Journal of the Neurological Sciences 462 (2024) 123091

Contents lists available at ScienceDirect

Journal of the Neurological Sciences

journal homepage: www.elsevier.com/locate/jns

Unraveling sex differences in Parkinson's disease through explainable machine learning

Gianfrancesco Angelini [a], Antonio Malvaso [b,c], Aurelia Schirripa [c], Francesca Campione [c], Sebastian Luca D'Addario [c,d], Nicola Toschi [a,e], Daniele Caligiore [c,f,*]

[a] Medical Physics Section, Department of Biomedicine and Prevention, University of Rome Tor Vergata, Via Montpellier, 1, 00133 Rome, Italy
[b] Department of Brain and Behavioral Sciences, IRCCS Mondino Foundation, National Neurological Institute, University of Pavia, Via Mondino 2, 27100 Pavia, Italy
[c] Computational and Translational Neuroscience Laboratory, Institute of Cognitive Sciences and Technologies, National Research Council (CTNLab-ISTC-CNR), Via Gian Domenico Romagnosi, 18A, 00196 Rome, Italy
[d] IRCCS Fondazione Santa Lucia, Via Del Fosso di Fiorano, 64, 00143 Rome, Italy
[e] Athinoula A. Martinos Center for Biomedical Imaging and Harvard Medical School, Boston, MA, USA
[f] AI2Life s.r.l., Innovative Start-Up, ISTC-CNR Spin-Off, Via Sebino 32, 00199 Rome, Italy

ARTICLE INFO

Keywords:
Parkinson's disease
Gender differences
Heterogeneous data
Non-motor symptoms
Features importance
Explainable machine learning

ABSTRACT

Sex differences affect Parkinson's disease (PD) development and manifestation. Yet, current PD identification and treatments underuse these distinctions. Sex-focused PD literature often prioritizes prevalence rates over feature importance analysis. However, underlying aspects could make a feature significant for predicting PD, despite its score. Interactions between features require consideration, as do distinctions between scoring disparities and actual feature importance. For instance, a higher score in males for a certain feature doesn't necessarily mean it's less important for characterizing PD in females. This article proposes an explainable Machine Learning (ML) model to elucidate these underlying factors, emphasizing the importance of features. This insight could be critical for personalized medicine, suggesting the need to tailor data collection and analysis for males and females. The model identifies sex-specific differences in PD, aiding in predicting outcomes as "Healthy" or "Pathological". It adopts a system-level approach, integrating heterogeneous data - clinical, imaging, genetics, and demographics - to study new biomarkers for diagnosis. The explainable ML approach aids non-ML experts in understanding model decisions, fostering trust and facilitating interpretation of complex ML outcomes, thus enhancing usability and translational research. The ML model identifies muscle rigidity, autonomic and cognitive assessments, and family history as key contributors to PD diagnosis, with sex differences noted. The genetic variant SNCA-rs356181 may be more significant in characterizing PD in males. Interaction analysis reveals a greater occurrence of feature interplay among males compared to females. These disparities offer insights into PD pathophysiology and could guide the development of sex-specific diagnostic and therapeutic approaches.

1. Introduction

Parkinson's disease (PD) is a notable neurodegenerative disorder characterized by motor symptoms including tremors, rigidity, and bradykinesia [1,2,3]. Additionally, PD manifests non-motor symptoms such as cognitive changes, depression, anxiety, sleep disturbances, autonomic dysfunction, psychosis, and impulse control disorders [4,5]. The impact of PD is significant due to its high prevalence and the resulting strain on individuals, families, and healthcare systems. It affects around 1 % of people aged 60 and above, making it among the most prevalent neurodegenerative disorders [6]. With the world population aging, PD rates are projected to rise, creating considerable challenges for healthcare systems globally.

