)	7,69% (2/26)	----
<b>Cognition</b>							
MCI* (yes)	37,82% (59/156)	13,64% (9/66)	60% (18/30)	0% (0/18)	56,25% (9/16)	88,46% (23/26)	----
<i>Single</i>	19,87% (31/156)	7,58 (5/66)	30% (9/30)	0% (0/18)	37,5% (6/16)	42,31% (11/26)	----
<i>Multiple</i>	17,95% (28/156)	6,06% (4/66)	30% (9/30)	0% (0/18)	18,75% (3/16)	46,15% (12/26)	----

<i>Executive</i>	12,18% (19/156)	1,52% (1/66)	16,67% (5/30)	0% (0/18)	31,25% (5/16)	30,77% (8/26)	----
<i>Memory</i>	5,77% (9/156)	4,55% (3/66)	10% (3/30)	0% (0/18)	0% (0/16)	11,54% (3/26)	----
<i>Visuospatial</i>	1,92% (3/156)	1,52% (1/66)	3,33% (1/30)	0% (0/18)	6,25% (1/16)	0% (0/26)	----
<i>Executive + memory</i>	10,26% (16/156)	3,03% (2/66)	26,67% (8/30)	0% (0/18)	12,5% (2/16)	15,38% (4/26)	----
<i>Executive + visuospatial</i>	3,21% (5/156)	3,03% (2/66)	3,33% (1/30)	0% (0/18)	0% (0/16)	7,69% (2/26)	----
<i>Memory + visuospatial</i>	0% (0/156)	0% (0/66)	0% (0/30)	0% (0/18)	0% (0/16)	0% (0/26)	----
<i>Exe + mem + visuo</i>	4,49% (7/156)	0% (0/66)	0% (0/30)	0% (0/18)	6,25% (1/16)	23,08% (6/26)	----
<i>Amnestic</i>	20,51% (32/156)	7,58% (5/66)	36,67% (11/30)	0% (0/18)	18,75% (3/16)	50% (13/26)	----
<i>Non-amnestic</i>	17,31% (27/156)	6,06% (4/66)	23,33% (7/30)	0% (0/18)	37,5% (6/16)	38,46% (10/26)	----
<b>Screening</b>							
MMSE	28,42 +/- 4,69	28,70 +/- 5,09	27,97 +/- 1,52	29,06 +/- 0,80	28,94 +/- 1,0	27,42 +/- 7,56	28,82 +/- 11,10
MoCA	25,91 +/- 8,47	26,87 +/- 6,78	24,93 +/- 3,11	26,73 +/- 10,44	26,08 +/- 11,76	23,7 +/- 10,38	27,14 +/- 13,60
<b>Attention and executive functions</b>							
WAIS-III Digit Span forward (z)	-0,13 +/- 0,91	0,057 +/- 0,82	-0,281 +/- 1,02	0,11 +/- 0,92	-0,59 +/- 0,71	-0,32 +/- 0,99	0,078 +/- 0,79
WAIS-III Digit Span backward (z)	0,025 +/- 0,90	0,13 +/- 0,86	0,048 +/- 1,04	0,33 +/- 0,93	-0,32 +/- 0,79	-0,25 +/- 0,84	0,31 +/- 0,95
WAIS-III Digit Span total (z)	-0,019 +/- 0,95	0,12 +/- 0,92	-0,054 +/- 1,14	0,33 +/- 0,87	-0,54 +/- 0,82	-0,24 +/- 0,82	0,22 +/- 0,86
TMT A (z)	-0,45 +/- 2,05	0,12 +/- 1,50	-0,42 +/- 1,63	0,02 +/- 1,42	-0,78 +/- 1,22	-2,06 +/- 3,33	0,13 +/- 1,0
TMT B (z)	-0,76 +/- 1,67	-0,17 +/- 1,54	-1,06 +/- 1,61	-0,034 +/- 1,38	-1,41 +/- 1,47	-2,00 +/- 1,46	0,06 +/- 1,23
TMT B-A (z)	-0,73 +/- 2,01	0,06 +/- 1,07	-1,25 +/- 2,3	-0,09 +/- 1,03	-1,53 +/- 2,48	-2,15 +/- 2,67	0,044 +/- 1,19
TMT infractions	0,625 +/- 1,19	0,31 +/- 0,63	0,69 +/- 1,12	0,29 +/- 0,46	0,75 +/- 1,39	1,52 +/- 1,96	0,35 +/- 0,58
Semantic Fluency (z)	-0,87 +/- 0,84	-0,53 +/- 0,69	-0,95 +/- 0,82	-0,78 +/- 0,87	-1,51 +/- 0,85	-1,36 +/- 0,84	-0,46 +/- 0,64
Animals (z)	-0,062 +/- 1,26	0,21 +/- 1,1	0,023 +/- 1,15	0,31 +/- 1,67	-0,61 +/- 1,28	-0,78 +/- 1,12	0,37 +/- 0,85
Phonetic Fluency (z)	-0,42 +/- 0,94	-0,24 +/- 0,90	-0,509 +/- 0,90	-0,002 +/- 0,91	-0,876 +/- 0,83	-0,83 +/- 1,01	-0,08 +/- 0,85
<b>Episodic verbal memory</b>							
RAVLT 1-5 (z)	-0,11 +/- 1,31	0,17 +/- 1,14	-0,70 +/- 1,38	0,71 +/- 0,96	-0,10 +/- 1,51	-0,74 +/- 1,26	0,94 +/- 1,08

RAVLT Trial 1 (z)	-0,60 +/- 1,08	-0,57 +/- 0,97	-0,78 +/- 1,15	-0,045 +/- 0,96	-0,64 +/- 1,45	-0,85 +/- 1,03	-0,07 +/- 1,12
RAVLT B (z)	-0,24 +/- 1,21	0,03 +/- 1,13	-0,91 +/- 1,06	0,41 +/- 0,94	-0,70 +/- 1,37	-0,31 +/- 1,24	0,32 +/- 1,19
RAVLT IR (z)	-0,27 +/- 1,06	0,072 +/- 0,94	-0,74 +/- 1,04	0,11 +/- 0,87	-0,30 +/- 1,02	-0,83 +/- 1,11	0,39 +/- 0,89
RAVLT DR (z)	-0,09 +/- 1,20	0,34 +/- 1,12	-0,67 +/- 1,01	0,19 +/- 1,01	-0,145 +/- 1,20	-0,64 +/- 1,28	0,59 +/- 1,11
RAVLT Recognition (z)	-0,016 +/- 1,24	0,23 +/- 1,08	-0,139 +/- 1,35	0,26 +/- 1,08	-0,56 +/- 1,27	-0,36 +/- 1,43	0,39 +/- 0,87
RAVLT False Recognition	2,55 +/- 2,70	1,97 +/- 1,98	3,63 +/- 3,72	1,89 +/- 1,71	2,63 +/- 2,33	3,19 +/- 3,23	1,43 +/- 2,04
RAVLT Intrusions	1,63 +/- 2,65	1,28 +/- 1,81	1,5 +/- 3,09	1,17 +/- 1,46	1,56 +/- 1,63	3,08 +/- 4,29	1,5 +/- 1,85
<b>Visuospatial abilities</b>							
Rey Complex Figure Copy (z)	0,078 +/- 1,05	0,39 +/- 0,82	-0,186 +/- 0,98	0,42 +/- 0,55	0,008 +/- 1,33	-0,64 +/- 1,30	0,32 +/- 0,82
WAIS-III Block Design (z) or VOSP	0,11 +/- 0,81	0,41 +/- 0,74	-0,15 +/- 0,67	0,33 +/- 0,97	0,29 +/- 0,85	-0,50 +/- 0,81	0,49 +/- 0,95
Bells test (z)	-0,01 +/- 0,89	0,067 +/- 0,74	-0,035 +/- 0,92	0,27 +/- 1,02	-0,39 +/- 0,96	-0,13 +/- 1,07	1,62 +/- 2,01
<b>Psychiatric, sleep and markers</b>							
<b>Psychiatric</b>							
BDI-II	10,23 +/- 7,77	9,22 +/- 6,98	9,86 +/- 8,36	8,56 +/- 7,40	10 +/- 6,69	14,89 +/- 9,50	5,69 +/- 5,57
BDI-PC	2,40 +/- 2,53	2,25 +/- 2,62	2,33 +/- 2,24	2,06 +/- 2,38	1,54 +/- 1,95	3,63 +/- 2,95	1,53 +/- 2,05
BAI	7,02 +/- 6,93	5,85 +/- 5,57	7,09 +/- 6,20	5,8 +/- 5,67	5,18 +/- 5,06	12,21 +/- 10,91	4,38 +/- 4,42
BAI-PC	2,62 +/- 2,86	2,09 +/- 2,16	2,13 +/- 2,19	3,33 +/- 3,57	1,55 +/- 1,95	4,74 +/- 4,28	1,55 +/- 1,91
UPDRS I – Hallucinations	0,20 +/- 0,53	0,08 +/- 0,32	0,3 +/- 0,58	0,11 +/- 0,47	0,19 +/- 0,40	0,46 +/- 0,85	----
UPDRS I – Depression	0,47 +/- 0,77	0,40 +/- 0,74	0,53 +/- 0,78	0,39 +/- 0,61	0,69 +/- 0,95	0,54 +/- 0,81	----
UPDRS I – Apathy	0,40 +/- 0,74	0,23 +/- 0,52	0,37 +/- 0,81	0,61 +/- 0,85	0,75 +/- 0,93	0,5 +/- 0,86	----
Depression* (yes)	33,33% (52/156)	30,30% (20/66)	40% (12/30)	22,22% (4/18)	31,25% (5/16)	42,31% (11/26)	5,41% (4/74)
Anxiety*(yes)	19,87% (31/156)	18,18% (12/66)	23,33% (7/30)	11,11% (2/18)	12,5% (2/16)	30,77% (8/26)	2,70% (2/74)
Psychiatric History* (yes)	42,94% (67/156)	46,97% (31/66)	46,67% (14/30)	27,78% (5/18)	43,75% (7/16)	38,46% (10/26)	----
Antidepressant* (yes)	25,64% (40/156)	24,24% (16/66)	30% (9/30)	11,11% (2/18)	31,25% (5/16)	30,77% (8/26)	----
Anxiolytics*(yes)	11,54% (18/156)	9,09% (6/66)	23,33% (7/30)	0% (0/18)	12,5% (2/16)	11,54% (3/26)	----

Psychiatric Medication * (yes)	32,05% (50/156)	30,30% (20/66)	43,33% (13/30)	11,11% (2/18)	37,5% (6/16)	34,62% (9/26)	----
<b>Sleep</b>							
ISI	10,23 +/- 6,75	9,54 +/- 6,40	12,05 +/- 7,61	8,25 +/- 6,53	10,86 +/- 6,96	11,5 +/- 6,66	6,15 +/- 4,74
ESS	7,79 +/- 5,35	7,85 +/- 5,38	7,54 +/- 6,77	6,77 +/- 5,30	9,5 +/- 5,85	7,93 +/- 5,58	6,36 +/- 4,22
<b>Perceptual</b>							
UPSIT	0,72 +/- 0,28	0,77 +/- 0,27	0,68 +/- 0,31	0,71 +/- 0,26	0,76 +/- 0,30	0,615 +/- 0,27	0,97 +/- 0,35
FM-100	1,06 +/- 0,51	0,96 +/- 0,45	1,05 +/- 0,57	1,09 +/- 0,59	1,23 +/- 0,38	1,21 +/- 0,55	0,82 +/- 0,34
<b>Motor</b>							
UPDRS-III	3,08 +/- 3,39	2,03 +/- 2,20	3,03 +/- 2,39	4,11 +/- 4,63	3,81 +/- 4,31	4,68 +/- 4,48	----
Purdue Pegboard	23,32 +/- 4,77	24,05 +/- 3,51	22,69 +/- 4,97	24,39 +/- 3,36	22,5 +/- 3,63	21,83 +/- 7,12	25,6 +/- 8,86
ATT	180,73 +/- 35,95	186,89 +/- 28,06	187,11 +/- 56,72	178,34 +/- 29,57	176,91 +/- 17,18	162,25 +/- 31,30	194,04 +/- 69,90
Timed Up and Go	6,45 +/- 1,33	6,24 +/- 1,06	6,23 +/- 1,84	6,39 +/- 0,83	6,56 +/- 1,21	7,18 +/- 1,26	6,25 +/- 2,16
<b>Autonomic</b>							
Orthostatic hypotension	0,33 +/- 0,56	0,24 +/- 0,48	0,38 +/- 0,58	0,22 +/- 0,43	0,28 +/- 0,45	0,6 +/- 0,76	0,1 +/- 0,12
SBPD	11,96 +/- 14,32	9,46 +/- 11,17	9,93 +/- 16,47	12,39 +/- 13,66	14,87 +/- 19,83	18,44 +/- 14,18	5,67 +/- 2,72
Urinary	0,47 +/- 0,67	0,43 +/- 0,66	0,45 +/- 0,81	0,39 +/- 0,50	0,47 +/- 0,50	0,68 +/- 0,69	0,00 +/- 0,00
Sexuality (Erectile Sx)	1,59 +/- 1,45	1,4 +/- 1,44	1,48 +/- 1,50	1,57 +/- 1,35	1,92 +/- 1,55	2,00 +/- 1,50	0,125 +/- 0,12
Constipation	0,80 +/- 0,91	0,66 +/- 0,89	0,81 +/- 0,91	0,64 +/- 0,68	0,93 +/- 0,88	1,17 +/- 1,09	0,2 +/- 0,23

Anxiety, by the use of anxiolytics or a severity score of moderate at the MDS-UPDRS-I-4; ATT, Alternate Tap Test; BAI, Beck Anxiety Inventory; BAI-PC; BAI Primary Care; BDI-II, Beck Depression Inventory, Second Edition; BDI-PC, BDI Primary Care; Depression, by the use of antidepressants or a severity score of moderate at the UPDRS-I-3; EMG, Electromyography; ESS, Epworth Sleepiness Scale; FM-100, Farnsworth-Munsell 100-hue test; ISI, Insomnia Severity Index; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; Phonetic Fluency, letters P, F, L in French or F, A, S in English, one minute each; Q, question; RAVLT 1-5, Rey Auditory-Verbal Learning Test sum of trials 1 to 5; RAVLT B, RAVLT List B; RAVLT IR, RAVLT immediate recall; RAVLT DR, RAVLT delayed recall; RBD Medication, if the patient is taking clonazepam, dopaminergic agonists, or melatonin; RBD, REM Sleep Behavior Disorder;



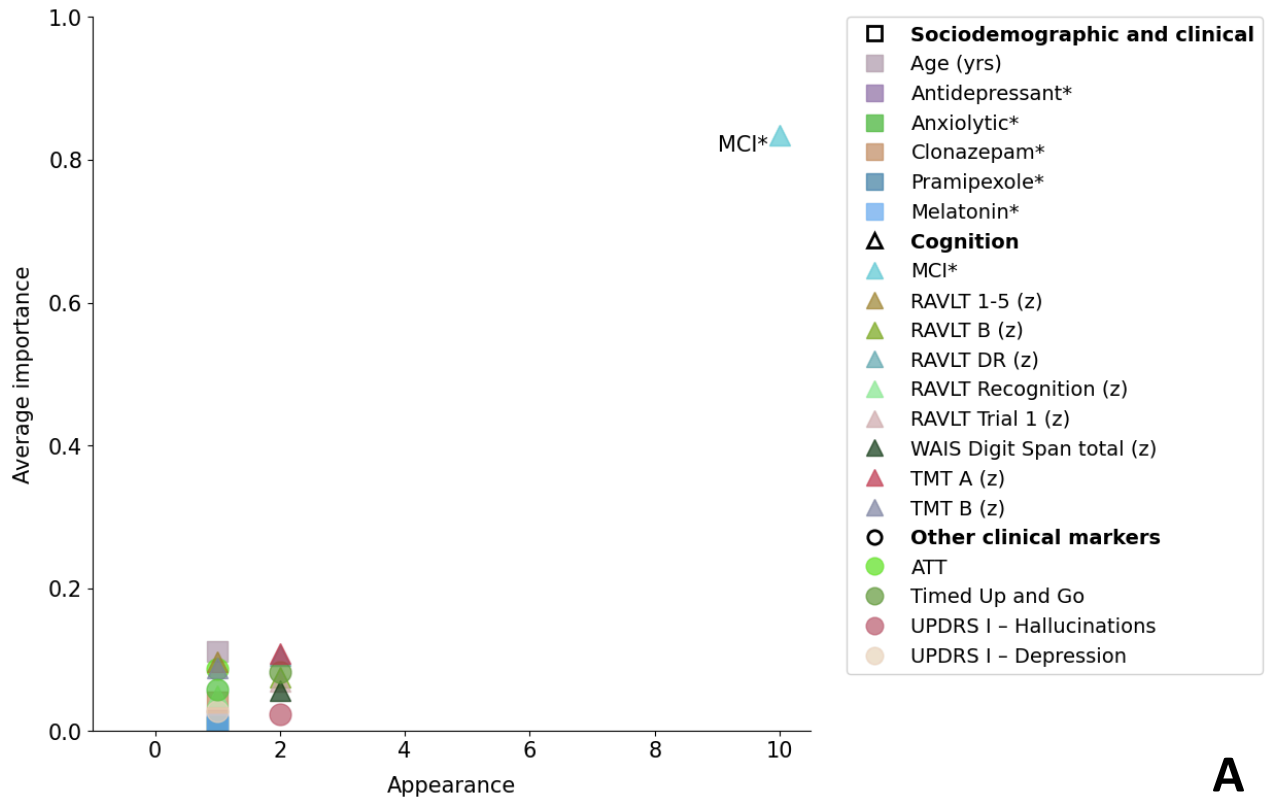
Rey Figure Copy, Rey-Osterrieth Complex Figure copy time; Semantic Fluency, animals or fruits/vegetables, one minute each; Sx, Symptoms; SBPD, Systolic blood pressure drop ;TMT A, Trail Making Test part A; TMT B-A, TMT part B minus part A; TMT B, TMT part B; UPDRS, Unified Parkinson's Disease Rating Scale; UPSIT, University of Pennsylvania Smell Identification Test; VOSP, Visual Object and Space Perception Battery (number location and object decision); WAIS-III, Wechsler Adult Intelligence Scale Third Edition; (z), Z score; \*, categorical variable; \*\*, continuous and categorical variables (e.g. for BAI\*\*(>=10 and >=19), there are three variables: 1) BAI raw score (continuous), 2) BAI proportion of individuals with a score >=10 (categorical) and 3) BAI proportion of individuals with a score >=19 (categorical)); <sup>1</sup>assessed using a structured interview modeled on the Multiple System Atrophy Rating Scale (Wenning et al., 2004).

**S2 Table** – Features left after data cleaning for all classifications – used as input for the ML.

	<b>Features</b>
<b>DLB vs PD noMCI</b>	Age (yrs), ATT, Antidepressant*, Anxiolytic*, Clonazepam*, Constipation, Pramipexole*, RBD duration Dx, Gender*, RBD familial history*, Melatonin*, RAVLT 1-5 (z), RAVLT B (z), RAVLT DR (z), RAVLT Recognition (z), RAVLT Trial 1 (z), RAVLT False Recognition, Education (yrs), WAIS Digit Span total (z), Timed Up and Go, TMT A (z), TMT B (z), UPDRS I – Hallucinations, UPDRS I – Depression, UPSIT.
<b>DLB vs iRBD-noMCI</b>	Age (yrs), Gender*, Education (yrs), RBD duration Dx, Antidepressant*, Anxiolytic*, Clonazepam*, Melatonin*, Pramipexole*, WAIS Digit Span total (z), TMT B (z), RAVLT 1-5 (z), RAVLT B (z), RAVLT DR (z), RAVLT Recognition (z), UPDRS I – Depression, UPDRS I – Hallucinations, UPDRS I – Apathy, UPSIT, ATT.
<b>DLB vs Controls</b>	Age (yrs), Gender*, Education (yrs), WAIS Digit Span total (z), TMT A (z), TMT B (z), RAVLT 1-5 (z), RAVLT Trial 1 (z), RAVLT B (z), RAVLT DR (z), RAVLT Recognition (z), RAVLT False Recognition, MCI*.
<b>DLB vs all PD</b>	Age (yrs), Gender*, Education (yrs), RBD duration Dx, RBD familial history*, Antidepressant*, Anxiolytic*, Clonazepam*, Melatonin*, Pramipexole*, WAIS Digit Span total (z), TMT A (z), TMT B (z), RAVLT 1-5 (z), RAVLT Trial 1 (z), RAVLT B (z), RAVLT DR (z), RAVLT Recognition (z), RAVLT False Recognition, MCI*, UPDRS I – Depression, UPDRS I – Hallucinations, UPDRS I – Apathy, UPSIT, ATT, Timed Up and Go.
<b>DLB vs all RBD</b>	Age (yrs), Gender*, Education (yrs), Antidepressant*, Anxiolytic*, Clonazepam*, Melatonin*, Pramipexole*, WAIS Digit Span total (z), TMT B (z), RAVLT 1-5 (z), RAVLT B (z), RAVLT DR (z), RAVLT Recognition (z), MCI*, UPDRS I – Depression, UPDRS I – Hallucinations, UPDRS I – Apathy, UPSIT.
<b>DLB vs RBD-MCI</b>	Age (yrs), Gender*, Education (yrs), RBD familial history*, Antidepressant*, Anxiolytic*, Clonazepam*, Melatonin*, Pramipexole*, WAIS Digit Span total (z), TMT A (z), TMT B (z), RAVLT 1-5 (z), RAVLT B (z), RAVLT DR (z), RAVLT Recognition (z), RAVLT False Recognition, UPDRS I – Depression, UPDRS I – Hallucinations, UPDRS I – Apathy, UPSIT.
<b>DLB vs PD-MCI</b>	Age (yrs), Gender*, Education (yrs), RBD duration Dx, RBD familial history*, Antidepressant*, Anxiolytic*, Clonazepam*, Melatonin*, Pramipexole*, WAIS Digit Span total (z), TMT A (z), TMT B (z), RAVLT 1-5 (z), RAVLT Trial 1 (z), RAVLT B (z), RAVLT DR (z), RAVLT Recognition (z), RAVLT False Recognition, UPDRS I – Depression, UPDRS I – Hallucinations, UPDRS I – Apathy, UPSIT, ATT, Timed Up and Go.