In recent years, comprehension of PD has advanced beyond viewing it solely as a dysfunction of the basal ganglia. Several works acknowledge PD as a disorder involving the intricate interplay within the basal ganglia-cortex-cerebellum system, influenced by various factors such as brain-body interactions, lifestyle choices, and environmental conditions

* Corresponding author at: Computational and Translational Neuroscience Laboratory, Institute of Cognitive Sciences and Technologies, National Research Council (CTNLab-ISTC-CNR), Via Gian Domenico Romagnosi, 18A, 00196 Rome, Italy.
E-mail address: daniele.caligiore@istc.cnr.it (D. Caligiore).

https://doi.org/10.1016/j.jns.2024.123091
Received 26 March 2024; Received in revised form 24 May 2024; Accepted 6 June 2024
Available online 8 June 2024

[7,8,9,10,11,12]. This *system-level* perspective underscores the significance of investigating PD subtypes based on genetic, environmental, and sex disparities, as these elements can significantly influence the manifestation and progression of the disease [13,14].

Recent evidence suggests significant sex-based differences in PD manifestation and progression, highlighting the necessity for personalized, sex-specific approaches to both diagnosis and treatment [15,16,14,17,18,19]. Theoretical explanations for sex differences in PD manifestation include hormonal influences, genetic susceptibility, and differences in brain structure and function [20,21,22]. Estrogen, for example, has been shown to have neuroprotective effects, which may contribute to the lower incidence of PD in premenopausal females. Additionally, genetic studies have identified sex-specific risk factors for PD [23,24,25].

Despite the growing recognition of sex-based differences in PD, there remains a significant gap in the literature regarding the underlying mechanisms by which sex influences disease manifestation. Understanding these mechanisms is critical for developing personalized, sex-specific approaches to both diagnosis and treatment. In addition, sex-focused PD literature often prioritizes prevalence rates over the analysis of feature importance. Nevertheless, *underlying factors may still render a feature significant in predicting PD, regardless of its prevalence rate*. It should be considered how features interact, along with the differences between scoring and the actual importance of features. For example, a higher score in males for a feature does not necessarily diminish its importance in characterizing PD in females. Typically, the literature exploring sex disparities in PD correlates variations in test scores between males and females with their respective significance in predicting the likelihood of developing the disease. This article shows that explainable Machine Learning (ML) approaches could reveal that this implicit correlation is not always accurate. For instance, consider a scenario where a high score on a particular test implies an increased risk of PD. In males, a higher score on this test does not necessarily indicate greater importance of the test as a marker for the disease compared to females. It is plausible that in females, achieving a score slightly lower than what males typically achieve still signifies a heightened risk of developing PD. This discrepancy could arise due to various factors, such as different baseline levels of the feature being measured or intricate relationships with other unmeasured features. Consequently, the feature assessed by the test may hold significant or even heightened importance for females despite their scores being comparatively lower than those of males. Furthermore, despite some efforts utilizing ML to investigate sex differences in PD [26,27,28], there remains a significant gap in research addressing the most important heterogeneous features for characterizing PD in females and males.

This article pioneers the investigation of sex-specific features in PD by employing a two-fold approach. First, it adopts a *system-level analysis*, integrating and studying the *interactions among heterogeneous data* drawn from the Parkinson's Progression Markers Initiative dataset, encompassing demographic, genetic, biological, and clinical information related to both motor and non-motor features. In this way, it could be possible to better understand the complexity of PD and potentially identify novel biomarkers for diagnosis and prognosis [7,13]. Second, it utilizes an *explainable ML* approach to uncover patterns and relationships within the data, shedding light on the distinct features defining PD in females and males. By combining these methodologies, this study aims to advance our understanding of PD heterogeneity and pave the way for more tailored diagnostic and therapeutic interventions. Ultimately, elucidating sex-specific aspects of PD will contribute to optimizing symptom management and enhancing the quality of life for individuals living with this debilitating condition.

As previously demonstrated in the context of medical research, explainable ML approaches enable researchers to gain a deeper understanding of complex phenomena in neurological diseases [29,30]. An explainable ML algorithm provides clear and understandable explanations about how it reached a result. This transparency helps humans, especially non-ML experts, grasp the reasoning behind the model decisions, fostering trust and facilitating interpretation of complex ML outcomes.