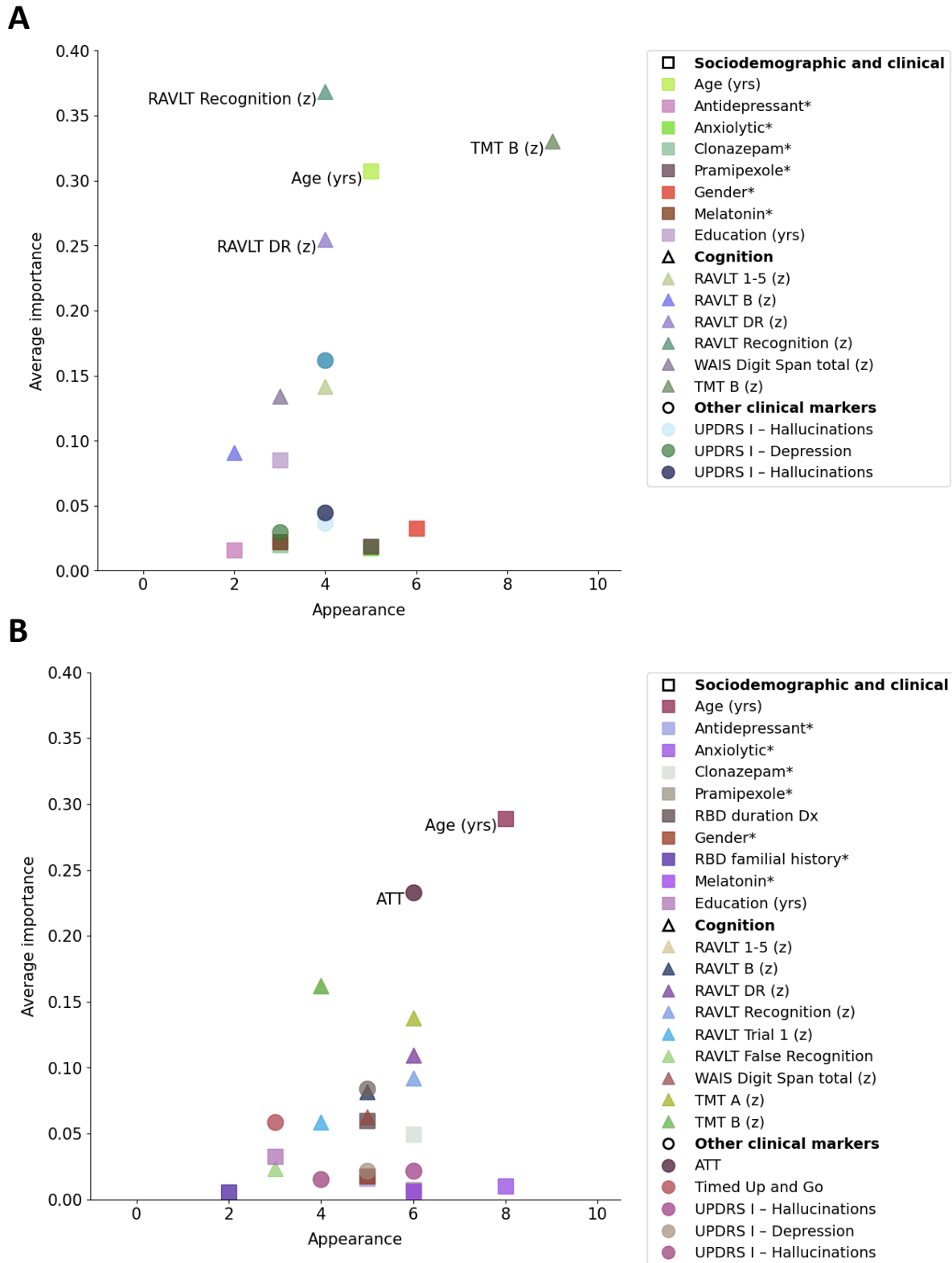
ATT, Alternate Tap Test; DLB, Dementia with Lewy Bodies; MCI, Mild Cognitive Impairment; PD, Parkinson's Disease; RAVLT 1-5, Rey Auditory-Verbal Learning Test sum of trials 1 to 5; RAVLT B, RAVLT List B; RAVLT DR, RAVLT delayed recall; RBD, REM sleep behavior disorder; REM, rapid eye movement; TMT A, Trail Making Test part A; TMT B, TMT part B; UPDRS I, Unified Parkinson's Disease Rating Scale Part I; UPSIT, University of Pennsylvania Smell Identification Test; WAIS, Wechsler Adult Intelligence Scale; (z), z score; \*, categorical variable.

**S1 Figure** – Feature Appearance vs Importance through 10 seeds for the DLB vs all PD classification.



**Note.** The scatterplot presents the selected variables after exhaustive feature selection. Each point represents a distinct feature in this multivariate model, with its importance to the model indicated by its vertical position and its frequency of selection represented by its horizontal position. Features with an importance score greater than 0.5 are labeled within the plot to emphasize their relevance. The performance of the model also appears inside the scatter plot. ATT, Alternate Tap Test; MCI, Mild Cognitive Impairment; RAVLT 1-5, Rey Auditory-Verbal Learning Test sum of trials 1 to 5; RAVLT B, RAVLT List B; RAVLT IR, RAVLT immediate recall; RAVLT DR, RAVLT delayed recall; TMT A, Trail Making Test part A; TMT B, TMT part B; UPDRS, Unified Parkinson’s Disease Rating Scale; WAIS, Wechsler Adult Intelligence Scale Third Edition; (z), z score; \*, categorical variable.

**S2 Figure** – Feature Appearance vs Importance through 10 seeds for the A) DLB vs all iRBD classification and B) DLB vs all PD classification, without using the MCI variable.



**Note.** The scatterplot presents the selected variables after exhaustive feature selection. Each point represents a distinct feature in this multivariate model, with its importance to the model indicated by its vertical position and its frequency of selection represented by its horizontal position. Features with an importance score greater than 0.5 are labeled within the plot to emphasize their relevance. The performance of the model also appears inside the scatter plot. ATT, Alternate Tap Test; MCI, Mild Cognitive Impairment; RAVLT 1-5, Rey Auditory-Verbal Learning Test sum of trials 1 to 5; RAVLT B, RAVLT List B; RAVLT IR, RAVLT immediate recall; RAVLT DR, RAVLT delayed recall; REM, rapid eye movement; RBD, REM sleep behaviour disorder; TMT A, Trail Making Test part A; TMT B, TMT part B; UPDRS, Unified Parkinson's Disease Rating Scale; WAIS, Wechsler Adult Intelligence Scale Third Edition; (z), z score; \*, categorical variable.

## **Chapitre 4 – Discussion générale**

## **Sommaire des résultats principaux de la thèse**

### **Rappel du but général de la thèse**

L'objectif général de cette thèse était d'évaluer la pertinence et le rôle de l'AM dans l'investigation de la prédiction de la démence dans les synucléinopathies et un de leur stade prodromal, le TCSPi. Deux articles découlent de cette thèse : 1) un premier visant à recenser la littérature ayant utilisé l'AM (avec techniques de validation du modèle) en lien avec des données EEG ou cognitives pour identifier et prédire le déclin cognitif dans les synucléinopathies, et 2) un deuxième article empirique visant à identifier, avec l'AM (avec techniques de validation du modèle), la meilleure combinaison de variables sociodémographiques, cliniques et cognitives permettant de prédire la démence dans le TCSPi.

### **Retour sur l'article 1**

Dans ce premier article, nous avons effectué une recherche exhaustive de la littérature scientifique en sélectionnant les études qui ont utilisé l'AM et les variables EEG et cognitives en lien avec le déclin cognitif ou la démence dans les synucléinopathies (TCSPi, MP et DCL). Nous avons retenu les études ayant utilisé un algorithme d'AM pour lequel la capacité de généralisation hors-échantillon d'entraînement a été évaluée (ex : par validation de modèle). La recherche initiale sur Pubmed a généré 3579 études, parmi lesquelles 33 répondaient aux critères d'inclusion, soit : 1) une étude originale, 2) la présence d'une de ces catégories cliniques : TCSPi, MP avec ou sans stade prodromal de la démence défini comme TCSP ou TCL, ou DCL, 3) l'utilisation de données EEG à l'état de repos et/ou de données cognitives (tests neuropsychologiques spécifiques ou de dépistage), 4) l'utilisation d'algorithmes d'AM ayant utilisé une méthode de validation de modèle (ou *cross-validation*), et 5) langue de publication en anglais. La majorité des études retenues ont été conduites entre 2017-2022 : 20 utilisant des données EEG et 14 des données cognitives. Les informations suivantes ont été récoltées pour chaque articles sélectionnés : 1) la classification binaire étudiée, 2) les variables (ou *features*) utilisées, 3) la méthode de sélection des variables (ou *feature selection*), 4) l'algorithme d'AM utilisé, 5) la méthode de validation du modèle, 6) les performances de classification, et 7) les métriques et nombre de sujets. Nous avons considéré toutes

les classifications binaires possibles dans l'étendue des principales synucléinopathies, en plus de celles avec des sujets sains (contrôles). Diverses classifications sont abordées dans les 33 études, majoritairement à partir de données EEG (59%). Les six catégories principales sont : 1) MP vs contrôles (n=13 ; 12 EEG), 2) DCL vs contrôles (n=1 ; 1 EEG), 3) MP vs MP avec démence ou DCL (n=5 ; 0 EEG), 4) MP vs MP avec stade prodromal (n=7 ; 3 EEG), 5) MP évoluant vers un déclin cognitif ou une démence (n=6 ; 2 EEG), et 6) TCSP évoluant vers une MP ou une DCL (n=1 ; 1 EEG).

L'article 1 montre une grande variété de classifications, de métriques et de variables utilisées au sein d'une même modalité. Ainsi, dans la première partie de l'article, nous avons examiné une classification à la fois, présentant, pour une question de clarté, uniquement l'algorithme le plus performant des études ayant testé plusieurs algorithmes. Cela a permis une investigation rigoureuse et une description qualitative des caractéristiques les plus performantes pour chaque classification. Pour distinguer les patients avec une MP des sujets contrôles plusieurs études utilisent différentes mesures spectrales en EEG, mais également des mesures de complexité, ainsi qu'une extraction automatique des composantes par des réseaux de neurones (Khare et al., 2021; Lee et al., 2021; Yuvaraj et al., 2016). Le ralentissement EEG permet de distinguer de les patients avec une DCL des sujets contrôles (ratio  $\theta/\alpha$ ) (Dauwan et al., 2016), ainsi que l'évolution vers un déclin cognitif dans la MP (ex : *global relative median power* en  $\theta$  (Cozac et al., 2016)). Une seule étude a investigué l'évolution du TCSPi vers une MP ou une DCL, à partir de spectrogrammes en utilisant une seule électrode de la région pariétale, obtenant une performance modérée (Ruffini et al., 2019). Les mesures de connectivité cérébrale en EEG sont quant à elles peu utilisées, mais sont sélectionnées comme importantes par les modèles d'AM dans la classification de la présence du TCL ou du TCSP dans la MP, respectivement (Cesari et al., 2021; Chaturvedi et al., 2019). En ce qui concerne les études à partir de données cognitives, ces dernières montrent une importance notable dans la classification du statut cognitif dans la MP, qui est supérieure à celle des données sociodémographiques et cliniques, lorsque combinées dans un modèle d'AM (ex : la fluence verbale sémantique, dans la classification de la MP avec/sans démence) (Byeon, 2020a; Dujardin et al., 2010; Lee et al., 2016). L'ensemble des études qui se sont penchées sur la classification des stades prodromaux dans la MP (ex : TCSPi ou TCL) (Betrouni et al., 2019; Byeon, 2020b; Cesari et al., 2021; Chaturvedi et al., 2019; Chong-Wen et



al., 2022; Jeon et al., 2022; Ortelli et al., 2022) sans l'EEG ont utilisé au moins un test de dépistage cognitif (MoCA, Mini-Mental State Examination (MMSE), 5 items questionnaire) au sein d'un ensemble de variables sociodémographiques, cliniques (ex : motrices, olfactives) et cognitives (dans une seule étude; le Clinical Dementia Rating (CDR) *sum of boxes*). L'article identifie que la performance de classification est améliorée lorsque les mesures de dépistage sont utilisées de pair avec des mesures cognitives.

La deuxième partie de l'article 1 s'est concentrée sur l'ensemble des algorithmes testés dans les études sélectionnées (n=101), pour les deux métriques les plus utilisées (précision ou *accuracy* et AUC). Nous présentons les contraintes à une analyse quantitative des algorithmes les plus performants, selon le type de modalité utilisée (c.à.d., EEG ou cognitive) pour les différentes classifications. Dix catégories d'algorithmes ont été examinés, RF (18,46 %) et SVM (15,38 %) étant les plus utilisés. De plus, la performance de 32,31% des algorithmes utilisés (soit 6 études) a été mesurée par la métrique de précision, à partir de classes débalancées (c.à.d., nombre de sujets par classes non-identique – dans un problème binaire). Nous avons ensuite présenté les résultats des études qui ont comparé les performances de différents algorithmes, à partir de la même cohorte et des mêmes variables. Nos résultats indiquent qu'aucun algorithme n'est systématiquement plus performant que les autres pour les différentes classifications. Le choix de l'algorithme le plus performant varie en fonction des caractéristiques spécifiques, du type de données et de la tâche de classification à accomplir.

## **Retour sur l'article 2**

Dans la deuxième étude, nous avons effectué un suivi moyen de 5.6 ans auprès de 156 individus avec un TCSPi, parmi lesquels 60 ont développé une maladie neurodégénérative : 34 une MP et 26 une DCL. Nous avons utilisé un modèle d'AM, le RF, ainsi qu'une méthodologie rigoureuse pour contrôler le déséquilibre de classes, la valeur significative (valeur P) et la généralisation hors-échantillon des résultats. Respectivement, avec du *bootstrapping*, une moyenne à travers plusieurs grains (ou *seeds*; l'aléatoire généré par l'ordinateur) et une validation croisée imbriquée. Une optimisation des variables a été effectuée par une sélection exhaustive des attributs, ainsi qu'une

optimisation des hyperparamètres du RF via une *GridSearch*. Trois catégories principales de classifications ont été déterminées : 1) DCL vs TCSPi (avec ou sans TCL), 2) DCL vs MP (avec ou sans TCL) et 3) DCL vs contrôles. Les données lors de l'évaluation initiale (temps de base) ont été utilisées pour prédire la condition clinique au suivi (toujours TCSPi, avec un TCL, évolution vers une MP ou évolution vers une DCL). Les variables utilisées appartenaient à trois catégories : 1) sociodémographique et clinique, 2) cognitive, et 3) autres marqueurs cliniques (psychiatrique, sommeil, sensoriel, moteur et autonome). Nous avons présenté les variables calculées comme ayant la plus grande « importance » pour chacune des classifications, c'est-à-dire que chacune d'entre elles a une contribution élevée à l'amélioration moyenne de la classification au sein d'une partition lorsque l'arbre est divisé en sélectionnant cette variable.

Les résultats de notre étude montrent que le modèle d'AM a permis de distinguer efficacement les patients avec un TCSPi qui ont d'abord développé une démence de ceux qui ont d'abord développé une MP (0.80 +/- 0.03,  $p=2.22e-16$ ), de ceux qui sont demeurés idiopathiques (0,78 +/- 0,05,  $p=2,22 e-16$ ), de ceux qui n'ont pas développé de troubles cognitifs (c'est-à-dire, MP sans TCL et TCSPi sans TCL ; 0.76 +/- 0.04,  $p=1.46e-11$  et 0.71 +/- 0.05,  $p=2.22e-16$ ) ou encore des sujets contrôles en santé (0.79 +/- 0.02,  $p=2.22e-16$ ). Notre étude montre le rôle crucial, dans les modèles d'AM, des marqueurs cognitifs des fonctions exécutives (Trail Making Test; TMT), de la mémoire (test d'apprentissage des 15 mots de Rey; RAVLT) et du TCL dans la distinction des différentes trajectoires pathologiques. Plus précisément, dans le modèle RF multivarié, les principaux prédicteurs du développement de la DCL au suivi chez les patients atteints d'un TCSPi sont les mesures de mémoire épisodique verbale (RAVLT), suivies des mesures d'attention et des fonctions exécutives (TMT). La variable catégorielle du TCL émerge comme étant une variable décisive, afin de distinguer les patients qui développeront une démence au suivi de ceux qui développeront d'abord une MP ou qui demeureront idiopathiques.

Dans les modèles RF à une seule variable, la performance de la variable du TCL seule était équivalente à son utilisation dans un modèle RF multivarié pour classifier l'évolution vers la démence, par rapport à celle vers la MP ou le maintien du statut idiopathique. De même, le RAVLT (mais pas le TMT) a montré une efficacité comparable lorsqu'il était utilisé seul, plutôt que dans

un modèle multivarié, pour distinguer les patients qui ont développé une démence des sujets contrôles, renforçant son potentiel pour une utilisation en clinique. Enfin, nos résultats montrent que les modèles multivariés RF surpassent les modèles univariés RF en termes de précision lorsqu'il s'agit de discerner les statuts cognitifs plus spécifiques, tels que l'absence de TCL au moment du suivi. Ils se distinguent également par la présence de plusieurs variables à forte importance issues d'une interaction synergique entre les modalités sociodémographiques, cliniques et cognitives.

Dans la prochaine section, les résultats de cette thèse seront discutés en termes d'implications cliniques et théoriques. Nous détaillerons ensuite les forces et limites des deux articles et présenterons une réflexion sur les avenues de recherche futures.

## **Contributions théoriques, méthodologiques et empiriques**

### **Avancées, manquements et potentiel des classifications investiguant l'évolution des synucléinopathies**

*Portrait des enjeux freinant la synthèse et l'extraction de tendances à travers les études en AM – divergences méthodologiques multiples et domaine émergent*

L'article 1 a dévoilé de nombreux défis dans l'extraction et l'analyse de tendances, dans les études en AM sur le déclin cognitif lié aux synucléinopathies et à leurs stades prodromaux. La difficulté principale est la variété des classifications binaires étudiées, ainsi que des modalités et des variables utilisées. Cela pourrait être expliqué par le nombre limité d'études que nous avons pu recenser, confirmant que les applications méthodologiques étudiées pour explorer ces conditions cliniques sont à leur balbutiement – de l'exploration et du développement sont nécessaires dans ce domaine émergent. À noter que la population atteinte de synucléinopathies est relativement rare, ce qui rend d'autant plus difficile la réalisation d'études à grande échelle, limitant également le nombre d'études multicentriques.

Alors que les données disponibles sont déjà limitées, nous avons également constaté que la diversité des métriques utilisées à travers les différentes études constitue un frein important pour la synthèse et l'interprétation des résultats. Le choix des métriques d'évaluation est un aspect crucial dans l'étude des performances d'un modèle d'AM (Kingsmore et al., 2021; Lever, 2016; Tholke et al., 2023). De nombreuses métriques existent et le choix dépend de divers facteurs, tels que la complexité du modèle ou le type de données à traiter (Hossin et Sulaiman, 2015; Tholke et al., 2023). Chaque métrique cible des aspects distincts et leurs valeurs ne sont pas directement comparables (Hossin et Sulaiman, 2015; Lever, 2016; Tholke et al., 2023). Par exemple, la justesse (ou *accuracy*) est une mesure qui correspond à la fraction des prédictions correctes, mais ne donne pas d'information sur les faux négatifs ou les faux positifs (Lever, 2016; Luque et al., 2019). La sensibilité quant à elle mesure la proportion de vrais positifs correctement identifiés, sans tenir compte des vrais négatifs et des faux positifs (Lever, 2016). Par conséquent, une étude qui ne présente qu'une seule métrique laisse en suspens des questionnements concernant la performance

du classificateur s'il était évalué selon d'autres métriques (Lever, 2016). Nous suggérons que les futures études présentent leurs matrices de confusion, afin de permettre l'extraction de différentes métriques de performance.

Nous avons recensé toutes les métriques utilisées dans les 33 études retenues dans la revue de littérature et avons observé que la métrique la plus fréquemment utilisée était la précision/justesse (*accuracy*), suivie de l'AUC (*Area Under the Curve*), puis de la sensibilité/spécificité. Une recherche menée dans notre laboratoire suggère que selon la métrique utilisée, l'équilibre ou le déséquilibre des classes d'une classification binaire affectera différemment la performance obtenue. Effectivement, dans le cas de données équilibrées où chaque classe comprend un nombre identique de sujets, l'AUC sera légèrement supérieure à la justesse (Tholke et al., 2023). La justesse tendra plutôt à surestimer la performance lorsque les données sont déséquilibrées, ce qui est fréquent dans les études sur les synucléinopathies en raison des défis liés au recrutement des patients. Par ailleurs, l'article 1 souligne que 52,72 % des études ont effectué des classifications binaires à partir de deux classes déséquilibrées (c.à.d., n'ayant pas le même nombre de sujets). Plus précisément, 20,69 % des études utilisent la métrique de justesse avec des données déséquilibrées. Cette asymétrie dans la distribution des classes est un facteur critique à prendre en compte, car cela peut introduire un biais dans la modélisation prédictive, compromettant potentiellement la généralisation des résultats.

Ainsi, le premier article de cette thèse met en évidence les défis méthodologiques rencontrés dans la synthèse des connaissances en AM dans l'étude de l'EEG et de la cognition dans les synucléinopathies et suggère différentes façons d'y pallier. En effet, nous suggérons que davantage d'études se penchent sur la question en utilisant une variété de métriques et en remédiant au débalancement des classes (ex : sous-échantillonnage aléatoire ou *bootstrapping*). Cela pourrait permettre à une future méta-analyse d'identifier les variables clés – soit celles qui sont retrouvées de manière récurrente comme significatives. Cette compréhension est cruciale pour élaborer des modèles de prédiction optimaux.

### *Intérêt plutôt diagnostique que prédictif de l'AM dans l'étude des synucléinopathies*

Les deux études de cette thèse mettent en évidence le fort potentiel des classifications utilisant l'AM issues d'études longitudinales pour prédire l'évolution des pertes cognitives des patients atteints d'une synucléinopathies (TCSPi ou MP). Toutefois, nous constatons dans l'article 1 un manque significatif de telles études dans la littérature scientifique actuelle. La plupart des recherches se concentrent sur la distinction entre MP vs contrôles, en raison de l'intérêt clinique d'un tel diagnostic et donc de cette classification (Mei et al., 2021; Rana et al., 2022). Les autres classifications sont moins étudiées, tel que démontré dans notre article 1. Notons par exemple, la distinction entre la MP avec TCL et la MP sans TCL. Pourtant, il s'agit d'un résultat significatif de notre étude 2, dans la classification longitudinale des patients TCSPi qui développent une DCL vs une MP sans TCL. Ainsi, l'application méthodologique à ce type de questions cliniques et neuroscientifiques n'en est qu'à ses balbutiements, voire à un stade embryonnaire pour certaines conditions. Nous soulignons l'intérêt d'explorer davantage ces classifications.

Il est important de mentionner que le déséquilibre dans la répartition des études par catégorie diagnostique entraîne une représentativité inégale, affectant la fiabilité de certaines catégories. Cela est entre autres dû à l'intérêt récent (bien que croissant) pour les méthodes d'AM dans l'étude des synucléinopathies. Par exemple, les catégories DCL versus Contrôles et TCSPi qui développeront une MP/DCL ne comptent que pour 3% des études. Dans l'optique d'enrichir les futures revues systématiques, nous suggérons d'élargir les critères d'inclusion afin d'inclure davantage d'études pertinentes. Cela permettrait d'optimiser l'exploitation des données disponibles en AM sur le déclin cognitif dans les synucléinopathies, tout en préservant la qualité de l'information. Nous suggérons d'inclure les états de sommeil en EEG, ainsi que d'autres méthodes de neuroimagerie et d'électrophysiologie, bien que cela puisse limiter les applications cliniques en raison de leur coût plus élevé et de leur accessibilité réduite. Il serait également bénéfique d'élargir les types de classifications considérées, afin de considérer la classification binaire entre la DCL et d'autres types de démences, telle que la maladie d'Alzheimer, autant dans un devis transversal que longitudinal. Enfin, il convient de noter que de nombreuses études qui satisfaisaient plusieurs critères d'inclusion (c.à.d., modalités, classifications et AM) ont dues être écartées en raison d'une

absence de spécification claire sur les méthodes de validation de modèle, limitant ainsi notre visée de généralisation clinique. Nous suggérons que la rigueur dans la sélection des études soit maintenue dans les revues systématiques futures, afin d'assurer la généralisation des résultats.

### *L'apport de tracer un continuum via notre approche globale*

Cette thèse se distingue en étant la première à tracer un continuum des marqueurs tout au long de l'évolution du TCSPi et de la MP. L'absence d'une telle approche globale dans la littérature existante en AM limite la capacité à comprendre la progression des synucléinopathies. En évaluant les performances et les variables significatives dans diverses classifications à travers le spectre évolutif de ces pathologies, nous offrons une perspective novatrice sur ce domaine. Ce point revêt une importance particulière étant donné la fréquence élevée et la variabilité du déclin cognitif dans les synucléinopathies. Par ailleurs, l'article 2 se penche spécifiquement sur la sélection des variables les plus pertinentes selon la classification recherchée. L'étude couvre un nombre important de conditions cliniques du « spectre » des synucléinopathies que les patients avec un TCSPi sont susceptibles de développer. Nous avons ainsi identifié diverses combinaisons de tests pouvant prédire la progression du TCSPi et avons découvert que les combinaisons de variables varient en fonction des pathologies cliniques posées : évolution vers la MP, vers une DCL, ou une progression sans développer comme tel une synucléinopathie clinique (donc qui demeure avec la nomenclature de TCSPi).

En somme, notre recherche souligne l'importance de mener davantage d'études longitudinales en utilisant les outils d'AM et d'élargir le spectre des évolutions cliniques considérées, pour une compréhension et une prise en charge optimale de ces troubles neurodégénératifs.

### *Construire un pont entre les statistiques classiques et l'AM*

La frontière entre les statistiques classiques et l'AM est sujette à débat (Bzdok, 2017). Les statistiques classiques offrent une compréhension mécanistique, généralement via des analyses de régression et des tests d'hypothèse nulle (ANOVA, tests T, etc.) pour isoler les variables

individuellement importantes. L'AM, quant à lui, excelle dans la modélisation qui vise une prédiction précise, identifiant des ensembles de variables pour prédire précisément des résultats, souvent par le biais d'algorithmes de type "boîte noire" (Abu-Mostafa et al., 2012; Bzdok, 2017; Bzdok et Meyer-Lindenberg, 2018). Ces approches se complètent : des variables importantes pour les statistiques classiques ne peuvent parfois pas optimiser les performances prédictives en AM, et vice-versa. Étant donné le contexte clinique dans lequel s'inscrit cette thèse, nous avons tenté d'investiguer cette complémentarité avec rigueur, en recherchant les variables permettant une modélisation prédictive précise, indépendamment de leur lien causal avec le résultat. Cette approche ne minimise pas l'importance de comprendre les mécanismes pathophysiologiques de ces troubles neurodégénératifs, mais la complète.

*Multimodalité et mesures notables en AM à ce jour – dans la classification des conditions du continuum des synucléinopathies*

En premier lieu, l'article 1 montre que dans la littérature scientifique en AM, différentes modalités sont utilisées de façon prédominante, selon la classification d'intérêt. Les synucléinopathies ont principalement été étudiées à partir de variables électroencéphalogramme (EEG) pour les classifications impliquant un groupe témoin (avec la MP ou la DCL) ou l'évolution du TCSPi vers la MP/DCL, alors que les tests cognitifs, de pair avec des variables cliniques et sociodémographiques sont souvent étudiés pour la classification de la MP avec une cognition normale vs la MP avec des troubles cognitifs.

Dans l'article 1, nous montrons qu'au sein d'un modèle multivarié, certaines modalités se distinguent d'autres, selon la classification étudiée. En effet, l'article 1 s'est penché sur les études en AM qui ont comparé l'efficacité de différentes modalités, à partir du même modèle algorithmique et de la même cohorte. Il s'agit surtout d'études se penchant sur la classification du statut cognitif de patients ayant la MP, ou encore de la classification DCL vs Contrôles. L'article 1 montre que les variables d'EEG se distinguent souvent comme étant les plus importantes pour diverses classifications multimodales qui intègrent également des mesures cognitives, cliniques, sociodémographiques ou parfois même d'autres méthodes neuroimagerie (ex : IRM, PET scan).



Les caractéristiques EEG, telles que la puissance médiane relative globale à travers les bandes de fréquences, la fréquence moyenne postérieure (MP → déclin cognitif), ainsi que le rapport de l'activité theta/alpha (DCL vs Contrôles), contribuent d'une façon plus importante à la performance des modèles d'AM et dépassent la performance d'autres modalités lorsqu'utilisées séparément.

L'article 1 nous permet également de démontrer qu'au sein même des mesures EEG, certaines se démarquent dans l'optimisation de la performance de prédiction. Prenons les classifications en AM qui visent à distinguer la MP de sujets contrôles, soit la condition la plus étudiée dans la littérature recensée. Notre article 1 démontre que plusieurs variables présentent des résultats prometteurs pour la classification de la MP – allant des caractéristiques spectrales dans différentes régions cérébrales (bande thêta principalement en temporal et frontal, de manière bilatérale), à la connectivité inter-fréquences (modulation des hautes fréquences par les basses fréquences dans les régions frontales et centrales), jusqu'aux mesures de complexité (irrégularité et imprévisibilité du signal). L'importance des mesures de complexité souligne le rôle crucial de la dynamique complexe et non linéaire de l'activité cérébrale au repos dans la MP. Ce constat est renforcé par les études spectrales qui ont identifié plusieurs caractéristiques de puissance spectrale à travers différentes régions cérébrales. Ainsi, nous démontrons qu'une variété de mesures cérébrales sont importantes, ce qui nous amène à souligner l'intérêt de certains modèles d'AM, spécifiquement les réseaux de neurones, qui ont permis dans de nombreuses études de l'article 1 de déterminer de manière autonome les caractéristiques les plus importantes du signal EEG, permettant d'en exploiter sa richesse. Effectivement, ces algorithmes utilisent le signal EEG brut, sans nécessiter de pré-sélection manuelle par le chercheur. Au niveau clinique, la capacité à distinguer de façon significative les sujets contrôles des patients atteints de la MP à partir du signal EEG au repos souligne que cette maladie fait partie d'un spectre neurodégénératif. Cela corrobore l'hypothèse de Braak et al. (Braak et al., 2003) selon laquelle une neurodégénérescence importante affecte déjà le cerveau au moment du diagnostic clinique.

L'article 1 nous permet également de confirmer l'importance en AM de certains biomarqueurs identifiés dans des études précédentes en statistiques classiques (SC), tel que le ralentissement de l'EEG. Effectivement, ce phénomène a été identifié comme caractérisant le

déclin cognitif dans la MP et est observé dans la DCL dans les études précédentes (rythme alpha ralenti dans les régions occipitales (Bonanni et al., 2016; Bonanni et al., 2015; Bonanni et al., 2008) et temporales (Briel et al., 1999; Stylianou et al., 2018)). Notre article révèle que le ratio theta/alpha permet de classifier la DCL de sujets sains, à partir de l'AM, corroborant les trouvailles précédentes. Par ailleurs, ce ralentissement est également retrouvé dans la classification en AM de la MP de sujets sains (ratio alpha1/theta; (Chaturvedi et al., 2017)), tout comme dans les études en SC étudiant la sévérité des atteintes cognitives dans la MP (ratio ondes rapides/lentes (Eichelberger et al., 2017; Guner et al., 2017; Kamei et al., 2010; Latreille et al., 2016; Morita et al., 2011) et fréquence dominante supérieure (Caviness et al., 2007; Latreille et al., 2016; Soikkeli et al., 1991; Stanzione et al., 1996)). Quant au TCSPi, le ralentissement de l'EEG à l'éveil (central et occipital droit) a été également associé en SC au TCL (Iranzo et al., 2010) et au développement futur de la DCL (Rodrigues Brazete et al., 2016). Malgré ces corrélations connues, notre revue met en lumière une absence d'études examinant en AM le ralentissement de l'EEG en tant que prédicteur de la démence dans le TCSPi, bien que cette condition soit un facteur de risque majeur de la DCL. Cela pourrait être expliqué par l'utilisation récente des outils d'AM dans l'étude des synucléinopathies. L'article 1 suggère donc qu'il serait d'un grand intérêt d'explorer d'avantage ce ralentissement à partir d'outils en AM dans l'investigation de l'ensemble du spectre des synucléinopathies.

L'article 1 dévoile d'autres mesures EEG d'intérêt, soit certaines mesures de connectivité cérébrale - les mesures de complexité. Nous suggérons qu'elles soient d'avantage investiguées dans les études en AM, étant donné qu'elles ressortent comme cruciales dans les modèles d'AM. Par exemple, l'Index de Déphasage, une mesure de connectivité EEG, présente une efficacité supérieure pour distinguer la MP avec TCL de la MP sans TCL lorsqu'utilisé seule, plutôt qu'avec des mesures spectrales (c.à.d., puissance des bandes de fréquence) (Chaturvedi et al., 2019). Cela montre que les mesures de connectivité ont une importance élevée dans les modèles d'AM et que l'ajout de mesures spectrales peut y introduire du bruit. Nous montrons que, contrairement aux mesures de complexité, les mesures de cohérence n'offrent pas de précision de classification significative pour la classification de la DCL vs contrôles, même lorsqu'elles sont combinées avec des variables spectrales (Dauwan et al., 2016). Ainsi, au sein des mesures de connectivité, certaines semblent plus pertinentes que d'autres selon la classification (Cesari et al., 2021). Il est donc

essentiel d'approfondir les recherches en AM portant sur les interactions entre les différentes régions cérébrales, car les mesures de connectivité semblent détenir des informations supplémentaires précieuses pour comprendre la progression de la maladie.

Enfin, dans l'article 1, nous montrons l'utilité et la pertinence de combiner diverses modalités – puisque chacune capture des informations différentes, qui en synergie peuvent contribuer à une meilleure classification des sous-types de synucléinopathies (TCSPi vs MP vs DCL) de même qu'à une meilleure prédiction des individus les plus susceptibles de convertir. Ce constat a pu être fait à travers les études cognitives, qui sont principalement multimodales, intégrant une variété de variables, telles que des données cliniques et démographiques. Par exemple, les variables optimales pour différencier la MP avec et sans démence dans les modèles d'AM étudiés englobent un mélange d'évaluations cognitives, d'évaluations de la fonction motrice, ainsi qu'un ensemble complet de variables démographiques, cliniques, de style de vie et fonctionnelles. Notre étude montre que cette combinaison de variables diverses semble importante pour une classification précise, indiquant que le modèle bénéficie significativement d'un spectre varié d'informations. Plus précisément, notre revue indique que les études utilisant des mesures cognitives pour étudier le statut ou le déclin cognitif dans la MP démontrent une précision de classification améliorée par l'inclusion de données démographiques et cliniques aux données cognitives. À noter que les mesures EEG ne bénéficient pas de l'ajout des mesures démographiques (MP → TCL et MP → DCL). Bien que les données cliniques et sociodémographiques améliorent la performance de classification, au sein de ces modèles multivariés, les mesures cognitives sont systématiquement identifiées comme étant les plus importantes dans l'optimisation de la prédiction du modèle d'AM. Ces mesures cognitives sont variées – notamment tests pour les fonctions exécutives, la mémoire de travail (dans l'évaluation de la progression vers un déclin cognitif ou de la démence chez dans la MP), ainsi que des tâches de mémoire et verbales (MP avec et sans démence). Par ailleurs, notre revue dévoile l'intérêt d'inclure des tests de dépistage plus larges, tels que le MoCA. Plus précisément, nous montrons que les tests de dépistage améliorent la performance de classification lorsqu'ils sont combinés avec des mesures cognitives spécifiques (MP vs MP+ démence (Lee et al., 2016) ou MP+TCL (Byeon, 2020b); fluence verbale avec questionnaire 5 items PDD-5S et MMSE avec le *Sum of boxes*, respectivement). Pour mieux comprendre quelles fonctions cognitives sont les plus importantes dans les modèles de prédiction

d'AM du déclin cognitif dans la MP et dans le TCSPi, nous suggérons une évaluation approfondie de l'ensemble des domaines cognitifs dans les expérimentations futures, ainsi qu'une plus grande investigation de la combinaison optimale de ces divers marqueurs.

Nous suggérons que la modalité la plus déterminante diverge selon la classification d'intérêt et que les études multimodales pourraient permettre de clarifier l'apport de chaque modalité (et de mesures au sein de ces modalités), lorsqu'utilisées conjointement. En bref, l'article 1 propose une vision globale des performances de diverses variables à travers les modalités et classifications, et fournit une base solide pour les futures études en multimodalité. L'intégration de ces différentes modalités pourrait non seulement aider à affiner la classification des synucléinopathies, mais aussi à améliorer notre compréhension de la progression de ces maladies neurodégénératives.

#### *Regarder de plus près les algorithmes – sous un angle quantitatif*

L'article 1 montre une diversité d'algorithmes dans la recherche en synucléinopathies. Nous trouvons qu'une dizaine de catégories d'algorithmes sont utilisées, mais notons une utilisation particulièrement élevée des algorithmes d'apprentissage supervisé, plus spécifiquement les RF, suivis des SVM, possiblement en raison de leur efficacité dans le traitement des données de grandes dimensions, ainsi que leur plus grande résistance au surapprentissage.

Tous les algorithmes examinés présentent des performances significatives et la majorité d'entre eux affiche des performances de satisfaisantes (>80%) à excellentes (>95%). Toutefois, il est crucial de prendre en compte l'influence potentielle d'un biais de publication lors de l'interprétation de ces résultats. Ce phénomène se produit lorsque les études produisant des performances significatives ou élevées sont publiées de façon préférentielle, alors que celles dont la performance n'est pas significative ou faible ne sont pas publiées, les rendant donc sous-représentées dans la littérature.

Ensuite, l'article 1 souligne à nouveau le manque d'études dans le domaine, mais surtout les contraintes que cela représente dans les tentatives de comparaison directe des performances

algorithmiques entre les études. D'abord, lorsque les études sont séparées en fonction de nos six types de classification, des algorithmes et des métriques de performance, la taille de l'échantillon devient insuffisante (moyenne  $n=3$ ), compromettant la possibilité d'effectuer une analyse statistique qui permettrait de se prononcer sur les algorithmes les plus performants. De plus, nous n'avons pas accès aux données de la majorité des études, ce qui est souvent le cas dans la recherche clinique. Finalement, il n'est pas possible de comparer directement les performances des différents algorithmes, car des facteurs tels que la complexité du problème, la qualité des données, la taille de l'échantillon et l'optimisation des hyperparamètres peuvent varier de manière significative, ce qui rend les comparaisons directes non-interprétable, en raison du grand nombre de variables confondantes. De plus, malgré que les études des catégories "EEG" et "cognition" ont effectivement utilisé ces modalités, les variables examinées sont très diverses (par exemple, EEG : connectivité, puissance, rapports ; cognition : tests de dépistage, combinaisons avec des variables sociodémographiques ou exclusivement des tests cognitifs). Cela complique davantage les comparaisons directes, ce qui renforce notre propos selon lequel une comparaison quantitative ou statistique des performances entre les études serait erronée.

Nous avons donc plutôt identifié un sous-ensemble d'études qui ont effectué des comparaisons intra-études de plusieurs algorithmes d'AM pour une même classification, en utilisant les mêmes cohortes et une même métrique de performance. Cette approche permet l'évaluation et la comparaison des performances algorithmiques, car elle garantit l'uniformité des données et techniques de mesure de performances utilisés pour les différents algorithmes qui sont comparés. Les résultats de l'article 1 suggèrent qu'aucun algorithme n'est systématiquement plus performant que les autres pour chacune des classifications. Ainsi, notre article 1 démontre que le choix d'un algorithme devrait être effectué à partir des particularités des données en main et de la question clinique posée. La diversité des performances algorithmiques d'une étude à l'autre souligne la nécessité d'approches personnalisées d'analyse des données dans la recherche sur la MP. Les études futures pourraient bénéficier de l'exploration d'un éventail plus large d'algorithmes, y compris ceux sous-représentés dans la littérature actuelle, afin d'établir des bases plus complètes quant à la performance des algorithmes dans les tâches de classification des synucléinopathies.

## **Avancée dans l'étude de la démence liée au TCSPi : première exploration de l'AM pour une prédiction précise et multivariée basée sur les données cognitives**

*Vers une optimisation de l'AM dans la prédiction de la démence liée au TCSPi : considérer les enjeux méthodologiques et exploiter la force des données cognitives*

À notre connaissance, l'article 2 est la première étude longitudinale à utiliser l'AM pour optimiser la prédiction de la conversion vers la démence chez des individus avec un TCSPi, principalement à partir de données cognitives. L'AM nous permet d'aller au-delà de l'analyse statistique univariée traditionnelle, en montrant comment une variable peut non seulement apporter une meilleure compréhension mécaniste lorsqu'elle est utilisée individuellement, mais parfois également optimiser la performance de prédiction lorsqu'utilisée dans un algorithme d'AM dans un contexte multivarié, en interaction avec d'autres variables.

Dans le cadre de l'article 2, nous avons procédé à une étude rigoureuse du potentiel de prédiction de l'AM ainsi que de ses capacités de généralisation, à travers diverses méthodes d'optimisation, tout en considérant et en réduisant plusieurs biais. Tout d'abord, l'un des principaux défis de la recherche sur des populations cliniques réside dans la gestion des données manquantes, qui peut survenir pour diverses raisons, y compris l'absence de suivi. Grâce aux avantages de l'AM, qui permet entre autres d'utiliser conjointement des variables continues avec des variables catégorielles, nous avons pu en partie surmonter ce biais et améliorer la performance de nos modèles. Par ailleurs, nous n'avons pas fait d'imputation moyenne de données pour les données manquantes, afin de ne pas biaiser le modèle d'AM en accentuant des tendances existantes (White et al., 2010). Puis, afin de réduire la multicollinéarité, nous avons exclu les variables présentant un coefficient de corrélation supérieur à 0,85 pour chacune des classifications binaires. Ensuite, nous avons pris en compte une autre contrainte souvent présente dans l'étude de pathologies (ex : en raison de l'imprévisibilité des trajectoires évolutives), celle des classes non équilibrées dans un problème de classification binaire, par le biais du sous-échantillonnage aléatoire (ou *bootstrapping*), qui nous a permis de sélectionner un nombre égal de sujets de chaque classe lors de la classification. Malgré un coût en temps de calcul accru, nous avons effectué une sélection exhaustive des

variables, ce qui nous a permis d'identifier la meilleure combinaison de variables, éliminant celles qui étaient redondantes ou non pertinentes. La réduction de dimensionnalité obtenue dans l'ensemble de nos classifications nous permet d'indiquer que plusieurs variables introduisent du bruit dans l'effort d'optimiser la prédiction des modèles d'AM. Par ailleurs, nous avons mis en évidence les variables les plus « importantes ». Chaque variable a été évaluée en fonction de sa contribution à l'amélioration moyenne de la classification lorsqu'une partition de l'arbre est divisé en sélectionnant cette variable, et ce, pour plusieurs grains (ou *seeds*; valeurs initiales pour la génération de nombres aléatoires). Cela fut possible puisque malgré que le modèle RF présente un mode de fonctionnement qualifié de « boîte noire », il fait partie des algorithmes d'AM qui permettent le meilleur équilibre entre la complexité et la compréhension de ce qui sous-tend les décisions de l'algorithme, principalement grâce à ce calcul de l'importance des variables (Bzdok et Ioannidis, 2019). Enfin, afin de maximiser le potentiel du modèle basé sur le RF, nous avons optimisé ses hyperparamètres – dans le cas des modèles multivariés, mais également univariés. En dernier lieu, afin de s'assurer que la performance du modèle n'est pas le résultat de surapprentissage (ou *overfitting*) et peut se généraliser à de nouvelles données, nous avons procédé à une validation croisée stratifiée à cinq plis lors de l'évaluation du modèle d'AM à travers les différentes étapes d'optimisation des hyperparamètres, ainsi que des variables. Ce souci quant au surapprentissage transparait également dans l'article 1, où seules les études utilisant des techniques de validation de modèle ont été sélectionnées. La plupart des études de notre revue exhaustive de la littérature ont eu recours à une méthode de validation croisée à 10 plis pour évaluer le modèle entraîné sur un ensemble de test.

En abordant directement les enjeux méthodologiques et en exploitant habilement les données cognitives, nous avons non seulement mis en lumière le potentiel d'une approche en AM, mais aussi suggéré comment les défis liés à la gestion des données manquantes et à l'équilibre des classes peuvent être atténués. Enfin, cette étude ouvre la voie à l'utilisation plus systématique de l'AM dans la recherche sur les synucléinopathies. En proposant des pratiques explicitement décrites, nous espérons favoriser leur adoption par les cliniciens et chercheurs du domaine, et ainsi contribuer à des avancées significatives dans la prédiction du déclin cognitif dans le TCSPi.

### *Cibler la détection précoce et spécifique de la DCL dans le TCSPi*

Les études antérieures ont identifié plusieurs marqueurs potentiels de la phénoconversion vers une synucléinopathie dans le TCSPi (Miglis et al., 2021). Cependant, peu de marqueurs n'atteignent un ensemble combiné de facteurs tels que de précision, d'exactitude, de disponibilité et de rentabilité, permettant d'en faire un marqueur idéal. Par ailleurs, le moment d'apparition d'une anomalie dans ces marqueurs, ainsi que dans l'évolution de la pathologie varie grandement, certains se manifestant tôt et évoluant lentement (ex : l'hyposmie et la dysautonomie), tandis que d'autres apparaissent plus près du moment de la phénoconversion (ex : atteinte motrice et cognitive). De plus, bien que les modèles pathophysiologiques ont amélioré notre compréhension de la progression des synucléinopathies, ils ne permettent pas à ce jour d'expliquer les différentes trajectoires évolutives (ex : MP, DCL, ou idiopathique au suivi), étant donnée cette variabilité importante des stades prodromes de la DCL et de la MP, aussi bien au niveau de l'apparition et de la progression des symptômes décrite ci-haut, mais également, au niveau de la propagation pathologique, ainsi que des marqueurs d'imagerie, dues à une variété de mécanismes pathophysiologiques (Berg et al., 2021). Notre thèse permet de suggérer que l'AM, via les capacités de « boîte noire », pourrait contribuer à capturer la complexité des facteurs sous-jacents menant aux présentations prodromales vers les différentes trajectoires évolutives. En effet, dans l'article 2, nous obtenons des performances significatives ( $p < 0.05$ ) dans la classification de la progression de vers la DCL ( $0.79 \pm 0.02$ ), la MP ( $0.80 \pm 0.03$ ) ou le maintien du statut idiopathique ( $0.78 \pm 0.05$ ).

Dans l'article 2, nous avons ciblé l'identification de la combinaison de variables cognitives, cliniques et sociodémographiques permettant d'optimiser l'algorithme RF en AM, et avons pu prédire avec précision la conversion vers une démence jusqu'à 5 ans avant la conversion. Nous avons inclus dans le modèle d'AM de nombreuses variables qui ont été individuellement liées au développement des synucléinopathies – telles que celles identifiées par Postuma et al. (2019) ainsi que Fereshtehnejad et al. (2019). (Postuma et al., 2019) ont regroupé les données de 24 centres (pour un total 1280 patients avec un TCSPi) et ont montré que plusieurs marqueurs utilisés dans l'article 2 sont liés à une augmentation du risque de phénoconversion vers une synucléinopathie – soit les déficits de l'olfaction, des dysfonctions érectiles, une perte d'atonie musculaire dans le sommeil REM, le TCL, l'âge, ainsi que des symptômes moteurs. L'évolution des manifestations



précoces des synucléinopathies dans le TCSPi jusqu'au diagnostic clinique a été décrite par (Fereshtehnejad et al., 2019). Ils ont constaté que la perte de l'olfaction et la baisse de la pression sanguine orthostatique apparaissent plus de 20 ans avant le diagnostic de synucléinopathie. Par la suite, les troubles de la discrimination des couleurs, la constipation et la dysfonction érectile apparaissent 10 à 16 ans avant le diagnostic, suivis d'une légère dysfonction urinaire et d'un déclin cognitif subtil, tel que mesuré par le MoCA, apparaissant 7 à 9 ans avant le diagnostic. Nous confirmons l'importance des variables cognitives en AM, par rapport aux mesures cliniques et sociodémographiques, 5 ans avant le diagnostic de démence. Ces mesures cognitives seront précisées de façon intégrative dans les deux prochaines sections. Cette thèse suggère que certaines des variables précédemment identifiées en SC sont cruciales pour les prédictions en AM, lorsqu'utilisées au sein d'une combinaison de marqueurs qui sont en interaction. Par ailleurs, les variables importantes varient selon la classification en question dans le continuum des synucléinopathies.