Employing advanced analytical techniques, particularly leveraging SHapley Additive exPlanations (SHAP) beeswarm plots, the ML model uncovers the top nine features, over 59 initially examined, that exert the most significant impact on PD diagnosis, with a specific focus on sex-based disparities. The ML model highlights muscle rigidity, autonomic symptoms, cognitive assessments, and family history as crucial factors for PD prediction, acknowledging the presence of sex differences. The genetic variant SNCA-rs356181 appears to play a more significant role in characterizing PD in males. Furthermore, interaction analysis indicates a higher frequency of feature interplays among males as opposed to females.

Our research presents a sex-based diagnostic model for PD, unveiling substantial differences between healthy controls and PD classification based on sex. Furthermore, it deepens comprehension of sex-specific PD characteristics, offering insights into early disease detection and bolstering ML-based diagnostic approaches.

2. Materials and methods

2.1. PPMI dataset

This paper utilizes data sourced from the Parkinson's Progression Markers Initiative (PPMI) database,[1] a resource for both PD researchers and healthcare professionals. The dataset contains a wide range of information, including genetics, neuroimaging scans, cerebrospinal fluid samples, and clinical assessments, collected from individuals diagnosed with idiopathic and genetic PD, as well as data from healthy controls. It serves as a repository for studying the natural progression of PD, identifying biomarkers, and enhancing diagnostic and therapeutic approaches. The original identifiers from the PPMI dataset have been preserved for the features discussed in this paper.

2.2. Cohort chosen for the study

Variables without relevant information were excluded from the PPMI dataset to prevent unnecessary noise and maintain the clarity of the analysis. Furthermore, columns deemed unnecessary were omitted to streamline the dataset and focus our investigation on key factors directly related to our research objectives. Variables not yet recorded in the literature were also excluded to maintain alignment with established scientific knowledge and to uphold the validity of our findings. This selection process aimed to prioritize variables with demonstrated relevance to our research questions, thereby enhancing the robustness and interpretability of our analyses. This selection process resulted in a dataset composed of 59 variables, and their description is summarized in Table 1.

2.3. Data pre-processing

The Categorical Boosting (CatBoost) algorithm was used to handle missing values in the dataset, leveraging its inherent capability to manage them without requiring imputation. This functionality enables users to work with the data in its original form (refer to Section 2.4.1). The problem has been framed as a binary classification, with the group membership as the target variable. Initially, several data cleaning operations were performed, followed by separation based on biological sex before further preparation and division into sets for cross-validation. Unuseful categories were removed from the visit label (EVENT_ID) and target label (APPRDX) columns. In particular, "ST" was removed from EVENT_ID, "SWEDD" and "PRODROMAL" were removed from

[1] https://www.ppmi-info.org/.

Table 1
PPMI features used in our study.