Ainsi, la présente thèse apporte une contribution innovante en combinant les variables précédemment identifiées pour se concentrer spécifiquement à la distinction de la conversion à la DCL (plutôt qu'aux 'synucléinopathies') par rapport aux autres trajectoires évolutives.

*Le TCL: une variable importante en AM dans la détection précoce des différentes trajectoires évolutives des synucléinopathies*

Le TCL est fréquent dans le TCSPi, étant confirmé chez le tiers des patients (Gagnon et al., 2009; Zhang et al., 2016), ce que nous confirmons dans notre étude avec une proportion de 38%. Toutefois, bien qu'il s'agisse d'un des facteurs de risque bien connus de la démence (Genier Marchand et al., 2017; Miglis et al., 2021), tous les patients n'évoluent pas dans cette direction (Fischer et al., 2007; Ganguli et al., 2004; Gauthier et al., 2006; Joza, 2023). Effectivement, il est rapporté que certains patients avec un TCL à corps de Lewy ou un autre sous-type de TCL peuvent retrouver une cognition normale avec le temps, bien qu'en restant à risque de démence (McKeith et al., 2020).

L'article 2 montre que le TCL est la variable calculée par le RF comme étant la plus importante dans un modèle multivarié pour distinguer deux types de progressions dans le TCSPi : vers 1) la démence, par rapport à 2) la MP ou au maintien du statut idiopathique. De plus, la performance de précision pour ces classifications demeure élevée lorsque le TCL est utilisé seul dans un algorithme de RF univarié, soulignant son potentiel clinique. Par ailleurs, nous notons une fréquence supérieure du TCL au temps de base chez les patients avec un TCSPi qui développeront une démence (88%) comparativement à ceux qui développeront une MP (26%) ou demeureront idiopathiques (31%). Ainsi, cette thèse renforce la valeur prédictive du TCL, tout en précisant qu'en AM, le TCL serait une variable pivot permettant de distinguer spécifiquement l'évolution vers la démence. Nous confirmons également les résultats précédents selon lesquels les patients avec un TCSPi qui développeront une MP seraient similaires à ceux qui demeureront idiopathiques, en ce qui concerne leur performance aux tests cognitifs, ainsi qu'à la fréquence du diagnostic de TCL (Genier Marchand et al., 2017). Nos résultats concordent également avec les études indiquant une association entre le TCL et la probabilité plus élevée de phénoconversion vers la démence (Genier Marchand et al., 2017; Genier Marchand et al., 2018; Joza, 2023; Postuma et al., 2019).

La distinction entre l'évolution vers la DCL vs la MP a également été caractérisée dans la littérature scientifique par des déficits attentionnels, exécutifs et visuospatiaux (Genier Marchand et al., 2017). Les variables cognitives ont aussi été trouvées comme étant plus efficaces pour distinguer ces deux groupes que les variables autonomiques (Postuma et al., 2019). Cette thèse offre une perspective nuancée sur leur utilisation et leur apport dans un contexte d'optimisation de prédiction en AM. Nous montrons que ces variables perdent de leur importance prédictive en AM dans la distinction de l'évolution vers la démence ou la MP, lorsque la variable du TCL est introduite, mais également lorsqu'elle est retirée. Effectivement, lorsqu'elle est retirée du modèle d'AM, la performance est considérablement réduite (16-17% en moins) et il s'agit plutôt des variables d'âge et motrices qui ressortent comme étant les plus importantes, bien qu'en interaction avec des variables de mémoire, d'attention et de fonctions exécutives, mais avec une importance significativement moindre dans le modèle RF. Ces changements mettent en évidence une idée intéressante : les variables cognitives, bien qu'essentielles, peuvent ne pas agir de manière isolée. L'interaction entre les facteurs cognitifs, moteurs et liés à l'âge semble plus complexe dans le contexte d'AM que dans les analyses statistiques traditionnelles.

Quant aux distinctions entre les sous-types de TCL associés au développement d'une DCL dans le TCSPi, la littérature scientifique indique des associations avec les deux sous-types, soit amnésique et non-amnésique (Genier Marchand et al., 2017; Postuma et al., 2019; Terzaghi et al., 2019). Certaines études suggèrent que le TCL non-amnésique à domaines multiples pourrait être un indicateur plus prédictif de la progression vers la DCL (Jicha et al., 2010; Kondo et al., 2016). Une distinction est établie entre le TCL associé à la maladie d'Alzheimer et le TCL associé aux maladies à corps de Lewy (TCL-LB). Ce dernier serait souvent non-amnésique ou amnésique à domaines multiples (Armstrong, 2021; McKeith et al., 2020). Il est suggéré que pour tout sous-type de TCL, son association à un TCSPi pourrait indiquer une pathologie à corps de Lewy sous-jacente (McKeith et al., 2020). Enfin, des travaux récents ont montré que les patients avec un TCSPi et un TCL concomitant ont des altérations cérébrales plus importantes que celles des patients sans TCL (Miglis et al., 2021; Rahayel et al., 2018; Rahayel et al., 2021; Remillard-Pelchat et al., 2022). Le TCL chez les patients avec un TCSPi est également lié à d'autres modifications cérébrales plus sévères, notamment une perfusion sanguine réduite et une activité EEG plus lente (Rodrigues Brazete et al., 2013; Vendette et al., 2012).

*La contribution des marqueurs cognitifs d'apprentissage – pour distinguer la démence de l'absence d'atteintes cognitives au suivi*

Via l'algorithme RF, nous avons identifié dans l'étude 2 deux prédicteurs cruciaux dans la distinction de la trajectoire vers la démence, de trois trajectoires sans atteintes cognitives : 1) les sujets contrôles, 2) les patients qui ont développé une MP sans TCL, ainsi que 3) les patients toujours idiopathiques sans TCL. Ces prédicteurs sont les mesures d'apprentissage (RAVLT 1-5; somme des essais 1 à 5 du RAVLT) et des fonctions exécutives (TMT B; partie B du TMT). Les études longitudinales qui se sont intéressées aux changements cognitifs dans le TCSPi avant la transition vers la DCL rapportent différents déficits cognitifs à mesure que le patient se rapproche du développement de la DCL. L'attention et les fonctions exécutives (TMT B) étaient altérées six ans avant le diagnostic de DCL, alors que des déficits significatifs au niveau de la mémoire épisodique (RAVLT) et de la fluence verbale sémantique étaient identifiés environ deux ans avant la conversion (Genier Marchand et al., 2018). Cette thèse permet de dévoiler une atteinte précoce

de la mémoire et des fonctions exécutives 5 ans avant la conversion et montre que ces variables contribuent non seulement à une meilleure compréhension mécanistique du problème lorsqu'elles sont examinées individuellement en statistiques classiques, mais optimisent également la précision prédictive lorsqu'elles sont utilisées ensemble dans un contexte multivarié d'AM.

Plus précisément, dans la prédiction de la DCL (DCL versus sujets contrôles), la mesure de mémoire (RAVLT 1-5) se démarque comme étant la plus notable (c.à.d., haut score d'importance et souvent sélectionnées) du modèle RF multivarié, suivi d'une mesure des fonctions exécutives (TMT B), puis d'une mesure attentionnelle (TMT A; partie A du TMT) et de mesures sociodémographiques et cliniques présentant une importance moindre. Ces mesures distinguent également les patients qui ont développé une démence des deux autres classifications mentionnées précédemment. À noter que le fait que les mesures cognitives se démarquent au sein de ce modèle multimodal d'AM est en accord avec les conclusions de l'article 1. Ainsi, cette thèse souligne que l'association des variables de mémoire et des fonctions exécutives augmente la précision de la prédiction quant à l'évolution vers une DCL chez un patient ayant un TCSPi.

En ce qui concerne le potentiel clinique individuel de ces mesures d'apprentissage, l'article 2 dévoile l'importance du RAVLT 1-5 en AM, mais pas du TMT B. Effectivement, lorsqu'utilisé seul dans un modèle d'AM, le RAVLT 1-5 présente des performances similaires au modèle multivarié, ce qui n'est pas le cas pour le TMT B. En effet, le TMT B est moins performant lorsqu'utilisé seul en AM, bien que les études précédentes aient montré qu'il s'agit de la première mesure à s'altérer significativement chez les patients qui développeront la DCL. Cette thèse suggère que 5 ans avant le développement de la démence, le TMT B est plus efficace dans l'optimisation de la prédiction en AM lorsqu'il est utilisé en combinaison à d'autres variables cognitives, principalement au RAVLT 1-5.

## *Nouvelle perspective sur les trajectoires évolutives des synucléinopathies*

Le modèle d'AM identifié dans l'étude 2 apporte un éclairage nouveau sur le pronostic des trajectoires évolutives dans les synucléinopathies. Ce modèle permet de prédire le type de synucléinopathie que les patients sont susceptibles de développer 5 ans avant leur conversion.

Ainsi, cette thèse permet de démontrer l'apport de l'AM et des données cognitives précédemment étudiées (provenant de batteries neuropsychologiques), dans la classification significative de presque l'ensemble des conditions du continuum des synucléinopathies par rapport à la démence : principalement l'apport du TCL pour distinguer de la trajectoire vers la MP ou qui demeureront idiopathiques et du RAVLT 1-5 pour distinguer des sujets contrôles. La seule distinction des phénotypes prodromiques qui fut non-significative, fut celle de la démence vs de la MP avec un TCL. À noter que la distinction entre les patients atteints de TCSPi qui ont développé une MP avec une cognition normale de ceux avec un TCL a rarement été conduite dans les études longitudinales sur la conversion du TCSPi - nous avons comblé ce manquement en étudiant la présence de la condition du TCL dans le TCSPi et dans la MP au moment du suivi, ce qui constitue une première dans l'application de l'AM à de telles données. L'étude 2 montre que l'efficacité du RF est non significative, lorsque nous tentons de classer les patients avec un TCSPi qui développeront une MP accompagnée d'un TCL de ceux qui développeront une DCL. Cela pourrait suggérer qu'ils partagent certaines similarités neuropathologiques, par exemple une trajectoire de neurodégénérescence similaire.

Quant aux classifications significatives, mais moins performantes, on retrouve celle de conditions cognitives plus précises et plus éloignées de la démence au moment du suivi – soit pour classer l'absence de TCL dans le TCSPi de la présence de DCL. En effet, dans ce modèle de classification multivarié, aucune variable ne se démarque particulièrement des autres. Nous montrons qu'une combinaison de variables issues des trois modalités (c.à.d., cognitives - mémoire et fonctions exécutives; sociodémographiques - âge; et cliniques – motricité et olfaction), avec une importance équivalente est nécessaire et issue d'une interaction synergique. Ainsi, l'article 2 montre que plus la condition s'éloigne de la DCL au suivi, plus un modèle complexe d'AM est nécessaire, confirmant la force d'une approche multimodale et multivariée. Par ailleurs, bien

qu'avec une importance bien moindre que les variables cognitives, lorsqu'elles sont disponibles, les mesures sociodémographiques et cliniques sont toujours sélectionnées comme optimisant le modèle d'AM pour l'ensemble des classifications multivariées dans le continuum des synucléinopathies, suggérant qu'elles ont une valeur ajoutée – confirmant ce constat de l'article 1.

Cette thèse confirme la valeur de l'AM pour différencier les trajectoires de synucléinopathie, et ce, spécialement dans un contexte de chevauchement significatif des évolutions des premiers stades neuropathologiques des synucléinopathies.

## **Implications, limites et directions futures**

### **Implications**

Cette thèse crée un cadre de réflexion et trace la voie aux avancées majeures dans le débat sur l'utilisation de l'AM pour prédire le déclin cognitif dans les synucléinopathies. Nous nous y penchons à travers un souci de rigueur méthodologique et de généralisation. C'est l'accès à l'ensemble de la littérature scientifique et à l'une des plus grandes cohortes à l'échelle internationale de patients avec un TCSPi du Centre d'Études Avancées en Médecine du Sommeil (CÉAMS) de l'Hôpital du Sacré-Cœur de Montréal qui a rendu cet effort possible. Au cours des dernières années, l'AM a été propulsé en neurosciences par l'augmentation de la puissance de calcul et l'ère des données massives, ouvrant une multitude de possibilités pour l'avancement des connaissances et pour des applications cliniques convoitées par les cliniciens (Magoulas, 1999). Cependant, l'AM est comparable à une épée à double tranchant : d'un côté, elle offre un éclairage inégalé sur la compréhension et les décisions cliniques, en révélant des prédictions basées sur des interactions complexes entre les données, mais d'un autre côté, son utilisation requiert une rigueur importante pour tirer pleinement parti de son potentiel. Cela s'applique non seulement dans le cadre d'une étude spécifique, mais aussi en tant que pièce contributive au « puzzle global » des connaissances en AM sur une question de recherche spécifique, soit ici la prédiction du déclin cognitif dans les synucléinopathies.

Comme l'AM est un outil de plus en plus utilisé par les cliniciens, nous considérons qu'il est important et d'actualité de faire le point de son utilisation dans la prédiction de la démence dans les synucléinopathies. C'est l'objectif de l'article 1. Dans cette revue systématique de la littérature, nous soulignons le potentiel prédictif de l'AM dans l'étude des synucléinopathies par les hautes performances de classifications rapportées dans les études. Toutefois, nous mentionnons aussi son utilisation récente et la difficulté à extraire des tendances à partir des études jusqu'à maintenant réalisées, étant donné la grande variabilité de métriques et de classifications s'étant intéressées au sujet. Nous appelons donc à la rigueur et aux nuances dans l'interprétation des résultats découlant des travaux qui ont utilisé l'AM dans l'étude des synucléinopathies. Cela nous permet de mettre en évidence le manque de recoupement des études existantes et de souligner l'importance d'utiliser plusieurs métriques et d'étudier les classifications cognitives qui sont manquantes, soit des études transversales et longitudinales investiguant le statut cognitif (TCL ou démence) dans le TCSPi et la MP. Nous montrons également l'intérêt et l'utilité prédictive des données cognitives en AM, bien qu'encore embryonnaires. Quant aux algorithmes, beaucoup d'études sont encore à faire, mais le RF et le SVM montrent un grand potentiel. Les autres algorithmes présentés dans l'étude 1 méritent plus d'investigations afin de se prononcer sur leur valeur et efficacité.

Nous avons également montré l'importance de considérer la progression du TCSPi selon un riche continuum de classifications des synucléinopathies, puisque ce point de vue plus large est essentiel pour améliorer notre compréhension de l'évolution cognitive de ces maladies à partir des outils d'AM. Comme nous démontrons également l'intérêt de la multimodalité, nous suggérons que notre premier article de thèse serve de base au développement de modèles d'AM multimodaux et multivariés, à partir du continuum de performances et de variables performantes présentées. C'est d'ailleurs pour cette raison que nous avons fourni une description exhaustive des méthodes d'analyse (variables, méthodes de sélection des variables, algorithmes, type de *cross-validation*, sélection et importance des attributs/variables), afin de favoriser le transfert des connaissances aux chercheurs voulant poursuivre une étude en AM auprès de cette population clinique. De plus, le code sera mis à disposition sur GitHub lors de la publication de l'article.

Cette perspective plus large nous a permis d'identifier des lacunes dans la littérature, en particulier l'absence d'études sur le déclin cognitif dans le TCSPi. En effet, la majorité des études comparaient la MP par rapport à un groupe contrôle sain, avec un manque de recherche sur l'étude du déclin cognitif dans le TCSPi. L'article 2 de cette thèse comble ce vide en suggérant que les données cognitives en AM ont effectivement un potentiel pour prédire l'évolution vers la démence dans le TCSPi et pour distinguer cette trajectoire des autres conditions cliniques possibles. Nous montrons ainsi l'importance des marqueurs cognitifs liés à l'apprentissage verbal (mémoire et fonctions exécutives) dans la prédiction de la démence dans le TCSPi à partir de l'AM. Nous montrons aussi que l'AM permet de distinguer le sous-type de trajectoire évolutive (c.à.d., vers la DCL ou vers la MP), dévoilant qu'en AM dans une visée prédictive, le TCL est une variable cruciale dans la distinction des phénotypes prodromiques, le modèle d'AM performant à 80% de précision de façon significative.

Notre étude met également en lumière les subtilités distinctives entre l'AM et les statistiques classiques. Nous avons souligné que cette frontière n'est pas absolue et que ces deux approches sont complémentaires. Toutefois, il est important de noter que l'AM vise davantage l'optimisation des prédictions (dans le cadre d'une généralisation au-delà des données utilisées pour créer le modèle prédictif), contrairement à la statistique classique qui a une fonction plus explicative. Par conséquent, les variables pertinentes pour une analyse univariée inférentielle ne seront pas forcément les mêmes pour une approche prédictive multivariée (ou même univariée). Par exemple, nos résultats montrent que les mesures visuospatiales, qui étaient significatives en statistiques classiques pour distinguer l'évolution du TCSPi vers la DCL par rapport à la MP (ou le TCSP demeuré idiopathique), perdent de leur importance lorsqu'on les intègre dans un modèle prédictif multivariée en AM. Toutefois, lorsqu'il s'agit de prédire l'évolution vers la DCL par rapport au TCSPi sans TCL, les mesures d'apprentissage verbal identifiées à partir des statistiques classiques (RAVLT) sont retenues dans le modèle AM, en combinaison à des mesures exécutives, sociodémographiques et cliniques, d'importance et d'apparition équivalente dans le modèle. Notre étude fournit donc une première mise en contexte des variables précédemment étudiées, en les intégrant dans un contexte d'optimisation prédictive en AM pour des tâches de classification.



En explorant les voies d'optimisation des prédictions, nous reconnaissons la variabilité considérable dans la présentation clinique et la progression des symptômes de ces maladies. D'après nous, cette complexité peut être capturée par des algorithmes d'AM. Des études comme la nôtre contribuent donc à l'élaboration de modèles performants. Cela peut ouvrir la voie à l'identification des patients avec un TCSPi à haut risque de développer une démence pour des essais cliniques de médicaments. Dans l'éventualité où un médicament réussirait à prévenir la progression vers la démence chez des patients préalablement identifiés comme étant à haut risque, via un algorithme d'AM, cela pourrait offrir une opportunité de "*reverse engineering*". Cette approche consisterait à analyser l'impact du médicament sur la prévention de la démence, ce qui pourrait contribuer à élucider le mécanisme déclencheur de la neurodégénérescence, un phénomène qui à ce jour demeure incompris. De plus, un dépistage précoce des individus avec un TCSPi à risque de développer une démence pourrait permettre une prise en charge plus adaptée et personnalisée de ces patients, en plus d'alléger le fardeau sur le système de santé. Par ailleurs, le coût de la prise en charge de la démence est sensiblement réduit lorsque le diagnostic est posé à un stade précoce (Leifer, 2003). C'est pourquoi il est crucial que les études actuelles accordent une attention particulière à la validation des modèles d'AM, pour favoriser leur applicabilité à une population plus large. C'est dans cette optique que nous avons utilisé une méthodologie rigoureuse dans les deux articles de cette thèse, en nous concentrant exclusivement sur les modèles d'AM avec validation de modèle (validation croisée ou *cross-validation*). Nous avons également privilégié l'utilisation d'outils non invasifs, accessibles et abordables, tels que les tests cognitifs, les variables cliniques et l'EEG. Ces outils facilitent le transfert des connaissances théoriques en applications cliniques. Par ailleurs, il convient d'exercer une grande prudence lors de l'utilisation des algorithmes développés, en particulier lorsqu'il s'agit de prendre des décisions cliniques.

En résumé, cette thèse a mis en évidence le potentiel et l'importance de l'application rigoureuse de l'AM pour prédire le déclin cognitif dans les synucléinopathies. Alors que nous progressons dans l'ère de l'information et de la médecine de précision, il est essentiel que nous continuions à développer et à affiner ces outils, en veillant à la fois à leur rigueur scientifique et à leur pertinence clinique. L'avenir de la prédiction du déclin cognitif dans les synucléinopathies, et dans d'autres maladies neurodégénératives dépendra en grande partie de notre capacité à relever ces défis. Bien que nous ayons rencontré de nombreuses limites et défis dans le développement de

ces deux articles, cela a fourni une information précieuse sur l'état actuel de la recherche dans ce domaine. Puis, par notre méthodologie rigoureuse, nous avons tracé la voie et fourni des orientations pour les recherches futures.

## Limites

Quelques limites de cette thèse sont à souligner. Tout d'abord, dans l'article 1, une limite majeure concerne le nombre restreint d'études qui ont exploré la trajectoire clinique des troubles cognitifs dans les synucléinopathies à partir de l'AM avec une validation croisée, ainsi qu'avec des données EEG et cognitives. En raison du nombre insuffisant d'études et de la variété des variables, des métriques et des algorithmes utilisés, il n'était pas possible de mener une méta-analyse ou des analyses quantitatives. Ainsi, bien que nous ayons pu offrir une vue d'ensemble, nous n'avons pu identifier de combinaison de variables et/ou d'algorithmes spécifiques permettant d'optimiser chacune des classifications.

En ce qui concerne l'article 2, plusieurs limites doivent être notées. Tout d'abord, au coût d'une optimisation exhaustive des variables (*exhaustive feature selection*; c.à.d., testant toutes les combinaisons possibles des variables disponibles) et des hyperparamètres à travers l'ensemble des classifications proposées, nous n'avons pas pu explorer d'autres options algorithmiques, ni de métriques de performance, en raison du temps de calcul élevé que cela a généré, bien que nous ayons utilisé le système de Calcul Canada. De plus, par soucis d'accessibilité clinique des méthodes sélectionnées, nous n'avons pas inclus différentes méthodes d'imagerie cérébrale. Nous n'avons donc pas présenté les résultats en AM provenant de la combinaison de diverses modalités d'imagerie, même s'il a été montré que cela pourrait améliorer les performances de classification dans d'autres maladies neurodégénératives (Sharma et Mandal, 2022; Venugopalan et al., 2021). Nous n'avons pas non plus abordé l'impact du sous-type de TCL sur les trajectoires cliniques des patients avec un TCSPi, en raison de la taille limitée des groupes, qui rendait l'utilisation des algorithmes d'AM inadéquate. Enfin, une limite importante est que malgré l'accès à un des plus grands ensembles de données longitudinales sur les individus avec un TCSPi, il y avait des données manquantes. Cela nous a contraint à retirer certains patients et/ou certaines variables dans toutes

nos classifications, compromettant l'uniformité des analyses conduites et causant une perte d'informations à la suite du nettoyage des données.

### **Pistes de recherche futures**

Cette thèse trace la voie pour les prochaines études qui s'intéresseront à la classification des synucléinopathies à partir de l'utilisation de l'AM. Nous suggérons d'utiliser différentes métriques, différents algorithmes et de donner accès aux matrices de confusion. Cela permettra d'améliorer notre compréhension des performances des différentes catégories d'algorithmes, selon les variables utilisées. Nous suggérons également l'exploration du continuum des classifications possibles des synucléinopathies en EEG et en neuropsychologie, surtout pour les classifications les moins étudiées, soit les études transversales et longitudinales investiguant le statut cognitif (TCL ou démence) dans le TCSPi et la MP. Ces futures études permettraient d'enrichir le portrait des connaissances existantes et éventuellement permettre d'optimiser davantage les algorithmes de prédiction en AM.