ai_caudate	It is a metric used in neuroimaging research. It focuses on the asymmetry in the size or volume of the caudate nucleus, a brain structure involved in motor control and cognitive functions. In the context of the PPMI database, this index is computed by comparing the volumes of the caudate nucleus between the left and right hemispheres of the brain. It is a continuous feature measured as a percentage.
ai_putamen	It serves as a quantitative measure to assess the imbalance between the left and right putamen regions within the brain. The putamen, situated in the basal ganglia, plays a pivotal role in modulating motor functions. This index is derived by evaluating the relative sizes or other metrics of the left and right putamen, typically acquired through various neuroimaging methodologies. It is a continuous feature measured as a percentage.
ai_striatum	It is a metric used to quantify the degree of asymmetry in the size of the striatum between the left and right hemispheres of the brain. The striatum is a critical component of the basal ganglia involved in various motor and cognitive functions. Researchers use various imaging techniques to measure striatal parameters and this index. It is a continuous feature measured as a percentage.
urate	Urate, also known as uric acid, is a chemical compound that is the final oxidation product of purine metabolism in humans. It is naturally present in the blood and is a potent antioxidant, protecting against oxidative stress. Serum Urate Acid (UA) levels can be influenced by various factors including diet, genetics, and certain medications. It is a continuous feature ranging from 0 to different levels of serum concentration.
PATNO	Patient number.
APPRDX	It is the "appropriate diagnosis" or "participant group". It encompasses the following categories: "PD Participant", "Healthy Control", "SWEDD" (Scan Without Evidence of Dopaminergic Deficit), and "Prodromal". It is a categorical feature measured as "1" for "PD participant"; "2" for "Healthy Controls"; "3" for SWEDD"; "4″ for "Prodromal".
EVENT_ID	It is the visit time. It is a categorical feature including the following categories: "BL": "BL Baseline"; "V04": "Visit 4 (Year 1)"; "V06": "Visit 6 (Year 2)"; "V08": "Visit 8 (Year 3)"; "V10": "Visit 10 (Year 4)"; "V12": "Visit 12 (Year 5)"; "ST": "Symptomatic Therapy Visit".
gen	Sex (male or female). It is categorical feature measured as "1": "Male"; "2": "Female"
ageonset	Age at PD symptom onset. It is a continuous feature measured as ordinal numbers.
agediag	Age at PD diagnosis. It is a continuous feature measured as ordinal numbers.
symptom1	Resting Tremor (initial symptom at PD diagnosis). It is a categorical feature measuring the symptom presence as "0": "absent"; "1": "present".
symptom2	Rigidity (initial symptom at PD diagnosis). It is a categorical feature measuring the symptom presence as "0": "absent"; "1": "present".
symptom3	Bradykinesia (initial symptom at PD diagnosis). It is a categorical feature measuring the symptom presence as "0": "absent"; "1": "Symptom3".
symptom4	Postural Instability (initial symptom at PD diagnosis). It is a categorical feature measuring the symptom presence as "0": "absent"; "1": "present".
symptom5	Other (initial symptom at PD diagnosis). It is a categorical feature measuring the symptom presence as "0": "absent"; "1": "present".
symptom6	Missing initial symptoms. It is a categorical feature measuring the symptoms presence as "0": "absent"; "1": "present".
DOMSIDE	Side most affected at PD symptom onset. It is a categorical feature measured as "Left", "Right" or "Symmetric".
CAUDATE_R	Striatal Binding Ratio. Brain region used for ROI sampling for right caudate. It is a continuous variable ranging from 0.06 to 5.09.
CAUDATE_L	Striatal Binding Ratio. Brain region used for ROI sampling for left caudate. It is a continuous variable ranging from 0.06 to 5.09.
MSEADLG	The Schwab and England Activities of Daily Living (ADL) scale is a tool used to assess the functional abilities of individuals with PD. It is commonly used in research and clinical settings to measure the level of independence in performing daily activities. The scale rates the patient's ability to perform various activities such as dressing, feeding, hygiene, walking, and other routine tasks. It is a continuous feature measured as a simple rating scale that ranges from 0% to 100%, with 100% indicating complete independence and 0% indicating complete dependence.
PD_MED_USE	This field pertains to the documentation of medication usage for the treatment of PD. It is a categorical feature measured as the following categories: "0" signifies individuals not taking medication for PD; "1" denotes Levodopa usage; "2" indicates Dopamine Agonist usage; "3" represents Other medications; "4" indicates a combination of Levodopa with Other medications; "5" signifies a combination of Levodopa with Dopamine Agonist; "6" represents a combination of Dopamine Agonist with Other medications; "7" indicates a combination of Levodopa, Dopamine Agonist, and Other medications.
APOE	Gene that encodes a protein involved in lipid metabolism and the transport of cholesterol in the body. It has different alleles, including APOE-ε2, APOE-ε3, and APOE-ε4. It is a categorical feature measured as six different categories according to different alleles: "ε3/ε3", "ε4/ε3", "ε3/ε2", "ε2/ε4", "ε4/ε4", "ε2/ε2".
APOE-ε4	It represents the count of APOE-ε4 alleles present in the individual. It is a categorical feature ranging from 0 to 2. The APOE-ε4 allele is associated with an increased risk of developing Alzheimer's disease and has also been studied in the context of PD. It, indeed, may influence the risk of developing PD.
EDUCYRS	Years of formal education a person has received.
HVLTFPRL	It is the score of the "Hopkins Verbal Learning Test (HVLT), False Alarms". It refers to a component of the HVLT, which is a cognitive assessment tool used to evaluate verbal learning and memory. The "False Alarms" feature in the HVLT typically measures a participant's tendency to incorrectly recognize or recall items that were not actually presented during the learning phase of the test. False alarms can indicate issues with memory, attention, or cognitive processing. The scoring of the test involves assessing the number of incorrect responses made by the participant during the recognition phase. It is a continuous variable ranging from 0 to 12.
HVLTRDLY	It is the score of the "Hopkins Verbal Learning Test, Delayed Recall". It refers to a component of the Hopkins Verbal Learning Test assessment. The test captures data related to a participant's ability to recall verbal information after a delay. The score involves assessing the number of words correctly recalled by the participant after a delay period. It is a continuous feature ranging from 0 to 36.
HVLTREC	It is the score of the "Hopkins Verbal Learning Test, Delayed Recognition". The test consists of memorization of a list of words to test the ability to recall immediately after memorization (immediate recall) and after a 20 min delay (delayed recall). It is a continuous feature ranging from 0 to 36.
MAPT	Gene that encodes the tau protein, which is involved in maintaining the structure of nerve cells. It is a categorical feature, including three different haplotypes: "H1/H2", "H1/H1", and "H2/H2".
SDMTOTAL	The Symbol Digit Modalities Test is a neuropsychological test that is generally interpreted as a measure of information processing speed, although several cognitive domains, such as lexical access speed and working memory contribute to its result. In its oral version, the patient is asked to verbally associate each symbol with the corresponding number, using a legend. The task is to correctly code as many symbols as possible in 90 s. The written modality consists of a sheet of paper with, at the top, a sequence of nine symbols and nine corresponding numbers (key). The task sequence consists of a series of symbols, each with a blank space underneath. Within a 90-s time limit the subject is required, consulting the key as necessary, to insert the numbers associated with the symbols. It is a continuous feature ranging from 0 to 110.
SNCA-rs356181	Specific genetic variant or single nucleotide polymorphism (SNP) within the SNCA gene. It is a categorical feature