Ensuite, inspiré par la *Parkinson's Progression Markers Initiative* pour l'étude de la MP, nous soulignons l'intérêt d'une uniformisation de la collecte des données dans l'étude du TCSPi. Par exemple, une base de données internationale accessible à la communauté d'intelligence artificielle pourrait accélérer le développement des connaissances dans le domaine de l'AM appliquée aux synucléinopathies, puisque les modèles d'AM sont encore plus performants et précis lorsqu'entraînés sur de plus grandes bases de données.

Enfin, l'adoption d'une approche multimodale serait prometteuse. Nous suggérons une intégration non seulement des biomarqueurs de diverses modalités, incluant la neuroimagerie, mais également des données cognitives, qui ont montré leur valeur prédictive en utilisant l'AM dans nos études. Les articles 1 et 2 ont mis en lumière les performances de diverses variables à travers différentes modalités et classifications, posant ainsi les fondations pour les études multimodales à venir.

## **Conclusion**

Cette thèse de doctorat explore l'application de l'AM pour prédire le déclin cognitif dans les synucléinopathies, en utilisant une approche méthodologique rigoureuse visant à maximiser le potentiel de généralisation des résultats hors-échantillon d'entraînement. À travers une revue systématique de la littérature et l'étude de la plus grande cohorte internationale de patients souffrant d'un TCSPi fournie par le CÉAMS, cette thèse identifie le potentiel inexploité des données neuropsychologiques dans le contexte de l'AM et la nécessité de continuer de développer des algorithmes plus optimaux. Plus précisément, ce travail met en lumière le potentiel de l'AM dans la prédiction de l'évolution de la démence dans le TCSPi et le rôle crucial des marqueurs cognitifs d'apprentissage (mémoire et fonctions exécutives) et du TCL pour distinguer les phénotypes prodromiques. Cette thèse va au-delà de l'exploration de l'AM dans les synucléinopathies, en contribuant à un cadre qui pourrait façonner les futurs modèles prédictifs du déclin cognitif dans les synucléinopathies, mais également dans d'autres maladies neurodégénératives. L'objectif ultime est de déposer les bases pour les futures recherches qui visent la prévention de la neurodénégescence, en tirant profit d'outils accessibles et de méthodes de calcul avancées.

## Références bibliographiques

- Aarsland, D. (2016, Jan). Cognitive impairment in Parkinson's disease and dementia with Lewy bodies. *Parkinsonism Relat Disord*, 22 Suppl 1, S144-148. <https://doi.org/10.1016/j.parkreldis.2015.09.034>
- Aarsland, D., Andersen, K., Larsen, J. P., Lolk, A., Nielsen, H. et Kragh-Sorensen, P. (2001, Mar 27). Risk of dementia in Parkinson's disease: a community-based, prospective study. *Neurology*, 56(6), 730-736. <https://doi.org/10.1212/wnl.56.6.730>
- Aarsland, D., Bronnick, K., Larsen, J. P., Tysnes, O. B., Alves, G. et Norwegian ParkWest Study, G. (2009, Mar 31). Cognitive impairment in incident, untreated Parkinson disease: the Norwegian ParkWest study. *Neurology*, 72(13), 1121-1126. <https://doi.org/10.1212/01.wnl.0000338632.00552.cb>
- Abu-Mostafa, Y. S., Magdon-Ismail, M. et Lin, H.-T. (2012). *Learning from data* (vol. 4). AMLBook New York.
- Adler, C. H. et Beach, T. G. (2016, Aug). Neuropathological basis of nonmotor manifestations of Parkinson's disease. *Mov Disord*, 31(8), 1114-1119. <https://doi.org/10.1002/mds.26605>
- Adler, J. (2010). *R in a nutshell: A desktop quick reference*. " O'Reilly Media, Inc."
- Al-Qazzaz, N. K., Ali, S. H., Ahmad, S. A., Chellappan, K., Islam, M. S. et Escudero, J. (2014). Role of EEG as biomarker in the early detection and classification of dementia. *ScientificWorldJournal*, 2014, 906038. <https://doi.org/10.1155/2014/906038>
- Albin, R., Koeppe, R., Chervin, R., Consens, F., Wernette, K., Frey, K. et Aldrich, M. (2000). Decreased striatal dopaminergic innervation in REM sleep behavior disorder. *Neurology*, 55(9), 1410-1412.
- Alexander, G. E., DeLong, M. R., & Strick, P. L. . (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience*, 9(1), 357-381.
- Anang, J. B., Gagnon, J. F., Bertrand, J. A., Romenets, S. R., Latreille, V., Panisset, M., Montplaisir, J. et Postuma, R. B. (2014, Sep 30). Predictors of dementia in Parkinson disease: a prospective cohort study. *Neurology*, 83(14), 1253-1260. <https://doi.org/10.1212/WNL.0000000000000842>
- Arisoy, E., Sainath, T. N., Kingsbury, B., & Ramabhadran, B. (2012, June). Deep neural network language models. Dans. Proceedings of the NAACL-HLT 2012 Workshop: Will We Ever Really Replace the N-gram Model? On the Future of Language Modeling for HLT.
- Armstrong, M. J. (2021). Advances in dementia with Lewy bodies. *Therapeutic Advances in Neurological Disorders*, 14, 17562864211057666.
- Arnulf, I., Bonnet, A.-M., Damier, P., Bejjani, B.-P., Seilhean, D., Derenne, J.-P. et Agid, Y. (2000). Hallucinations, REM sleep, and Parkinson's disease: a medical hypothesis. *Neurology*, 55(2), 281-288.
- Aston-Jones, G., Rajkowski, J. et Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biological psychiatry*, 46(9), 1309-1320.
- Austin, P. C., van Klaveren, D., Vergouwe, Y., Nieboer, D., Lee, D. S. et Steyerberg, E. W. (2016, Nov). Geographic and temporal validity of prediction models: different approaches were useful to examine model performance. *J Clin Epidemiol*, 79, 76-85. <https://doi.org/10.1016/j.jclinepi.2016.05.007>

- Baghdoyan, H. A. et Lydic, R. (2012). The neurochemistry of sleep and wakefulness. Dans *Basic neurochemistry* (p. 982-999). Elsevier.
- Bang, Y. R., Jeon, H. J., Youn, S. et Yoon, I. Y. (2017, Jan 10). Alterations of awake EEG in idiopathic REM sleep behavior disorder without cognitive impairment. *Neurosci Lett*, 637, 64-69. <https://doi.org/10.1016/j.neulet.2016.11.052>
- Beach, T. G., Adler, C. H., Lue, L., Sue, L. I., Bachalakuri, J., Henry-Watson, J., Sasse, J., Boyer, S., Shirohi, S. et Brooks, R. (2009). Unified staging system for Lewy body disorders: correlation with nigrostriatal degeneration, cognitive impairment and motor dysfunction. *Acta neuropathologica*, 117, 613-634.
- Bear, M. F., Connors, B. W. et Paradiso, M. A. (2016). *Neurosciences: à la découverte du cerveau* (vol. 4). John Libbey Eurotext.
- Belic, M., Bobic, V., Badza, M., Solaja, N., Duric-Jovicic, M. et Kostic, V. S. (2019, Sep). Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease-A review. *Clin Neurol Neurosurg*, 184, 105442. <https://doi.org/10.1016/j.clineuro.2019.105442>
- Berg, D., Borghammer, P., Fereshtehnejad, S.-M., Heinzl, S., Horsager, J., Schaeffer, E. et Postuma, R. B. (2021). Prodromal Parkinson disease subtypes—key to understanding heterogeneity. *Nature Reviews Neurology*, 17(6), 349-361.
- Berridge, C., Page, M., Valentino, R. et Foote, S. (1993). Effects of locus coeruleus inactivation on electroencephalographic activity in neocortex and hippocampus. *Neuroscience*, 55(2), 381-393.
- Bertrand, J. A., McIntosh, A. R., Postuma, R. B., Kovacevic, N., Latreille, V., Panisset, M., Chouinard, S. et Gagnon, J. F. (2016, Apr). Brain Connectivity Alterations Are Associated with the Development of Dementia in Parkinson's Disease. *Brain Connect*, 6(3), 216-224. <https://doi.org/10.1089/brain.2015.0390>
- Betrouni, N., Delval, A., Chaton, L., Defebvre, L., Duits, A., Moonen, A., Leentjens, A. F. G. et Dujardin, K. (2019, Feb). Electroencephalography-based machine learning for cognitive profiling in Parkinson's disease: Preliminary results. *Mov Disord*, 34(2), 210-217. <https://doi.org/10.1002/mds.27528>
- Bezdicek, O., Nikolai, T., Nepozitek, J., Perinova, P., Kemlink, D., Dusek, P., Prihodova, I., Dostalova, S., Ibarburu, V., Trnka, J., Kupka, K., Meckova, Z., Keller, J., Vymazal, J., Ruzicka, E., Sonka, K. et Dusek, P. (2018, Jul). Prospective memory impairment in idiopathic REM sleep behavior disorder. *Clin Neuropsychol*, 32(5), 1019-1037. <https://doi.org/10.1080/13854046.2017.1394493>
- Bhardwaj, A., Tiwari, A., Bhardwaj, H. et Bhardwaj, A. (2016). A genetically optimized neural network model for multi-class classification. *Expert Systems with Applications*, 60, 211-221. <https://doi.org/10.1016/j.eswa.2016.04.036>
- Bodkin, C. L. et Schenck, C. H. (2009). Rapid eye movement sleep behavior disorder in women: relevance to general and specialty medical practice. *Journal of Women's Health*, 18(12), 1955-1963.
- Boeve, B. F. (2010). REM sleep behavior disorder: updated review of the core features, the REM sleep behavior disorder-neurodegenerative disease association, evolving concepts, controversies, and future directions. *Annals of the New York Academy of Sciences*, 1184(1), 15-54.
- Bombois, S., Debette, S., Bruandet, A., Delbeuck, X., Delmaire, C., Leys, D. et Pasquier, F. (2008, Jul). Vascular subcortical hyperintensities predict conversion to vascular and mixed

- dementia in MCI patients. *Stroke*, 39(7), 2046-2051. <https://doi.org/10.1161/STROKEAHA.107.505206>
- Bonanni, L., Franciotti, R., Nobili, F., Kramberger, M. G., Taylor, J. P., Garcia-Ptacek, S., Falasca, N. W., Fama, F., Cromarty, R., Onofrj, M., Aarsland, D. et group, E. D. s. (2016, Oct 18). EEG Markers of Dementia with Lewy Bodies: A Multicenter Cohort Study. *J Alzheimers Dis*, 54(4), 1649-1657. <https://doi.org/10.3233/JAD-160435>
- Bonanni, L., Perfetti, B., Bifulchetti, S., Taylor, J. P., Franciotti, R., Parnetti, L., Thomas, A. et Onofrj, M. (2015, Jan). Quantitative electroencephalogram utility in predicting conversion of mild cognitive impairment to dementia with Lewy bodies. *Neurobiol Aging*, 36(1), 434-445. <https://doi.org/10.1016/j.neurobiolaging.2014.07.009>
- Bonanni, L., Thomas, A., Tiraboschi, P., Perfetti, B., Varanese, S. et Onofrj, M. (2008, Mar). EEG comparisons in early Alzheimer's disease, dementia with Lewy bodies and Parkinson's disease with dementia patients with a 2-year follow-up. *Brain*, 131(Pt 3), 690-705. <https://doi.org/10.1093/brain/awm322>
- Bougea, A., Efthymiopoulou, E., Spanou, I. et Zikos, P. (2022, May). A Novel Machine Learning Algorithm Predicts Dementia With Lewy Bodies Versus Parkinson's Disease Dementia Based on Clinical and Neuropsychological Scores. *J Geriatr Psychiatry Neurol*, 35(3), 317-320. <https://doi.org/10.1177/0891988721993556>
- Bourgouin, P. A., Rahayel, S., Gaubert, M., Arnaldi, D., Hu, M., Heidbreder, A., Remillard-Pelchat, D., Terzaghi, M. et Gagnon, J. F. (2019). Neuroimaging of Rapid Eye Movement Sleep Behavior Disorder. *Int Rev Neurobiol*, 144, 185-210. <https://doi.org/10.1016/bs.irn.2018.10.006>
- Braak, H., Del Tredici, K., Rüb, U., De Vos, R. A., Steur, E. N. J. et Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of aging*, 24(2), 197-211.
- Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3). <https://doi.org/10.1214/ss/1009213726>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*.
- Brennan, L., Devlin, K. M., Xie, S. X., Mechanic-Hamilton, D., Tran, B., Hurtig, H. H., Chen-Plotkin, A., Chahine, L. M., Morley, J. F., Duda, J. E., Roalf, D. R., Dahodwala, N., Rick, J., Trojanowski, J. Q., Moberg, P. J. et Weintraub, D. (2017). Neuropsychological Subgroups in Non-Demented Parkinson's Disease: A Latent Class Analysis. *J Parkinsons Dis*, 7(2), 385-395. <https://doi.org/10.3233/JPD-171081>
- Briel, R. C., McKeith, I. G., Barker, W. A., Hewitt, Y., Perry, R. H., Ince, P. G. et Fairbairn, A. F. (1999, Mar). EEG findings in dementia with Lewy bodies and Alzheimer's disease. *J Neurol Neurosurg Psychiatry*, 66(3), 401-403. <https://doi.org/10.1136/jnnp.66.3.401>
- Broeders, M., Velseboer, D. C., de Bie, R., Speelman, J. D., Muslimovic, D., Post, B., de Haan, R. et Schmand, B. (2013, Jul). Cognitive change in newly-diagnosed patients with Parkinson's disease: a 5-year follow-up study. *J Int Neuropsychol Soc*, 19(6), 695-708. <https://doi.org/10.1017/S1355617713000295>
- Brooks, P. L. et Peever, J. H. (2012). Identification of the transmitter and receptor mechanisms responsible for REM sleep paralysis. *Journal of Neuroscience*, 32(29), 9785-9795.
- Burges, C. J. C. (1998). *Data Mining and Knowledge Discovery*, 2(2), 121-167. <https://doi.org/10.1023/a:1009715923555>

- Buter, T. C., van den Hout, A., Matthews, F. E., Larsen, J. P., Brayne, C. et Aarsland, D. (2008, Mar 25). Dementia and survival in Parkinson disease: a 12-year population study. *Neurology*, 70(13), 1017-1022. <https://doi.org/10.1212/01.wnl.0000306632.43729.24>
- Byeon, H. (2020a, Nov 19). Best early-onset Parkinson dementia predictor using ensemble learning among Parkinson's symptoms, rapid eye movement sleep disorder, and neuropsychological profile. *World J Psychiatry*, 10(11), 245-259. <https://doi.org/10.5498/wjp.v10.i11.245>
- Byeon, H. (2020b, Apr 10). Is the Random Forest Algorithm Suitable for Predicting Parkinson's Disease with Mild Cognitive Impairment out of Parkinson's Disease with Normal Cognition? *Int J Environ Res Public Health*, 17(7). <https://doi.org/10.3390/ijerph17072594>
- Bzdok, D. (2017). Classical Statistics and Statistical Learning in Imaging Neuroscience. *Front Neurosci*, 11, 543. <https://doi.org/10.3389/fnins.2017.00543>
- Bzdok, D. et Ioannidis, J. P. (2019). Exploration, inference, and prediction in neuroscience and biomedicine. *Trends in neurosciences*, 42(4), 251-262.
- Bzdok, D. et Meyer-Lindenberg, A. (2018, Mar). Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging*, 3(3), 223-230. <https://doi.org/10.1016/j.bpsc.2017.11.007>
- Bzdok, D. et Yeo, B. T. T. (2017, Jul 15). Inference in the age of big data: Future perspectives on neuroscience. *Neuroimage*, 155, 549-564. <https://doi.org/10.1016/j.neuroimage.2017.04.061>
- Cagnin, A., Busse, C., Gardini, S., Jelcic, N., Guzzo, C., Gnoato, F., Mitolo, M., Ermani, M. et Caffarra, P. (2015, Sep-Dec). Clinical and Cognitive Phenotype of Mild Cognitive Impairment Evolving to Dementia with Lewy Bodies. *Dement Geriatr Cogn Dis Extra*, 5(3), 442-449. <https://doi.org/10.1159/000441184>
- Cai, J., Luo, J., Wang, S. et Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, 300, 70-79. <https://doi.org/10.1016/j.neucom.2017.11.077>
- Caviness, J. N., Hentz, J. G., Belden, C. M., Shill, H. A., Driver-Dunckley, E. D., Sabbagh, M. N., Powell, J. J. et Adler, C. H. (2015). Longitudinal EEG changes correlate with cognitive measure deterioration in Parkinson's disease. *J Parkinsons Dis*, 5(1), 117-124. <https://doi.org/10.3233/JPD-140480>
- Caviness, J. N., Hentz, J. G., Evidente, V. G., Driver-Dunckley, E., Samanta, J., Mahant, P., Connor, D. J., Sabbagh, M. N., Shill, H. A. et Adler, C. H. (2007, Aug). Both early and late cognitive dysfunction affects the electroencephalogram in Parkinson's disease. *Parkinsonism Relat Disord*, 13(6), 348-354. <https://doi.org/10.1016/j.parkreldis.2007.01.003>
- Caviness, J. N., Utianski, R. L., Hentz, J. G., Beach, T. G., Dugger, B. N., Shill, H. A., Driver-Dunckley, E. D., Sabbagh, M. N., Mehta, S. et Adler, C. H. (2016, Feb). Differential spectral quantitative electroencephalography patterns between control and Parkinson's disease cohorts. *Eur J Neurol*, 23(2), 387-392. <https://doi.org/10.1111/ene.12878>
- Cawley, G. C. T., N.L.C. (2010). On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *The Journal of Machine Learning Research*, 11, 2079-2107.
- Cesari, M., Christensen, J. A. E., Muntean, M. L., Mollenhauer, B., Sixel-Doring, F., Sorensen, H. B. D., Trenkwalder, C. et Jennum, P. (2021, Jan). A data-driven system to identify REM sleep behavior disorder and to predict its progression from the prodromal stage in Parkinson's disease. *Sleep Med*, 77, 238-248. <https://doi.org/10.1016/j.sleep.2020.04.010>



- Chang, C.-H., Lin, C.-H. et Lane, H.-Y. (2021). Machine Learning and Novel Biomarkers for the Diagnosis of Alzheimer's Disease. *International Journal of Molecular Sciences*, 22(5). <https://doi.org/10.3390/ijms22052761>
- Chaturvedi, M., Bogaarts, J. G., Kozak Cozac, V. V., Hatz, F., Gschwandtner, U., Meyer, A., Fuhr, P. et Roth, V. (2019, Oct). Phase lag index and spectral power as QEEG features for identification of patients with mild cognitive impairment in Parkinson's disease. *Clin Neurophysiol*, 130(10), 1937-1944. <https://doi.org/10.1016/j.clinph.2019.07.017>
- Chaturvedi, M., Hatz, F., Gschwandtner, U., Bogaarts, J. G., Meyer, A., Fuhr, P. et Roth, V. (2017). Quantitative EEG (QEEG) Measures Differentiate Parkinson's Disease (PD) Patients from Healthy Controls (HC). *Front Aging Neurosci*, 9, 3. <https://doi.org/10.3389/fnagi.2017.00003>
- Cholerton, B., Johnson, C. O., Fish, B., Quinn, J. F., Chung, K. A., Peterson-Hiller, A. L., Rosenthal, L. S., Dawson, T. M., Albert, M. S., Hu, S. C., Mata, I. F., Leverenz, J. B., Poston, K. L., Montine, T. J., Zabetian, C. P. et Edwards, K. L. (2018, May). Sex differences in progression to mild cognitive impairment and dementia in Parkinson's disease. *Parkinsonism Relat Disord*, 50, 29-36. <https://doi.org/10.1016/j.parkreldis.2018.02.007>
- Chong-Wen, W., Sha-Sha, L. et Xu, E. (2022). Predictors of rapid eye movement sleep behavior disorder in patients with Parkinson's disease based on random forest and decision tree. *PLoS one*, 17(6), e0269392. <https://doi.org/10.1371/journal.pone.0269392>
- Combrisson, E. et Jerbi, K. (2015, Jul 30). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J Neurosci Methods*, 250, 126-136. <https://doi.org/10.1016/j.jneumeth.2015.01.010>
- Cozac, V. V., Chaturvedi, M., Hatz, F., Meyer, A., Fuhr, P. et Gschwandtner, U. (2016). Increase of EEG Spectral Theta Power Indicates Higher Risk of the Development of Severe Cognitive Decline in Parkinson's Disease after 3 Years. *Front Aging Neurosci*, 8, 284. <https://doi.org/10.3389/fnagi.2016.00284>
- Cummings, M. P., Myers, D. S. et Mangelson, M. (2004). Applying permutation tests to tree-based statistical models: extending the R package rpart. Dans *Tech Rep CS-TR-4581, UMIACS-TR-2004-24, Center for Bioinformatics and Computational Biology, Institute for Advanced Computer Studies, University of Maryland*. Citeseer.
- Dalrymple-Alford, J. C., Livingston, L., MacAskill, M. R., Graham, C., Melzer, T. R., Porter, R. J., Watts, R. et Anderson, T. J. (2011, Mar). Characterizing mild cognitive impairment in Parkinson's disease. *Mov Disord*, 26(4), 629-636. <https://doi.org/10.1002/mds.23592>
- Dauvilliers, Y. (2019). *Les troubles du sommeil*. Elsevier Health Sciences.
- Dauwan, M., van der Zande, J. J., van Dellen, E., Sommer, I. E., Scheltens, P., Lemstra, A. W. et Stam, C. J. (2016). Random forest to differentiate dementia with Lewy bodies from Alzheimer's disease. *Alzheimers Dement (Amst)*, 4, 99-106. <https://doi.org/10.1016/j.dadm.2016.07.003>
- Dauwels, J., Vialatte, F. et Cichocki, A. (2010, Sep). Diagnosis of Alzheimer's disease from EEG signals: where are we standing? *Curr Alzheimer Res*, 7(6), 487-505. <https://doi.org/10.2174/156720510792231720>
- De Roy, J., Postuma, R. B., Brillion-Corbeil, M., Montplaisir, J., Genier Marchand, D., Escudier, F., Panisset, M., Chouinard, S. et Gagnon, J. F. (2020). Detecting the Cognitive Prodrome of Dementia in Parkinson's Disease. *J Parkinsons Dis*, 10(3), 1033-1046. <https://doi.org/10.3233/JPD-191857>