(*continued on next page*)

Table 1 (*continued*)

	measured as the following four categories: C/C, C/T, T/T, and None. These notations refer to the different genotypes individuals can have at that specific SNP locus. In particular, C/C indicates that an individual has two copies of the "C" allele at the SNP locus. In genetic notation, each individual inherits one allele from each parent, so C/C means both copies of the gene at that locus are the C allele. This is also referred to as homozygous for the C allele. C/T indicates that an individual has one copy of the "C" allele and one copy of the "T" allele at the SNP locus. This is also referred to as heterozygous for the SNP locus. T/T indicates that an individual has two copies of the "T" allele at the SNP locus. Similar to C/C, this is also homozygous but for the T allele. None indicates that the genotype could not be determined for that individual at the SNP locus.
SNCA-rs3910105	Specific genetic variant or single nucleotide polymorphism (SNP) within the SNCA gene. It is a categorical feature with the same categories of SNCA-rs356181.
VLTANIM	Semantic Verbal Fluency (SVF) test, animal semantics. The SVF consists of giving the person 60 s to verbally list as many things as possible in a category (animal semantics). To score the test, count up the total number of animals or words that the individual is able to produce. A score of under 17 indicates concern, although some practitioners use 14 as a cutoff. It is a continuous feature ranging from 0 to 36.
VLTFRUIT	SVF test, fruits semantics. It consists of giving the person 60 s to verbally list as many things as possible in a category (fruits semantics). As for VLTANIM, it is a continuous feature ranging from 0 to 36.
age	Age. It is a continuous feature.
age_cat	Age Category at enrollment. In our dataset it is a categorical feature measured as "1": "<56 yrs", "2": "56–65 yrs", and "3": ">65 yrs".
bjlot	The Benton Judgment of Line Orientation Test (BJLOT) is a standardized test of visuospatial skills commonly associated with functioning of the parietal lobe in the right hemisphere. The test measures a person's ability to match the angle and orientation of lines in space. Scores classify individuals as normal, borderline (or mild), moderate, and severely impaired. A score of 17 or less is considered a sign of severe deficit. It is a continuous feature ranging from 0 to a maximum depending on numbers of short-forms utilized (usually 15-items short forms).
changedx	Change in primary diagnosis from Baseline to current visit.
educ	Education. It is a categorical feature measured as: "1": "<13 yrs", "2": "13–23 yrs", and "3": ">23 yrs".
fampd_new	Family history of disease is a categorical feature measured as: "1": "First Degree Family with PD"; "2": "Non-1st Degree Family with PD"; "3": "No Family with PD".
gds_cat	The Geriatric Depression Scale (GDS) is a self-reported measure of depression symptoms in elderly persons. It consists of 15 yes/no questions. Normally it is a continuous feature but in our dataset it is a categorical feature measured as "0": "<5 (not depressed)" and "1": "5 or greater (depressed)".
hvlt_discrimination	It is the score of the "Hopkins Verbal Learning Test, Discrimination Recognition". It refers to a component of the Hopkins Verbal Learning Test assessment. It is a continuous feature ranging from 0 to 36.
hvlt_immediaterecall	It is the score of the "Hopkins Verbal Learning Test, Immediate/Total Recall". It refers to a component of the Hopkins Verbal Learning Test assessment. It is a continuous feature ranging from 0 to 36.
hvlt_retention	It is the score of the "Hopkins Verbal Learning Test, Retention". It refers to a component of the Hopkins Verbal Learning Test assessment. It is a continuous feature ranging from 0 to 36.
quip	It is the score of the Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease (QUIP), which has 3 sections: Section 1 assesses four ICDs (involving gambling, sexual, buying, and eating behaviors), Section 2 other compulsive behaviors (punding, hobbyism and walkabout) and Section 3 compulsive medication use. It is a continuous feature ranging from 0 to 5.
quip_buy	Section n. 1/3 of QUIP, Buy compulsion.
quip_eat	Section n. 1/3 of QUIP, Eating compulsion.
quip_gamble	Section n. 1/3 of QUIP, Gambling compulsion.
quip_sex	Section n. 1/3 of QUIP, Sex compulsion.
race	Race. It is a categorical feature measured as "1": "White", "2": "Black", "3": "Asian", and "4": "Other".
rigidity	It is represented by the Total Rigidity Score (OFF State). Rigidity is scored based on the passive movement of extremities when participants are in a relaxed sitting position, with higher scores reflecting more rigidity, according to the UPDRS-III scale. It is a continuous feature ranging from to 100.
rigidity_on	It is represented by the Total Rigidity Score (ON State). As for rigidity (OFF state), rigidity_on is scored based on the passive movement of extremities when participants are in a relaxed sitting position, with higher scores reflecting more rigidity, according to the UPDRS-III scale. It is a continuous feature ranging from to 100.
scopa	It refers to the total score of "Scales for Outcomes in Parkinson's Disease - Autonomic Dysfunction" (SCOPA-AUT). The SCOPA-AUT is a component of the wider research effort called SCales for Outcomes in PArkinson's disease (SCOPA), which selects or develops clinimetric and practical sound instruments for all key areas in PD. The scale consists of 25 items assessing the following regions: gastrointestinal, urinary, cardiovascular, thermoregulatory, pupillomotor and sexual (2 items for males and 2 items for females) dysfunction [31]. It is a continuous feature ranging from 0 to 125.
scopa_cv	It is the score for SCOPA-AUT, Cardiovascular dysfunction score. It is a continuous feature ranging from 0 to 15.
scopa_gi	It is the score for SCOPA-AUT, Gastrointestinal score. It is a continuous feature ranging from 0 to 35.
scopa_pm	It is the score for SCOPA-AUT, Pupillomotor dysfunction score. It is a continuous feature ranging from 0 to 5.
scopa_sex	It is the score for SCOPA-AUT, Sexual dysfunction score. It is a continuous feature ranging from 0 to 10.
scopa_therm	It is the score for SCOPA-AUT, Thermoregulatory dysfunction score. It is a continuous feature ranging from 0 to 20.
scopa_ur	It is the score for SCOPA-AUT, Urinary dysfunction score. It is a continuous feature ranging from 0 to 30.
sft	It is the Semantic Fluency Task score, calculated as the sum of the scores obtained from the categories VLTANIM, VLTFRUIT, and VLTVEG. This task serves to assess both lexical organization and executive function by prompting participants to generate as many words as they can from a given semantic category within a designated time frame. The resulting score is a continuous measure, ranging from 0 to 108.