- de Weerd, A. W., Perquin, W. V. M. et Jonkman, E. J. (1990). Role of the EEG in the Prediction of Dementia in Parkinson's Disease. *Dementia and Geriatric Cognitive Disorders*, 1(2), 115-118. <https://doi.org/10.1159/000107129>
- Del Tredici, K., Rub, U., De Vos, R. A., Bohl, J. R. et Braak, H. (2002, May). Where does parkinson disease pathology begin in the brain? *J Neuropathol Exp Neurol*, 61(5), 413-426. <https://doi.org/10.1093/jnen/61.5.413>
- Delazer, M., Hogl, B., Zamarian, L., Wenter, J., Ehrmann, L., Gschliesser, V., Brandauer, E., Poewe, W. et Frauscher, B. (2012, May 1). Decision making and executive functions in REM sleep behavior disorder. *Sleep*, 35(5), 667-673. <https://doi.org/10.5665/sleep.1828>
- Djulgovic, B. et Ioannidis, J. P. A. (2019). Precision medicine for individual patients should use population group averages and larger, not smaller, groups. *European Journal of Clinical Investigation*, 49(1). <https://doi.org/10.1111/eci.13031>
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78-87. <https://doi.org/10.1145/2347736.2347755>
- Donaghy, P. C., O'Brien, J. T. et Thomas, A. J. (2015, Jan). Prodromal dementia with Lewy bodies. *Psychol Med*, 45(2), 259-268. <https://doi.org/10.1017/S0033291714000816>
- Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4), 745-766.
- Dujardin, K., Defebvre, L., Grunberg, C., Becquet, E. et Destee, A. (2001, Feb). Memory and executive function in sporadic and familial Parkinson's disease. *Brain*, 124(Pt 2), 389-398. <https://doi.org/10.1093/brain/124.2.389>
- Dujardin, K., Dubois, B., Tison, F., Durif, F., Bourdeix, I., Pere, J. J., Duhamel, A. et group, E. s. (2010, Dec 15). Parkinson's disease dementia can be easily detected in routine clinical practice. *Mov Disord*, 25(16), 2769-2776. <https://doi.org/10.1002/mds.23391>
- Eichelberger, D., Calabrese, P., Meyer, A., Chaturvedi, M., Hatz, F., Fuhr, P. et Gschwandtner, U. (2017). Correlation of Visuospatial Ability and EEG Slowing in Patients with Parkinson's Disease. *Parkinsons Dis*, 2017, 3659784. <https://doi.org/10.1155/2017/3659784>
- Eisensehr, I., Linke, R., Noachtar, S., Schwarz, J., Gildehaus, F. et Tatsch, K. (2000). Reduced striatal dopamine transporters in idiopathic rapid eye movement sleep behaviour disorder: comparison with Parkinson's disease and controls. *Brain*, 123(6), 1155-1160.
- Elkhani, N. et Muniyandi, R. C. (2017). Membrane computing inspired feature selection model for microarray cancer data. *Intelligent Data Analysis*, 21, S137-S157. <https://doi.org/10.3233/ida-170875>
- Emre, M., Aarsland, D., Brown, R., Burn, D. J., Duyckaerts, C., Mizuno, Y., Broe, G. A., Cummings, J., Dickson, D. W., Gauthier, S., Goldman, J., Goetz, C., Korczyn, A., Lees, A., Levy, R., Litvan, I., McKeith, I., Olanow, W., Poewe, W., Quinn, N., Sampaio, C., Tolosa, E. et Dubois, B. (2007, Sep 15). Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Mov Disord*, 22(12), 1689-1707; quiz 1837. <https://doi.org/10.1002/mds.21507>
- English, M., Kumar, C., Ditterline, B. L., Drazin, D. et Dietz, N. (2022). Machine Learning in Neuro-Oncology, Epilepsy, Alzheimer's Disease, and Schizophrenia. *Acta Neurochir Suppl*, 134, 349-361. [https://doi.org/10.1007/978-3-030-85292-4\\_39](https://doi.org/10.1007/978-3-030-85292-4_39)
- Escudero, J., Sanei, S., Jarchi, D., Abasolo, D. et Hornero, R. (2011, Aug). Regional coherence evaluation in mild cognitive impairment and Alzheimer's disease based on adaptively extracted magnetoencephalogram rhythms. *Physiol Meas*, 32(8), 1163-1180. <https://doi.org/10.1088/0967-3334/32/8/011>

- España, R. A. et Scammell, T. E. (2011). Sleep neurobiology from a clinical perspective. *Sleep*, 34(7), 845-858.
- Fantini, M., Gagnon, J.-F., Filipini, D. et Montplaisir, J. (2003). The effects of pramipexole in REM sleep behavior disorder. *Neurology*, 61(10), 1418-1420.
- Fantini, M. L., Farini, E., Ortelli, P., Zucconi, M., Manconi, M., Cappa, S., & Ferini-Strambi, L. (2011). Longitudinal Study of Cognitive Function in Idiopathic REM Sleep Behavior Disorder. *Sleep*, 34(5), 619-625. <https://doi.org/10.1093/sleep/34.5.619>
- Fereshtehnejad, S. M., Yao, C., Pelletier, A., Montplaisir, J. Y., Gagnon, J. F. et Postuma, R. B. (2019, Jul 1). Evolution of prodromal Parkinson's disease and dementia with Lewy bodies: a prospective study. *Brain*, 142(7), 2051-2067. <https://doi.org/10.1093/brain/awz111>
- Ferini-Strambi, L., Di Gioia, M. R., Castronovo, V., Oldani, A., Zucconi, M. et Cappa, S. F. (2004, Jan 13). Neuropsychological assessment in idiopathic REM sleep behavior disorder (RBD): does the idiopathic form of RBD really exist? *Neurology*, 62(1), 41-45. <https://doi.org/10.1212/01.wnl.0000101726.69701.fa>
- Ferini-Strambi, L., Fasiello, E., Sforza, M., Salsone, M. et Galbiati, A. (2019, Nov). Neuropsychological, electrophysiological, and neuroimaging biomarkers for REM behavior disorder. *Expert Rev Neurother*, 19(11), 1069-1087. <https://doi.org/10.1080/14737175.2019.1640603>
- Ferman, T. J., Smith, G. E., Boeve, B. F., Graff-Radford, N. R., Lucas, J. A., Knopman, D. S., Petersen, R. C., Ivnik, R. J., Wszolek, Z., Uitti, R. et Dickson, D. W. (2006, Dec). Neuropsychological differentiation of dementia with Lewy bodies from normal aging and Alzheimer's disease. *Clin Neuropsychol*, 20(4), 623-636. <https://doi.org/10.1080/13854040500376831>
- Fischer, P., Jungwirth, S., Zehetmayer, S., Weissgram, S., Hoenigschnabl, S., Gelpi, E., Krampla, W. et Tragl, K. (2007). Conversion from subtypes of mild cognitive impairment to Alzheimer dementia. *Neurology*, 68(4), 288-291.
- Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2), 179-188. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- Foltynie, T., Brayne, C. E., Robbins, T. W. et Barker, R. A. (2004, Mar). The cognitive ability of an incident cohort of Parkinson's patients in the UK. The CamPaIGN study. *Brain*, 127(Pt 3), 550-560. <https://doi.org/10.1093/brain/awh067>
- Fonseca, L. C., Tedrus, G. M., Letro, G. H. et Bossoni, A. S. (2009, Jul). Dementia, mild cognitive impairment and quantitative EEG in patients with Parkinson's disease. *Clin EEG Neurosci*, 40(3), 168-172. <https://doi.org/10.1177/155005940904000309>
- Frauscher, B., Gschliesser, V., Brandauer, E., Marti, I., Furtner, M. T., Ulmer, H., Poewe, W. et Högl, B. (2010). REM sleep behavior disorder in 703 sleep-disorder patients: the importance of eliciting a comprehensive sleep history. *Sleep medicine*, 11(2), 167-171.
- Fries, P. (2005, Oct). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci*, 9(10), 474-480. <https://doi.org/10.1016/j.tics.2005.08.011>
- Fusar-Poli, P., Hijazi, Z., Stahl, D. et Steyerberg, E. W. (2018, Dec 1). The Science of Prognosis in Psychiatry: A Review. *JAMA Psychiatry*, 75(12), 1289-1297. <https://doi.org/10.1001/jamapsychiatry.2018.2530>
- Gagnon, J.-F., Bertrand, J.-A. et Génier Marchand, D. (2012). Cognition in rapid eye movement sleep behavior disorder. *Frontiers in neurology*, 3, 82.

- Gagnon, J. F., Bourgooin, P. A., De Roy, J., & Génier Marchand, D. . (2019). Neuropsychological aspects: cognition in RBD. *Rapid-eye-movement sleep behavior disorder*, 491-507. [https://doi.org/https://doi.org/10.1007/978-3-319-90152-7\\_34](https://doi.org/https://doi.org/10.1007/978-3-319-90152-7_34)
- Gagnon, J. F., Vendette, M., Postuma, R. B., Desjardins, C., Massicotte-Marquez, J., Panisset, M. et Montplaisir, J. (2009). Mild cognitive impairment in rapid eye movement sleep behavior disorder and Parkinson's disease. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 66(1), 39-47.
- Galtier, I., Nieto, A., Lorenzo, J. N. et Barroso, J. (2016). Mild cognitive impairment in Parkinson's disease: Diagnosis and progression to dementia. *J Clin Exp Neuropsychol*, 38(1), 40-50. <https://doi.org/10.1080/13803395.2015.1087465>
- Galtier, I., Nieto, A., Lorenzo, J. N. et Barroso, J. (2017, Jul). Mild Cognitive Impairment in Parkinson's Disease: Clustering and Switching Analyses in Verbal Fluency Test. *J Int Neuropsychol Soc*, 23(6), 511-520. <https://doi.org/10.1017/S1355617717000297>
- Galtier, I., Nieto, A., Lorenzo, J. N. et Barroso, J. (2019, Mar). Subjective cognitive decline and progression to dementia in Parkinson's disease: a long-term follow-up study. *J Neurol*, 266(3), 745-754. <https://doi.org/10.1007/s00415-019-09197-0>
- Gandal, M. J., Edgar, J. C., Klook, K. et Siegel, S. J. (2012, Mar). Gamma synchrony: towards a translational biomarker for the treatment-resistant symptoms of schizophrenia. *Neuropharmacology*, 62(3), 1504-1518. <https://doi.org/10.1016/j.neuropharm.2011.02.007>
- Ganguli, M., Dodge, H. H., Shen, C. et DeKosky, S. T. (2004). Mild cognitive impairment, amnesic type: an epidemiologic study. *Neurology*, 63(1), 115-121.
- Gauggel, S., Rieger, M. et Feghoff, T. A. (2004, Apr). Inhibition of ongoing responses in patients with Parkinson's disease. *J Neurol Neurosurg Psychiatry*, 75(4), 539-544. <https://doi.org/10.1136/jnnp.2003.016469>
- Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., Cummings, J. L., de Leon, M., Feldman, H., Ganguli, M., Hampel, H., Scheltens, P., Tierney, M. C., Whitehouse, P., Winblad, B. et International Psychogeriatric Association Expert Conference on mild cognitive, i. (2006, Apr 15). Mild cognitive impairment. *Lancet*, 367(9518), 1262-1270. [https://doi.org/10.1016/S0140-6736\(06\)68542-5](https://doi.org/10.1016/S0140-6736(06)68542-5)
- Genier Marchand, D., Montplaisir, J., Postuma, R. B., Rahayel, S. et Gagnon, J. F. (2017, Jan 1). Detecting the Cognitive Prodrome of Dementia with Lewy Bodies: A Prospective Study of REM Sleep Behavior Disorder. *Sleep*, 40(1). <https://doi.org/10.1093/sleep/zsw014>
- Genier Marchand, D., Postuma, R. B., Escudier, F., De Roy, J., Pelletier, A., Montplaisir, J. et Gagnon, J. F. (2018, May). How does dementia with Lewy bodies start? prodromal cognitive changes in REM sleep behavior disorder. *Ann Neurol*, 83(5), 1016-1026. <https://doi.org/10.1002/ana.25239>
- Geraedts, V. J., Boon, L. I., Marinus, J., Gouw, A. A., van Hilten, J. J., Stam, C. J., Tannemaat, M. R. et Contarino, M. F. (2018, Nov 6). Clinical correlates of quantitative EEG in Parkinson disease: A systematic review. *Neurology*, 91(19), 871-883. <https://doi.org/10.1212/WNL.0000000000006473>
- Gompf, H. S. et Anaclet, C. (2020). The neuroanatomy and neurochemistry of sleep-wake control. *Current opinion in physiology*, 15, 143-151.
- Gonthier, C., & Hot, P (2013). Apports de l'électroencéphalographie à la compréhension de la mémoire. *Revue de neuropsychologie*, 5(4), 243-254.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.



- Guner, D., Tiftikcioglu, B. I., Tuncay, N. et Zorlu, Y. (2017, Sep). Contribution of Quantitative EEG to the Diagnosis of Early Cognitive Impairment in Patients With Idiopathic Parkinson's Disease. *Clin EEG Neurosci*, 48(5), 348-354. <https://doi.org/10.1177/1550059416662412>
- Halliday, G., Hely, M., Reid, W. et Morris, J. (2008). The progression of pathology in longitudinally followed patients with Parkinson's disease. *Acta neuropathologica*, 115, 409-415.
- Halliday, G. M., Holton, J. L., Revesz, T. et Dickson, D. W. (2011). Neuropathology underlying clinical variability in patients with synucleinopathies. *Acta neuropathologica*, 122, 187-204.
- Hamilton, J. M., Landy, K. M., Salmon, D. P., Hansen, L. A., Masliah, E. et Galasko, D. (2012, Sep). Early visuospatial deficits predict the occurrence of visual hallucinations in autopsy-confirmed dementia with Lewy bodies. *Am J Geriatr Psychiatry*, 20(9), 773-781. <https://doi.org/10.1097/JGP.0b013e31823033bc>
- Hampel, H., Lista, S., Teipel, S. J., Garaci, F., Nistico, R., Blennow, K., Zetterberg, H., Bertram, L., Duyckaerts, C., Bakardjian, H., Drzezga, A., Colliot, O., Epelbaum, S., Broich, K., Lehericy, S., Brice, A., Khachaturian, Z. S., Aisen, P. S. et Dubois, B. (2014, Apr 15). Perspective on future role of biological markers in clinical therapy trials of Alzheimer's disease: a long-range point of view beyond 2020. *Biochem Pharmacol*, 88(4), 426-449. <https://doi.org/10.1016/j.bcp.2013.11.009>
- Han, C. X., Wang, J., Yi, G. S. et Che, Y. Q. (2013, Aug). Investigation of EEG abnormalities in the early stage of Parkinson's disease. *Cogn Neurodyn*, 7(4), 351-359. <https://doi.org/10.1007/s11571-013-9247-z>
- Hari, R., & Puce, A. (2017). *MEG-EEG Primer*. Oxford University Press.
- Hassan, M., Chaton, L., Benquet, P., Delval, A., Leroy, C., Plomhause, L., Moonen, A. J., Duits, A. A., Leentjens, A. F., van Kranen-Mastenbroek, V., Defebvre, L., Derambure, P., Wendling, F. et Dujardin, K. (2017). Functional connectivity disruptions correlate with cognitive phenotypes in Parkinson's disease. *Neuroimage Clin*, 14, 591-601. <https://doi.org/10.1016/j.nicl.2017.03.002>
- He, X., Zhang, Y., Chen, J., Xie, C., Gan, R., Wang, L. et Wang, L. (2017, Jan). Changes in theta activities in the left posterior temporal region, left occipital region and right frontal region related to mild cognitive impairment in Parkinson's disease patients. *Int J Neurosci*, 127(1), 66-72. <https://doi.org/10.3109/00207454.2016.1143823>
- Helkala, E. L., Laulumaa, V., Soikkeli, R., Partanen, J., Soininen, H. et Riekkinen, P. J. (1991, Jun). Slow-wave activity in the spectral analysis of the electroencephalogram is associated with cortical dysfunctions in patients with Alzheimer's disease. *Behav Neurosci*, 105(3), 409-415. <https://doi.org/10.1037//0735-7044.105.3.409>
- Hely, M. A., Reid, W. G., Adena, M. A., Halliday, G. M. et Morris, J. G. (2008, Apr 30). The Sydney multicenter study of Parkinson's disease: the inevitability of dementia at 20 years. *Mov Disord*, 23(6), 837-844. <https://doi.org/10.1002/mds.21956>
- Hendricks, J. C., Morrison, A. R. et Mann, G. L. (1982). Different behaviors during paradoxical sleep without atonia depend on pontine lesion site. *Brain research*, 239(1), 81-105.
- Hobson, P. et Meara, J. (2015, Oct). Mild cognitive impairment in Parkinson's disease and its progression onto dementia: a 16-year outcome evaluation of the Denbighshire cohort. *Int J Geriatr Psychiatry*, 30(10), 1048-1055. <https://doi.org/10.1002/gps.4261>

- Hogl, B., Stefani, A. et Videnovic, A. (2018, Jan). Idiopathic REM sleep behaviour disorder and neurodegeneration - an update. *Nat Rev Neurol*, 14(1), 40-55. <https://doi.org/10.1038/nrneurol.2017.157>
- Hogue, O., Fernandez, H. H. et Floden, D. P. (2018, Nov). Predicting early cognitive decline in newly-diagnosed Parkinson's patients: A practical model. *Parkinsonism Relat Disord*, 56, 70-75. <https://doi.org/10.1016/j.parkreldis.2018.06.031>
- Honeycutt, L., Gagnon, J. F., Pelletier, A., De Roy, J., Montplaisir, J. Y. et Postuma, R. B. (2020, Jun). Pareidolias and cognition in isolated REM sleep behavior disorder. *Parkinsonism Relat Disord*, 75, 76-79. <https://doi.org/10.1016/j.parkreldis.2020.05.017>
- Hong, J. Y., Sunwoo, M. K., Chung, S. J., Ham, J. H., Lee, J. E., Sohn, Y. H. et Lee, P. H. (2014, Jul). Subjective cognitive decline predicts future deterioration in cognitively normal patients with Parkinson's disease. *Neurobiol Aging*, 35(7), 1739-1743. <https://doi.org/10.1016/j.neurobiolaging.2013.11.017>
- Hornero, R., Escudero, J., Fernandez, A., Poza, J. et Gomez, C. (2008, Jun). Spectral and nonlinear analyses of MEG background activity in patients with Alzheimer's disease. *IEEE Trans Biomed Eng*, 55(6), 1658-1665. <https://doi.org/10.1109/tbme.2008.919872>
- Hossin, M. et Sulaiman, M. N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2), 01-11. <https://doi.org/10.5121/ijdkp.2015.5201>
- Hu, B., Steriade, M. et Deschenes, M. (1989). The cellular mechanism of thalamic ponto-geniculo-occipital waves. *Neuroscience*, 31(1), 25-35.
- Hu, M. T. (2020, Sep). REM sleep behavior disorder (RBD). *Neurobiol Dis*, 143, 104996. <https://doi.org/10.1016/j.nbd.2020.104996>
- Ij, H. (2018, 2018). Statistics versus machine learning. *Nat Methods*, 15(4), 233.
- Iranzo, A., Isetta, V., Molinuevo, J. L., Serradell, M., Navajas, D., Farre, R. et Santamaria, J. (2010, Jun). Electroencephalographic slowing heralds mild cognitive impairment in idiopathic REM sleep behavior disorder. *Sleep Med*, 11(6), 534-539. <https://doi.org/10.1016/j.sleep.2010.03.006>
- Iranzo, A., Molinuevo, J. L., Santamaria, J., Serradell, M., Marti, M. J., Valldeoriola, F. et Tolosa, E. (2006). Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: a descriptive study. *The Lancet Neurology*, 5(7), 572-577.
- Isella, V., Mapelli, C., Morielli, N., Siri, C., De Gaspari, D., Pezzoli, G., Antonini, A., Poletti, M., Bonuccelli, U., Picchi, L., Napolitano, A., Vista, M. et Appollonio, I. M. (2013, Dec). Diagnosis of possible mild cognitive impairment in Parkinson's disease: validity of the SCOPA-Cog. *Parkinsonism Relat Disord*, 19(12), 1160-1163. <https://doi.org/10.1016/j.parkreldis.2013.08.008>
- Janvin, C. C., Aarsland, D. et Larsen, J. P. (2005, Sep). Cognitive predictors of dementia in Parkinson's disease: a community-based, 4-year longitudinal study. *J Geriatr Psychiatry Neurol*, 18(3), 149-154. <https://doi.org/10.1177/0891988705277540>
- Jellinger, K. A. (2014). The pathomechanisms underlying Parkinson's disease. *Expert review of neurotherapeutics*, 14(2), 199-215.
- Jellinger, K. A. (2019). Is Braak staging valid for all types of Parkinson's disease? *Journal of neural transmission*, 126, 423-431.
- Jeon, J., Kim, K., Baek, K., Chung, S. J., Yoon, J. et Kim, Y. J. (2022, May). Accuracy of Machine Learning Using the Montreal Cognitive Assessment for the Diagnosis of Cognitive Impairment in Parkinson's Disease. *J Mov Disord*, 15(2), 132-139. <https://doi.org/10.14802/jmd.22012>

- Jicha, G. A., Schmitt, F. A., Abner, E., Nelson, P. T., Cooper, G. E., Smith, C. D. et Markesbery, W. R. (2010, Oct). Prodromal clinical manifestations of neuropathologically confirmed Lewy body disease. *Neurobiol Aging*, 31(10), 1805-1813. <https://doi.org/10.1016/j.neurobiolaging.2008.09.017>
- [Record #267 is using a reference type undefined in this output style.]
- Jouvet, M. (1965). Locus coeruleus et sommeil paradoxical. *CR Soc Biol (Paris)*, 159, 895-899.
- Joza, S. H., MT; Jung, KY; Kunz, D; Arnaldi, D; Lee, JY; Ferini-Strambi, L; Antelmi, E; Sixel-Döring, F; Cochen De Cock, V; Montplaisir, JY; Welch, J; Kim, HJ; Bes, F; Mattioli, P; Woo, KA; Marelli, S; Plazzi, G; Mollenhauer, B; Pelletier, A; Razzaque, J; Sunwoo, JS; Girtler, N; Trenkwalder, C; Gagnon, JF; Postuma, RB; International REM Sleep Behavior Disorder Study Group. (2023). Prodromal dementia with Lewy bodies in REM sleep behavior disorder: A multicenter study. *Alzheimer's & Dementia* <https://doi.org/10.1002/alz.13386>
- Jurcau, A. et Nunkoo, V. S. (2021). Clinical Markers May Identify Patients at Risk for Early Parkinson's Disease Dementia: A Prospective Study. *American Journal of Alzheimer's Disease & Other Dementiasr*, 36. <https://doi.org/10.1177/15333175211021369>
- Kai, T., Asai, Y., Sakuma, K., Koeda, T. et Nakashima, K. (2005, Oct 15). Quantitative electroencephalogram analysis in dementia with Lewy bodies and Alzheimer's disease. *J Neurol Sci*, 237(1-2), 89-95. <https://doi.org/10.1016/j.jns.2005.05.017>
- Kamei, S., Morita, A., Serizawa, K., Mizutani, T. et Hirayanagi, K. (2010, Jun). Quantitative EEG analysis of executive dysfunction in Parkinson disease. *J Clin Neurophysiol*, 27(3), 193-197. <https://doi.org/10.1097/WNP.0b013e3181dd4fdb>
- Kang, S.-H., Yoon, I.-Y., Lee, S. D., Han, J. W., Kim, T. H. et Kim, K. W. (2013). REM sleep behavior disorder in the Korean elderly population: prevalence and clinical characteristics. *Sleep*, 36(8), 1147-1152.
- Kataoka, H. et Sugie, K. (2021, May 26). Association between Fatigue and Hoehn-Yahr Staging in Parkinson's Disease: Eight-Year Follow-Up Study. *Neurol Int*, 13(2), 224-231. <https://doi.org/10.3390/neurolint13020023>
- Kerai, J. H., Bracewell, R. M., Hindle, J. V. et Leek, E. C. (2012). Visuospatial transformation impairments in Parkinson's disease. *J Clin Exp Neuropsychol*, 34(10), 1053-1064. <https://doi.org/10.1080/13803395.2012.716396>
- Khare, S. K., Bajaj, V. et Acharya, U. R. (2021). Detection of Parkinson's disease using automated tunable Q wavelet transform technique with EEG signals. *Biocybernetics and Biomedical Engineering*, 41(2), 679-689. <https://doi.org/10.1016/j.bbe.2021.04.008>
- Khatri, U. et Kwon, G. R. (2020). An Efficient Combination among sMRI, CSF, Cognitive Score, and APOE epsilon4 Biomarkers for Classification of AD and MCI Using Extreme Learning Machine. *Comput Intell Neurosci*, 2020, 8015156. <https://doi.org/10.1155/2020/8015156>
- Kim, H. M., Nazor, C., Zabetian, C. P., Quinn, J. F., Chung, K. A., Hiller, A. L., Hu, S. C., Leverenz, J. B., Montine, T. J., Edwards, K. L. et Cholerton, B. (2019). Prediction of cognitive progression in Parkinson's disease using three cognitive screening measures. *Clin Park Relat Disord*, 1, 91-97. <https://doi.org/10.1016/j.prdoa.2019.08.006>
- Kingsmore, K. M., Puglisi, C. E., Grammer, A. C. et Lipsky, P. E. (2021, Dec). An introduction to machine learning and analysis of its use in rheumatic diseases. *Nat Rev Rheumatol*, 17(12), 710-730. <https://doi.org/10.1038/s41584-021-00708-w>
- Klassen, B. T., Hentz, J. G., Shill, H. A., Driver-Dunckley, E., Evidente, V. G., Sabbagh, M. N., Adler, C. H. et Caviness, J. N. (2011, Jul 12). Quantitative EEG as a predictive biomarker

- for Parkinson disease dementia. *Neurology*, 77(2), 118-124. <https://doi.org/10.1212/WNL.0b013e318224af8d>
- Kleiner, S. et Bringmann, A. (1996). Nucleus basalis magnocellularis and pedunculopontine tegmental nucleus: control of the slow EEG waves in rats. *Archives italiennes de biologie*, 134(2), 153-167.
- Klimesch, W. (1996, Nov). Memory processes, brain oscillations and EEG synchronization. *Int J Psychophysiol*, 24(1-2), 61-100. [https://doi.org/10.1016/s0167-8760\(96\)00057-8](https://doi.org/10.1016/s0167-8760(96)00057-8)
- Kondo, D., Ota, K., Kasanuki, K., Fujishiro, H., Chiba, Y., Murayama, N., Sato, K., Hirayasu, Y., Arai, H. et Iseki, E. (2016, Oct 15). Characteristics of mild cognitive impairment tending to convert into Alzheimer's disease or dementia with Lewy bodies: A follow-up study in a memory clinic. *J Neurol Sci*, 369, 102-108. <https://doi.org/10.1016/j.jns.2016.08.011>
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Emerging artificial intelligence applications in computer engineering. Dans *Supervised machine learning: A review of classification techniques* (vol. 160, p. 3-24).
- Krenzer, M., Anacllet, C., Vetrivelan, R., Wang, N., Vong, L., Lowell, B. B., Fuller, P. M. et Lu, J. (2011). Brainstem and spinal cord circuitry regulating REM sleep and muscle atonia. *PLoS one*, 6(10), e24998.
- Kudlicka, A., Clare, L. et Hindle, J. V. (2011, Nov). Executive functions in Parkinson's disease: systematic review and meta-analysis. *Mov Disord*, 26(13), 2305-2315. <https://doi.org/10.1002/mds.23868>
- Latreille, V., Carrier, J., Gaudet-Fex, B., Rodrigues-Brazete, J., Panisset, M., Chouinard, S., Postuma, R. B. et Gagnon, J. F. (2016, Apr). Electroencephalographic prodromal markers of dementia across conscious states in Parkinson's disease. *Brain*, 139(Pt 4), 1189-1199. <https://doi.org/10.1093/brain/aww018>
- Lee, S., Hussein, R., Ward, R., Jane Wang, Z. et McKeown, M. J. (2021, Sep 1). A convolutional-recurrent neural network approach to resting-state EEG classification in Parkinson's disease. *J Neurosci Methods*, 361, 109282. <https://doi.org/10.1016/j.jneumeth.2021.109282>
- Lee, W. J., Wu, S. L., Li, J. Y., Chen, C. C., Lin, J. J., Peng, G. S., Lin, T. K., Wang, S. J., Chen, R. S., Yang, Y. H., Tsai, C. H., Hsu, J. L., Liao, Y. C., Chen, S. P. et Fuh, J. L. (2016, Apr). PDD-5S: A useful screening tool for Parkinson's disease dementia. *Parkinsonism Relat Disord*, 25, 85-90. <https://doi.org/10.1016/j.parkreldis.2016.02.001>
- Leifer, B. P. (2003, May). Early diagnosis of Alzheimer's disease: clinical and economic benefits. *J Am Geriatr Soc*, 51(5 Suppl Dementia), S281-288. <https://doi.org/10.1046/j.1532-5415.5153.x>
- Leuchter, A. F., Cook, I. A., Gilmer, W. S., Marangell, L. B., Burgoyne, K. S., Howland, R. H., Trivedi, M. H., Zisook, S., Jain, R., Fava, M., Iosifescu, D. et Greenwald, S. (2009, Sep 30). Effectiveness of a quantitative electroencephalographic biomarker for predicting differential response or remission with escitalopram and bupropion in major depressive disorder. *Psychiatry Res*, 169(2), 132-138. <https://doi.org/10.1016/j.psychres.2009.04.004>
- Lever, J. (2016). Classification evaluation: It is important to understand both what a classification metric expresses and what it hides. *Nature methods*, 13(8), 603-605.
- Levy, G., Jacobs, D. M., Tang, M. X., Cote, L. J., Louis, E. D., Alfaró, B., Mejia, H., Stern, Y. et Marder, K. (2002, Nov). Memory and executive function impairment predict dementia in Parkinson's disease. *Mov Disord*, 17(6), 1221-1226. <https://doi.org/10.1002/mds.10280>
- Lewis, S. J., Slabosz, A., Robbins, T. W., Barker, R. A. et Owen, A. M. (2005). Dopaminergic basis for deficits in working memory but not attentional set-shifting in Parkinson's disease.



<https://doi.org/10.1016/j.neuropsychologia.2004.10.001>

- Li, X., Zhou, Z., Jia, S., Hou, C., Zheng, W., Rong, P. et Jiao, J. (2016, Jul 15). Cognitive study on Chinese patients with idiopathic REM sleep behavior disorder. *J Neurol Sci*, 366, 82-86. <https://doi.org/10.1016/j.jns.2016.04.047>
- Li, Z., Jiang, X., Wang, Y. et Kim, Y. (2021, Dec 21). Applied machine learning in Alzheimer's disease research: omics, imaging, and clinical data. *Emerg Top Life Sci*, 5(6), 765-777. <https://doi.org/10.1042/ETLS20210249>
- Lichter, D. G., Benedict, R. H. B. et Hershey, L. A. (2018). Importance of Balance-Gait Disorder as a Risk Factor for Cognitive Impairment, Dementia and Related Non-Motor Symptoms in Parkinson's Disease. *J Parkinsons Dis*, 8(4), 539-552. <https://doi.org/10.3233/JPD-181375>
- Litvan, I., Aarsland, D., Adler, C. H., Goldman, J. G., Kulisevsky, J., Mollenhauer, B., Rodriguez-Oroz, M. C., Troster, A. I. et Weintraub, D. (2011, Aug 15). MDS Task Force on mild cognitive impairment in Parkinson's disease: critical review of PD-MCI. *Mov Disord*, 26(10), 1814-1824. <https://doi.org/10.1002/mds.23823>
- Litvan, I., Goldman, J. G., Troster, A. I., Schmand, B. A., Weintraub, D., Petersen, R. C., Mollenhauer, B., Adler, C. H., Marder, K., Williams-Gray, C. H., Aarsland, D., Kulisevsky, J., Rodriguez-Oroz, M. C., Burn, D. J., Barker, R. A. et Emre, M. (2012, Mar). Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Mov Disord*, 27(3), 349-356. <https://doi.org/10.1002/mds.24893>
- Llebaria, G., Pagonabarraga, J., Kulisevsky, J., Garcia-Sanchez, C., Pascual-Sedano, B., Gironell, A. et Martinez-Corral, M. (2008, Aug 15). Cut-off score of the Mattis Dementia Rating Scale for screening dementia in Parkinson's disease. *Mov Disord*, 23(11), 1546-1550. <https://doi.org/10.1002/mds.22173>
- Lotte, F., Congedo, M., Lecuyer, A., Lamarche, F. et Arnaldi, B. (2007, Jun). A review of classification algorithms for EEG-based brain-computer interfaces. *J Neural Eng*, 4(2), R1-R13. <https://doi.org/10.1088/1741-2560/4/2/R01>
- Luque, A., Carrasco, A., Martín, A. et de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, 216-231. <https://doi.org/10.1016/j.patcog.2019.02.023>
- Magoulas, G. D., & Prentza, A. (1999). Machine learning in medical applications. Dans *Advanced course on artificial intelligence* (p. 300-307). Springer Berlin Heidelberg. [https://doi.org/https://doi.org/10.1007/3-540-44673-7\\_19](https://doi.org/https://doi.org/10.1007/3-540-44673-7_19)
- Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9, 381-386. <https://doi.org/10.21275/ART20203995>
- Manrai, A. K., Patel, C. J. et Ioannidis, J. P. A. (2018, May 15). In the Era of Precision Medicine and Big Data, Who Is Normal? *JAMA*, 319(19), 1981-1982. <https://doi.org/10.1001/jama.2018.2009>
- Marques, A., Dujardin, K., Boucart, M., Pins, D., Delliaux, M., Defebvre, L., Derambure, P. et Monaca, C. (2010, Mar). REM sleep behaviour disorder and visuoperceptive dysfunction: a disorder of the ventral visual stream? *J Neurol*, 257(3), 383-391. <https://doi.org/10.1007/s00415-009-5328-7>
- Massicotte-Marquez, J., Decary, A., Gagnon, J. F., Vendette, M., Mathieu, A., Postuma, R. B., Carrier, J. et Montplaisir, J. (2008, Apr 8). Executive dysfunction and memory impairment in idiopathic REM sleep behavior disorder. *Neurology*, 70(15), 1250-1257. <https://doi.org/10.1212/01.wnl.0000286943.79593.a6>

- McKeith, I., Mintzer, J., Aarsland, D., Burn, D., Chiu, H., Cohen-Mansfield, J., Dickson, D., Dubois, B., Duda, J. E., Feldman, H., Gauthier, S., Halliday, G., Lawlor, B., Lippa, C., Lopez, O. L., Carlos Machado, J., O'Brien, J., Playfer, J., Reid, W. et International Psychogeriatric Association Expert Meeting on, D. L. B. (2004, Jan). Dementia with Lewy bodies. *Lancet Neurol*, 3(1), 19-28. [https://doi.org/10.1016/s1474-4422\(03\)00619-7](https://doi.org/10.1016/s1474-4422(03)00619-7)
- McKeith, I. G., Boeve, B. F., Dickson, D. W., Halliday, G., Taylor, J. P., Weintraub, D., Aarsland, D., Galvin, J., Attems, J., Ballard, C. G., Bayston, A., Beach, T. G., Blanc, F., Bohnen, N., Bonanni, L., Bras, J., Brundin, P., Burn, D., Chen-Plotkin, A., Duda, J. E., El-Agnaf, O., Feldman, H., Ferman, T. J., Ffytche, D., Fujishiro, H., Galasko, D., Goldman, J. G., Gomperts, S. N., Graff-Radford, N. R., Honig, L. S., Iranzo, A., Kantarci, K., Kaufer, D., Kukull, W., Lee, V. M. Y., Leverenz, J. B., Lewis, S., Lippa, C., Lunde, A., Masellis, M., Masliah, E., McLean, P., Mollenhauer, B., Montine, T. J., Moreno, E., Mori, E., Murray, M., O'Brien, J. T., Orimo, S., Postuma, R. B., Ramaswamy, S., Ross, O. A., Salmon, D. P., Singleton, A., Taylor, A., Thomas, A., Tiraboschi, P., Toledo, J. B., Trojanowski, J. Q., Tsuang, D., Walker, Z., Yamada, M. et Kosaka, K. (2017, Jul 4). Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium. *Neurology*, 89(1), 88-100. <https://doi.org/10.1212/WNL.0000000000004058>
- McKeith, I. G., Dickson, D. W., Lowe, J., Emre, M., O'Brien, J. T., Feldman, H., Cummings, J., Duda, J. E., Lippa, C., Perry, E. K., Aarsland, D., Arai, H., Ballard, C. G., Boeve, B., Burn, D. J., Costa, D., Del Ser, T., Dubois, B., Galasko, D., Gauthier, S., Goetz, C. G., Gomez-Tortosa, E., Halliday, G., Hansen, L. A., Hardy, J., Iwatsubo, T., Kalaria, R. N., Kaufer, D., Kenny, R. A., Korczyn, A., Kosaka, K., Lee, V. M., Lees, A., Litvan, I., Londos, E., Lopez, O. L., Minoshima, S., Mizuno, Y., Molina, J. A., Mukaetova-Ladinska, E. B., Pasquier, F., Perry, R. H., Schulz, J. B., Trojanowski, J. Q., Yamada, M. et Consortium on, D. L. B. (2005, Dec 27). Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology*, 65(12), 1863-1872. <https://doi.org/10.1212/01.wnl.0000187889.17253.b1>
- McKeith, I. G., Ferman, T. J., Thomas, A. J., Blanc, F., Boeve, B. F., Fujishiro, H., Kantarci, K., Muscio, C., O'Brien, J. T., Postuma, R. B., Aarsland, D., Ballard, C., Bonanni, L., Donaghy, P., Emre, M., Galvin, J. E., Galasko, D., Goldman, J. G., Gomperts, S. N., Honig, L. S., Ikeda, M., Leverenz, J. B., Lewis, S. J. G., Marder, K. S., Masellis, M., Salmon, D. P., Taylor, J. P., Tsuang, D. W., Walker, Z., Tiraboschi, P. et prodromal, D. L. B. D. S. G. (2020, Apr 28). Research criteria for the diagnosis of prodromal dementia with Lewy bodies. *Neurology*, 94(17), 743-755. <https://doi.org/10.1212/WNL.0000000000009323>
- McKenna, D. et Peever, J. (2017). Degeneration of rapid eye movement sleep circuitry underlies rapid eye movement sleep behavior disorder. *Movement Disorders*, 32(5), 636-644.
- Medicine, A. A. o. S. (2014). International classification of sleep disorders—third edition (ICSD-3). *AASM Resour Libr*, 281, 2313.
- Mei, J., Desrosiers, C. et Frasnelli, J. (2021). Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature. *Front Aging Neurosci*, 13, 633752. <https://doi.org/10.3389/fnagi.2021.633752>
- Miglis, M. G., Adler, C. H., Antelmi, E., Arnaldi, D., Baldelli, L., Boeve, B. F., Cesari, M., Dall'Antonia, I., Diederich, N. J., Doppler, K., Dušek, P., Ferri, R., Gagnon, J.-F., Gan-Or, Z., Hermann, W., Högl, B., Hu, M. T., Iranzo, A., Janzen, A., Kuzkina, A., Lee, J.-Y., Leenders, K. L., Lewis, S. J. G., Liguori, C., Liu, J., Lo, C., Ehgoetz Martens, K. A., Nepozitek, J., Plazzi, G., Provini, F., Puligheddu, M., Rolinski, M., Rusz, J., Stefani, A., Summers, R. L. S., Yoo, D., Zitser, J. et Oertel, W. H. (2021). Biomarkers of conversion to

- $\alpha$ -synucleinopathy in isolated rapid-eye-movement sleep behaviour disorder. *The Lancet Neurology*, 20(8), 671-684. [https://doi.org/10.1016/s1474-4422\(21\)00176-9](https://doi.org/10.1016/s1474-4422(21)00176-9)
- Mirzaei, G., Adeli, A. et Adeli, H. (2016, Dec 1). Imaging and machine learning techniques for diagnosis of Alzheimer's disease. *Rev Neurosci*, 27(8), 857-870. <https://doi.org/10.1515/revneuro-2016-0029>
- Mitchell, T. M. (1997). *Machine learning*. (vol. 45). McGraw Hill.
- Mohri, M., Rostamizadeh, A. et Talwalkar, A. (2012). Foundations of machine learning.[SI].
- Monchi, O., Petrides, M., Doyon, J., Postuma, R. B., Worsley, K. et Dagher, A. (2004, Jan 21). Neural bases of set-shifting deficits in Parkinson's disease. *J Neurosci*, 24(3), 702-710. <https://doi.org/10.1523/JNEUROSCI.4860-03.2004>
- Montplaisir, J., Gagnon, J. F., Fantini, M. L., Postuma, R. B., Dauvilliers, Y., Desautels, A., Rompré, S. et Paquet, J. (2010). Polysomnographic diagnosis of idiopathic REM sleep behavior disorder. *Movement Disorders*, 25(13), 2044-2051.
- Moore, S. F. et Barker, R. A. (2014, Jan). Predictors of Parkinson's disease dementia: towards targeted therapies for a heterogeneous disease. *Parkinsonism Relat Disord*, 20 Suppl 1, S104-107. [https://doi.org/10.1016/S1353-8020\(13\)70026-9](https://doi.org/10.1016/S1353-8020(13)70026-9)
- Morita, A., Kamei, S. et Mizutani, T. (2011, Aug). Relationship between slowing of the EEG and cognitive impairment in Parkinson disease. *J Clin Neurophysiol*, 28(4), 384-387. <https://doi.org/10.1097/WNP.0b013e3182273211>
- Morris, R. G., Downes, J. J., Sahakian, B. J., Evenden, J. L., Heald, A. et Robbins, T. W. (1988, Jun). Planning and spatial working memory in Parkinson's disease. *J Neurol Neurosurg Psychiatry*, 51(6), 757-766. <https://doi.org/10.1136/jnnp.51.6.757>
- Moruzzi, G. et Magoun, H. W. (1949). Brain stem reticular formation and activation of the EEG. *Electroencephalography and clinical neurophysiology*, 1(1-4), 455-473.
- Mostile, G., Nicoletti, A., Dibilio, V., Luca, A., Pappalardo, I., Giuliano, L., Cicero, C. E., Sciacca, G., Raciti, L., Contrafatto, D., Bruno, E., Sofia, V. et Zappia, M. (2015, Aug). Electroencephalographic lateralization, clinical correlates and pharmacological response in untreated Parkinson's disease. *Parkinsonism Relat Disord*, 21(8), 948-953. <https://doi.org/10.1016/j.parkreldis.2015.06.006>
- Muslimovic, D., Post, B., Speelman, J. D., De Haan, R. J. et Schmand, B. (2009, May). Cognitive decline in Parkinson's disease: a prospective longitudinal study. *J Int Neuropsychol Soc*, 15(3), 426-437. <https://doi.org/10.1017/S1355617709090614>
- Muslimovic, D., Post, B., Speelman, J. D. et Schmand, B. (2005, Oct 25). Cognitive profile of patients with newly diagnosed Parkinson disease. *Neurology*, 65(8), 1239-1245. <https://doi.org/10.1212/01.wnl.0000180516.69442.95>
- Myszczyńska, M. A., Ojamies, P. N., Lacoste, A. M. B., Neil, D., Saffari, A., Mead, R., Hautbergue, G. M., Holbrook, J. D. et Ferraiuolo, L. (2020, Aug). Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat Rev Neurol*, 16(8), 440-456. <https://doi.org/10.1038/s41582-020-0377-8>
- Naci, H. et Ioannidis, J. P. (2015). How good is "evidence" from clinical studies of drug effects and why might such evidence fail in the prediction of the clinical utility of drugs? *Annu Rev Pharmacol Toxicol*, 55, 169-189. <https://doi.org/10.1146/annurev-pharmtox-010814-124614>
- Naismith, S. L., Pereira, M., Shine, J. M. et Lewis, S. J. (2011, Jan). How well do caregivers detect mild cognitive change in Parkinson's disease? *Mov Disord*, 26(1), 161-164. <https://doi.org/10.1002/mds.23331>

- Neufeld, M. Y., Blumen, S., Aitkin, I., Parmet, Y. et Korczyn, A. D. (1994, Jan-Feb). EEG frequency analysis in demented and nondemented parkinsonian patients. *Dementia*, 5(1), 23-28. <https://doi.org/10.1159/000106690>
- Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press.
- Ohayon, M. M. et Schenck, C. H. (2010). Violent behavior during sleep: prevalence, comorbidity and consequences. *Sleep medicine*, 11(9), 941-946.
- Oikonomou, P., van Wamelen, D. J., Weintraub, D., Aarsland, D., Ffytche, D., Martinez-Martin, P., Rodriguez-Blazquez, C., Leta, V., Borley, C., Sportelli, C., Trivedi, D., Podlowska, A. M., Rukavina, K., Rizos, A., Lazcano-Ocampo, C. et Ray Chaudhuri, K. (2021, May). Nonmotor symptom burden grading as predictor of cognitive impairment in Parkinson's disease. *Brain Behav*, 11(5), e02086. <https://doi.org/10.1002/brb3.2086>
- Ortelli, P., Ferrazzoli, D., Versace, V., Cian, V., Zarucchi, M., Gusmeroli, A., Canesi, M., Frazzitta, G., Volpe, D., Ricciardi, L., Nardone, R., Ruffini, I., Saltuari, L., Sebastianelli, L., Baranzini, D. et Maestri, R. (2022, Apr 11). Optimization of cognitive assessment in Parkinsonisms by applying artificial intelligence to a comprehensive screening test. *NPJ Parkinsons Dis*, 8(1), 42. <https://doi.org/10.1038/s41531-022-00304-z>
- Owen, A. M. (2004, Dec). Cognitive dysfunction in Parkinson's disease: the role of frontostriatal circuitry. *Neuroscientist*, 10(6), 525-537. <https://doi.org/10.1177/1073858404266776>
- Palmqvist, S., Hertze, J., Minthon, L., Wattmo, C., Zetterberg, H., Blennow, K., Londos, E. et Hansson, O. (2012). Comparison of brief cognitive tests and CSF biomarkers in predicting Alzheimer's disease in mild cognitive impairment: six-year follow-up study. *PloS one*, 7(6), e38639. <https://doi.org/10.1371/journal.pone.0038639>
- Parkkinen, L., Pirttilä, T. et Alafuzoff, I. (2008). Applicability of current staging/categorization of  $\alpha$ -synuclein pathology and their clinical relevance. *Acta neuropathologica*, 115, 399-407.
- Pedersen, K. F., Larsen, J. P., Tysnes, O. B. et Alves, G. (2013, May). Prognosis of mild cognitive impairment in early Parkinson disease: the Norwegian ParkWest study. *JAMA Neurol*, 70(5), 580-586. <https://doi.org/10.1001/jamaneurol.2013.2110>
- Peever, J., Luppi, P.-H. et Montplaisir, J. (2014). Breakdown in REM sleep circuitry underlies REM sleep behavior disorder. *Trends in neurosciences*, 37(5), 279-288.
- Petrova, M., Mehrabian-Spasova, S., Aarsland, D., Raycheva, M. et Traykov, L. (2015, May-Aug). Clinical and Neuropsychological Differences between Mild Parkinson's Disease Dementia and Dementia with Lewy Bodies. *Dement Geriatr Cogn Dis Extra*, 5(2), 212-220. <https://doi.org/10.1159/000375363>
- Pezard, L., Jech, R. et Ruzicka, E. (2001, Jan). Investigation of non-linear properties of multichannel EEG in the early stages of Parkinson's disease. *Clin Neurophysiol*, 112(1), 38-45. [https://doi.org/10.1016/s1388-2457\(00\)00512-5](https://doi.org/10.1016/s1388-2457(00)00512-5)
- Pievani, M., de Haan, W., Wu, T., Seeley, W. W. et Frisoni, G. B. (2011, Sep). Functional network disruption in the degenerative dementias. *Lancet Neurol*, 10(9), 829-843. [https://doi.org/10.1016/S1474-4422\(11\)70158-2](https://doi.org/10.1016/S1474-4422(11)70158-2)
- Pillon, B., Ertle, S., Deweer, B., Bonnet, A. M., Vidailhet, M. et Dubois, B. (1997, Mar). Memory for spatial location in 'de novo' parkinsonian patients. *Neuropsychologia*, 35(3), 221-228. [https://doi.org/10.1016/s0028-3932\(96\)00091-7](https://doi.org/10.1016/s0028-3932(96)00091-7)
- Plomhause, L., Dujardin, K., Boucart, M., Herlin, V., Defebvre, L., Derambure, P. et Monaca Charley, C. (2014, May). Impaired visual perception in rapid eye movement sleep behavior disorder. *Neuropsychology*, 28(3), 388-393. <https://doi.org/10.1037/neu0000006>



- Poil, S. S., de Haan, W., van der Flier, W. M., Mansvelder, H. D., Scheltens, P. et Linkenkaer-Hansen, K. (2013). Integrative EEG biomarkers predict progression to Alzheimer's disease at the MCI stage. *Front Aging Neurosci*, 5, 58. <https://doi.org/10.3389/fnagi.2013.00058>
- Postuma, R., Gagnon, J., Vendette, M., Fantini, M., Massicotte-Marquez, J. et Montplaisir, J. (2009). Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder. *Neurology*, 72(15), 1296-1300.
- Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A. E., Halliday, G., Goetz, C. G., Gasser, T., Dubois, B., Chan, P., Bloem, B. R., Adler, C. H. et Deuschl, G. (2015, Oct). MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord*, 30(12), 1591-1601. <https://doi.org/10.1002/mds.26424>
- Postuma, R. B., Iranzo, A., Hu, M., Hogl, B., Boeve, B. F., Manni, R., Oertel, W. H., Arnulf, I., Ferini-Strambi, L., Puligheddu, M., Antelmi, E., Cochen De Cock, V., Arnaldi, D., Mollenhauer, B., Videnovic, A., Sonka, K., Jung, K. Y., Kunz, D., Dauvilliers, Y., Provini, F., Lewis, S. J., Buskova, J., Pavlova, M., Heidebreder, A., Montplaisir, J. Y., Santamaria, J., Barber, T. R., Stefani, A., St Louis, E. K., Terzaghi, M., Janzen, A., Leu-Semenescu, S., Plazzi, G., Nobili, F., Sixel-Doering, F., Dusek, P., Bes, F., Cortelli, P., Ehgoetz Martens, K., Gagnon, J. F., Gaig, C., Zucconi, M., Trenkwalder, C., Gan-Or, Z., Lo, C., Rolinski, M., Mahlknecht, P., Holzknacht, E., Boeve, A. R., Teigen, L. N., Toscano, G., Mayer, G., Morbelli, S., Dawson, B. et Pelletier, A. (2019, Mar 1). Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study. *Brain*, 142(3), 744-759. <https://doi.org/10.1093/brain/awz030>
- Pozzi, D., Petracchi, M., Sabe, L., Dancygier, G., Garcia, H. et Starkstein, S. (1994, Nov). Quantified electroencephalographic changes in Parkinson's disease with and without dementia. *Eur J Neurol*, 1(2), 147-152. <https://doi.org/10.1111/j.1468-1331.1994.tb00063.x>
- Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W., LaMantia, A.-S. et White, L. (2019). *Neurosciences*. De Boeck Supérieur.
- Rahayel, S., Postuma, R. B., Montplaisir, J., Marchand, D. G., Escudier, F., Gaubert, M., Bourgouin, P.-A., Carrier, J., Monchi, O. et Joubert, S. (2018). Cortical and subcortical gray matter bases of cognitive deficits in REM sleep behavior disorder. *Neurology*, 90(20), e1759-e1770.
- Rahayel, S., Postuma, R. B., Montplaisir, J., Mišić, B., Tremblay, C., Vo, A., Lewis, S., Matar, E., Ehgoetz Martens, K. et Blanc, F. (2021). A prodromal brain-clinical pattern of cognition in Synucleinopathies. *Annals of neurology*, 89(2), 341-357.
- Rahman, M. M., Usman, O. L., Muniyandi, R. C., Sahran, S., Mohamed, S. et Razak, R. A. (2020, Dec 7). A Review of Machine Learning Methods of Feature Selection and Classification for Autism Spectrum Disorder. *Brain Sci*, 10(12). <https://doi.org/10.3390/brainsci10120949>
- Rana, A., Dumka, A., Singh, R., Panda, M. K., Priyadarshi, N. et Twala, B. (2022, Aug 19). Imperative Role of Machine Learning Algorithm for Detection of Parkinson's Disease: Review, Challenges and Recommendations. *Diagnostics (Basel)*, 12(8). <https://doi.org/10.3390/diagnostics12082003>
- Remillard-Pelchat, D., Rahayel, S., Gaubert, M., Postuma, R. B., Montplaisir, J., Pelletier, A., Monchi, O., Brambati, S. M., Carrier, J. et Gagnon, J. F. (2022). Comprehensive Analysis of Brain Volume in REM Sleep Behavior Disorder with Mild Cognitive Impairment. *J Parkinsons Dis*, 12(1), 229-241. <https://doi.org/10.3233/JPD-212691>

- Rodrigues Brazete, J., Gagnon, J. F., Postuma, R. B., Bertrand, J. A., Petit, D. et Montplaisir, J. (2016, Jan). Electroencephalogram slowing predicts neurodegeneration in rapid eye movement sleep behavior disorder. *Neurobiol Aging*, 37, 74-81. <https://doi.org/10.1016/j.neurobiolaging.2015.10.007>
- Rodrigues Brazete, J., Montplaisir, J., Petit, D., Postuma, R. B., Bertrand, J. A., Genier Marchand, D. et Gagnon, J. F. (2013, Nov). Electroencephalogram slowing in rapid eye movement sleep behavior disorder is associated with mild cognitive impairment. *Sleep Med*, 14(11), 1059-1063. <https://doi.org/10.1016/j.sleep.2013.06.013>
- Rossini, P. M., Rossi, S., Babiloni, C. et Polich, J. (2007, Dec). Clinical neurophysiology of aging brain: from normal aging to neurodegeneration. *Prog Neurobiol*, 83(6), 375-400. <https://doi.org/10.1016/j.pneurobio.2007.07.010>
- Rouvière, L. (2010). *Introduction aux méthodes d'agrégation: boosting, bagging et forêts aléatoires. Illustrations avec R.*
- Ruffini, G., Ibanez, D., Castellano, M., Dubreuil-Vall, L., Soria-Frisch, A., Postuma, R., Gagnon, J. F. et Montplaisir, J. (2019). Deep Learning With EEG Spectrograms in Rapid Eye Movement Behavior Disorder. *Front Neurol*, 10, 806. <https://doi.org/10.3389/fneur.2019.00806>
- Rundo, J. V., & Downey III, R. Polysomnography. Dans *Handbook of clinical neurology* (vol. 160). <https://doi.org/https://doi.org/10.1016/B978-0-444-64032-1.00025-4>
- Sadiq, D., Whitfield, T., Lee, L., Stevens, T., Costafreda, S. et Walker, Z. (2017). Prodromal Dementia with Lewy Bodies and Prodromal Alzheimer's Disease: A Comparison of the Cognitive and Clinical Profiles. *J Alzheimers Dis*, 58(2), 463-470. <https://doi.org/10.3233/JAD-161089>
- Sasai, T., Matsuura, M. et Inoue, Y. (2013, Dec 1). Electroencephalographic findings related with mild cognitive impairment in idiopathic rapid eye movement sleep behavior disorder. *Sleep*, 36(12), 1893-1899. <https://doi.org/10.5665/sleep.3224>
- Sasai, T., Miyamoto, T., Miyamoto, M., Iwanami, M., Abe, T., Matsuura, M. et Inoue, Y. (2012, Mar). Impaired decision-making in idiopathic REM sleep behavior disorder. *Sleep Med*, 13(3), 301-306. <https://doi.org/10.1016/j.sleep.2011.09.012>
- Sateia, M. J. (2014). International classification of sleep disorders. *Chest*, 146(5), 1387-1394.
- Schenck, C. H., Garcia-Rill, E., Skinner, R. D., Anderson, M. L. et Mahowald, M. W. (1996, Sep 1). A case of REM sleep behavior disorder with autopsy-confirmed Alzheimer's disease: postmortem brain stem histochemical analyses. *Biol Psychiatry*, 40(5), 422-425. [https://doi.org/10.1016/0006-3223\(96\)00070-4](https://doi.org/10.1016/0006-3223(96)00070-4)
- Sharma, S. et Mandal, P. K. (2022). A Comprehensive Report on Machine Learning-based Early Detection of Alzheimer's Disease using Multi-modal Neuroimaging Data. *ACM Computing Surveys*, 55(2), 1-44. <https://doi.org/10.1145/3492865>
- Sherin, J. E., Elmquist, J. K., Torrealba, F. et Saper, C. B. (1998). Innervation of histaminergic tuberomammillary neurons by GABAergic and galaninergic neurons in the ventrolateral preoptic nucleus of the rat. *Journal of Neuroscience*, 18(12), 4705-4721.
- Siegel, J. (2011). Sleep mechanisms and phylogeny. *Principles and practice of sleep medicine*, 76-126.
- Sinforiani, E., Zangaglia, R., Manni, R., Cristina, S., Marchioni, E., Nappi, G., Mancini, F. et Pacchetti, C. (2006, Apr). REM sleep behavior disorder, hallucinations, and cognitive impairment in Parkinson's disease. *Mov Disord*, 21(4), 462-466. <https://doi.org/10.1002/mds.20719>

- Singh, A., Thakur, N., & Sharma, A. . (2016, March). A review of supervised machine learning algorithms. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom).Ieee.* , 1310-1315.
- Sleekier, T., Inglis, W., Winn, P. et Sahgal, A. (1994). The pedunclopontine tegmental nucleus: a role in cognitive processes? *Brain Research Reviews*, *19*(3), 298-318.
- Soikkeli, R., Partanen, J., Soininen, H., Pääkkönen, A. et Riekkinen, P. (1991). Slowing of EEG in Parkinson's disease. *Electroencephalography and clinical neurophysiology*, *79*(3), 159-165. [https://doi.org/10.1016/0013-4694\(91\)90134-p](https://doi.org/10.1016/0013-4694(91)90134-p)
- St Louis, E. K., Boeve, A. R. et Boeve, B. F. (2017). REM Sleep Behavior Disorder in Parkinson's Disease and Other Synucleinopathies. *Movement Disorders*, *32*(5), 645-658.
- Stanzione, P., Marciani, M. G., Maschio, M., Bassetti, M. A., Spanedda, F., Pierantozzi, M., Semprini, R. et Bernardi, G. (1996). Quantitative EEG changes in non-demented Parkinson's disease patients before and during L-dopa therapy. *European Journal of Neurology*, *3*(4), 354-362. <https://doi.org/10.1111/j.1468-1331.1996.tb00229.x>
- Stebbins, G. T., Gabrieli, J. D., Masciari, F., Monti, L. et Goetz, C. G. (1999, Apr). Delayed recognition memory in Parkinson's disease: a role for working memory? *Neuropsychologia*, *37*(4), 503-510. [https://doi.org/10.1016/s0028-3932\(98\)00068-2](https://doi.org/10.1016/s0028-3932(98)00068-2)
- Steriade, M., Domich, L., Oakson, G. et Deschenes, M. (1987). The deafferented reticular thalamic nucleus generates spindle rhythmicity. *Journal of neurophysiology*, *57*(1), 260-273.
- Stern, Y., Marder, K., Tang, M. X. et Mayeux, R. (1993, Sep). Antecedent clinical features associated with dementia in Parkinson's disease. *Neurology*, *43*(9), 1690-1692. <https://doi.org/10.1212/wnl.43.9.1690>
- Steyerberg, E. W. et Harrell, F. E., Jr. (2016, Jan). Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*, *69*, 245-247. <https://doi.org/10.1016/j.jclinepi.2015.04.005>
- Stoffers, D., Bosboom, J. L., Deijen, J. B., Wolters, E. C., Berendse, H. W. et Stam, C. J. (2007, Jul). Slowing of oscillatory brain activity is a stable characteristic of Parkinson's disease without dementia. *Brain*, *130*(Pt 7), 1847-1860. <https://doi.org/10.1093/brain/awm034>
- Stylianou, M., Murphy, N., Peraza, L. R., Graziadio, S., Cromarty, R., Killen, A., JT, O. B., Thomas, A. J., LeBeau, F. E. N. et Taylor, J. P. (2018, Jun). Quantitative electroencephalography as a marker of cognitive fluctuations in dementia with Lewy bodies and an aid to differential diagnosis. *Clin Neurophysiol*, *129*(6), 1209-1220. <https://doi.org/10.1016/j.clinph.2018.03.013>
- Sunwoo, J. S., Lee, S., Kim, J. H., Lim, J. A., Kim, T. J., Byun, J. I., Jeong, M. H., Cha, K. S., Choi, J. W., Kim, K. H., Lee, S. T., Jung, K. H., Park, K. I., Chu, K., Kim, M., Lee, S. K. et Jung, K. Y. (2017, Jun 1). Altered Functional Connectivity in Idiopathic Rapid Eye Movement Sleep Behavior Disorder: A Resting-State EEG Study. *Sleep*, *40*(6). <https://doi.org/10.1093/sleep/zsx058>
- Szymusiak, R., Alam, N., Steininger, T. L. et McGinty, D. (1998). Sleep-waking discharge patterns of ventrolateral preoptic/anterior hypothalamic neurons in rats. *Brain research*, *803*(1-2), 178-188.
- Tajiri, Y., Wada-Isoe, K., Tanaka, K., Adachi, T., Hanajima, R. et Nakashima, K. (2020, Feb). A Single-institution Study on Predictors of Short-term Progression from Mild Cognitive Impairment in Parkinson's Disease to Parkinson's Disease with Dementia. *Yonago Acta Med*, *63*(1), 28-33. <https://doi.org/10.33160/yam.2020.02.004>

- Takahashi, K., Lin, J.-S. et Sakai, K. (2009). Characterization and mapping of sleep-waking specific neurons in the basal forebrain and preoptic hypothalamus in mice. *Neuroscience*, 161(1), 269-292.
- Tanaka, H., Koenig, T., Pascual-Marqui, R. D., Hirata, K., Kochi, K. et Lehmann, D. (2000, Jan-Feb). Event-related potential and EEG measures in Parkinson's disease without and with dementia. *Dement Geriatr Cogn Disord*, 11(1), 39-45. <https://doi.org/10.1159/000017212>
- Tautan, A. M., Ionescu, B. et Santarnecchi, E. (2021, Jul). Artificial intelligence in neurodegenerative diseases: A review of available tools with a focus on machine learning techniques. *Artif Intell Med*, 117, 102081. <https://doi.org/10.1016/j.artmed.2021.102081>
- Terzaghi, M., Sinforiani, E., Zucchella, C., Zambrelli, E., Pasotti, C., Rustioni, V. et Manni, R. (2008, May). Cognitive performance in REM sleep behaviour disorder: a possible early marker of neurodegenerative disease? *Sleep Med*, 9(4), 343-351. <https://doi.org/10.1016/j.sleep.2007.06.013>
- Terzaghi, M., Toscano, G., Casoni, F., Picascia, M., Arnaldi, D., Rustioni, V., Versino, M., Sinforiani, E. et Manni, R. (2019). Assessment of cognitive profile as a prodromal marker of the evolution of rapid eye movement sleep behavior disorder. *Sleep*, 42(8), zsz103.
- Terzaghi, M., Zucchella, C., Rustioni, V., Sinforiani, E. et Manni, R. (2013, Oct 1). Cognitive performances and mild cognitive impairment in idiopathic rapid eye movement sleep behavior disorder: results of a longitudinal follow-up study. *Sleep*, 36(10), 1527-1532. <https://doi.org/10.5665/sleep.3050>
- Tholke, P., Mantilla-Ramos, Y. J., Abdelhedi, H., Maschke, C., Dehgan, A., Harel, Y., Kemtur, A., Mekki Berrada, L., Sahraoui, M., Young, T., Bellemare Pepin, A., El Khantour, C., Landry, M., Pascarella, A., Hadid, V., Combrisson, E., O'Byrne, J. et Jerbi, K. (2023, Jun 28). Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *Neuroimage*, 277, 120253. <https://doi.org/10.1016/j.neuroimage.2023.120253>
- Trambaiolli, L. R., Lorena, A. C., Fraga, F. J., Kanda, P. A., Anghinah, R. et Nitrini, R. (2011, Jul). Improving Alzheimer's disease diagnosis with machine learning techniques. *Clin EEG Neurosci*, 42(3), 160-165. <https://doi.org/10.1177/155005941104200304>
- Troster, A. I. (2008, Mar). Neuropsychological characteristics of dementia with Lewy bodies and Parkinson's disease with dementia: differentiation, early detection, and implications for "mild cognitive impairment" and biomarkers. *Neuropsychol Rev*, 18(1), 103-119. <https://doi.org/10.1007/s11065-008-9055-0>
- Tsiouris, K. M., Rigas, G., Gatsios, D., Antonini, A., Konitsiotis, S., Koutsouris, D. D. et Fotiadis, D. I. (2017, Jul). Predicting rapid progression of Parkinson's Disease at baseline patients evaluation. *Annu Int Conf IEEE Eng Med Biol Soc*, 2017, 3898-3901. <https://doi.org/10.1109/EMBC.2017.8037708>
- Turner, R., D'amato, C., Chervin, R. et Blaivas, M. (2000). The pathology of REM sleep behavior disorder with comorbid Lewy body dementia. *Neurology*, 55(11), 1730-1732.
- Uchiyama, M., Isse, K., Tanaka, K., Yokota, N., Hamamoto, M., Aida, S., Ito, Y., Yoshimura, M. et Okawa, M. (1995). Incidental Lewy body disease in a patient with REM sleep behavior disorder. *Neurology*, 45(4), 709-712.
- Utianski, R. L., Caviness, J. N., van Straaten, E. C., Beach, T. G., Dugger, B. N., Shill, H. A., ... & Hentz, J. G. (2014). Graph theory network function in Parkinson's disease assessed with electroencephalography. *Neuroreport*, 25(16), 1266.
- Vemuri, V. K. (2020). *The Hundred-Page Machine Learning Book: by Andriy Burkov*.



- Vendette, M., Gagnon, J. F., Decary, A., Massicotte-Marquez, J., Postuma, R. B., Doyon, J., Panisset, M. et Montplaisir, J. (2007, Nov 6). REM sleep behavior disorder predicts cognitive impairment in Parkinson disease without dementia. *Neurology*, 69(19), 1843-1849. <https://doi.org/10.1212/01.wnl.0000278114.14096.74>
- Vendette, M., Montplaisir, J., Gosselin, N., Soucy, J. P., Postuma, R. B., Dang-Vu, T. T. et Gagnon, J. F. (2012). Brain perfusion anomalies in rapid eye movement sleep behavior disorder with mild cognitive impairment. *Movement Disorders*, 27(10), 1255-1261.
- Venugopalan, J., Tong, L., Hassanzadeh, H. R. et Wang, M. D. (2021, Feb 5). Multimodal deep learning models for early detection of Alzheimer's disease stage. *Sci Rep*, 11(1), 3254. <https://doi.org/10.1038/s41598-020-74399-w>
- Wasserstein, R. L., and Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. *Am. Stat*, 70, 129–133. <https://doi.org/10.1080/00031305.2016.1154108>
- Weisman, D. et McKeith, I. (2007, Feb). Dementia with Lewy bodies. *Semin Neurol*, 27(1), 42-47. <https://doi.org/10.1055/s-2006-956754>
- White, I. R., Daniel, R. et Royston, P. (2010, Oct 1). Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. *Comput Stat Data Anal*, 54(10), 2267-2275. <https://doi.org/10.1016/j.csda.2010.04.005>
- Williams-Gray, C. H., Evans, J. R., Goris, A., Foltynie, T., Ban, M., Robbins, T. W., Brayne, C., Kolachana, B. S., Weinberger, D. R., Sawcer, S. J. et Barker, R. A. (2009, Nov). The distinct cognitive syndromes of Parkinson's disease: 5 year follow-up of the CamPaIGN cohort. *Brain*, 132(Pt 11), 2958-2969. <https://doi.org/10.1093/brain/awp245>
- Williams-Gray, C. H., Foltynie, T., Brayne, C. E., Robbins, T. W. et Barker, R. A. (2007, Jul). Evolution of cognitive dysfunction in an incident Parkinson's disease cohort. *Brain*, 130(Pt 7), 1787-1798. <https://doi.org/10.1093/brain/awm111>
- Williams-Gray, C. H., Mason, S. L., Evans, J. R., Foltynie, T., Brayne, C., Robbins, T. W. et Barker, R. A. (2013, Nov). The CamPaIGN study of Parkinson's disease: 10-year outlook in an incident population-based cohort. *J Neurol Neurosurg Psychiatry*, 84(11), 1258-1264. <https://doi.org/10.1136/jnnp-2013-305277>
- Winn, P. (1998). Frontal syndrome as a consequence of lesions in the pedunculopontine tegmental nucleus: a short theoretical review. *Brain Research Bulletin*, 47(6), 551-563.
- Xanthopoulos, P., Pardalos, P. M., & Trafalis, T. B. (2013). *Linear discriminant analysis*. Springer.
- Yilmaz, N. H., Calisoglu, P., Guntekin, B. et Hanoglu, L. (2020, Nov 1). Correlation between alpha activity and neuropsychometric tests in Parkinson's disease. *Neurosci Lett*, 738, 135346. <https://doi.org/10.1016/j.neulet.2020.135346>
- Youn, S., Kim, T., Yoon, I. Y., Jeong, J., Kim, H. Y., Han, J. W., Kim, J. M. et Kim, K. W. (2016, Aug). Progression of cognitive impairments in idiopathic REM sleep behaviour disorder. *J Neurol Neurosurg Psychiatry*, 87(8), 890-896. <https://doi.org/10.1136/jnnp-2015-311437>
- Yu, R. L., Lee, W. J., Li, J. Y., Chang, Y. Y., Chen, C. C., Lin, J. J., Sung, Y. F., Lin, T. K. et Fuh, J. L. (2020, Jan 23). Evaluating Mild Cognitive Dysfunction in Patients with Parkinson's Disease in Clinical Practice in Taiwan. *Sci Rep*, 10(1), 1014. <https://doi.org/10.1038/s41598-020-58042-2>
- Yuvaraj, R., Rajendra Acharya, U. et Hagiwara, Y. (2016). A novel Parkinson's Disease Diagnosis Index using higher-order spectra features in EEG signals. *Neural Computing and Applications*, 30(4), 1225-1235. <https://doi.org/10.1007/s00521-016-2756-z>
- Zgaljardic, D. J., Borod, J. C., Foldi, N. S. et Mattis, P. (2003). A review of the cognitive and behavioral sequelae of Parkinson's disease: relationship to frontostriatal circuitry. *Cognitive and behavioral neurology*, 16(4), 193-210.

- Zhang, J.-R., Chen, J., Yang, Z.-J., Zhang, H.-J., Fu, Y.-T., Shen, Y., He, P.-C., Mao, C.-J. et Liu, C.-F. (2016). Rapid eye movement sleep behavior disorder symptoms correlate with domains of cognitive impairment in Parkinson's disease. *Chinese Medical Journal*, 129(04), 379-385, C371.
- Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, 237, 350-361.  
<https://doi.org/https://doi.org/10.1016/j.neucom.2017.01.026>