APPRDX, focusing only on "PD Participant", "Healthy Control". Subsequently, features with near-zero variance within the target category were dropped (features that had target-conditional counts equal to target value counts), heavily biased features, single-valued features (such as sex, since has already been stratified by us), features biased towards a single target category (columns having a zero count only for one APPRDX value). Moreover, features with strong associations with the target variable were removed. Then, the target labels were transformed (APPRDX) into integer values (1 for "PD Participant" and 0 for "Healthy Control"), and the dataset was divided by biological sex into males (M) and females (F). A single visit label was selected for analysis, choosing the most common EVENT_ID among patients ("BL Baseline") while maintaining a valid distribution of APPRDX ("PD Participant", "Healthy Control"). Lastly, a single record per patient was chosen to minimize record dependence. This was a straightforward step as selecting a single EVENT_ID already resulted in a single record per patient.

This comprehensive data pre-processing ensured that the dataset was clean, well-structured, and suitable for subsequent analysis and modeling. Here a list of the starting features after cleaning: *APOE, APOE-ε4, EDUCYRS, HVLTFPRL, HVLTRDLY, HVLTREC, MAPT, SDMTOTAL, SNCA-rs356181, SNCA-rs3910105, VLTANIM, VLTFRUIT, age, age_cat, bjlot, changedx, educ, fampd-new, gds_cat, hvlt_discrimination, hvlt_immediaterecall, hvlt_retention, quip, quip_buy, quip_eat, quip_gamble, quip_sex,*

race, rigidity, rigidity_on, scopa, scopa_cv, scopa_gi, scopa_pm, scopa_sex, scopa_therm, scopa_ur, sft.

After this phase, the dataset was split into training (65%), validation (15%), and test (20%) sets. Record assignment ensured that each patient number (PATNO) appeared in only one of the three sets, based on a random allocation of PATNO. This allocation was facilitated by the dataset prior filtering to include one record per patient. Table 2 presents the distribution of the dataset following the pre-processing stage, while Table 3 delineates the dataset partitioning according to diagnosis.

The male cohort chosen for the training set had an average age of 61.2 years (with a range from 30.62 to 84.9 and a standard deviation of 10.23), coupled with an average education duration of 15.7 years (ranging from 5 to 26, with a standard deviation of 2.97). Conversely, the female counterpart had an average age of 60.0 years (ranging from 30.97 to 83.7, with a standard deviation of 10.31), alongside an average education duration of 15.4 years (ranging from 5 to 26, with a standard deviation of 2.82).

2.4. Sex-specific classification model and feature importance

2.4.1. CatBoost algorithm

The CatBoost algorithm was employed to construct a sex-based classification model capable of distinguishing between individuals who are healthy and those who have PD. Using the CatBoost algorithm in unraveling sex differences in PD offers several advantages compared to other ML algorithms. Firstly, CatBoost stands out due to its capability to directly handle categorical features without requiring one-hot encoding or label encoding, thereby simplifying the pre-processing step and potentially saving significant time and effort [32]. This feature makes it particularly well-suited for analyzing datasets that include numerous categorical variables such as sex, medication type, and symptom severity.

Secondly, CatBoost demonstrates remarkable robustness against overfitting. Leveraging gradient boosting, it inherently mitigates overfitting by amalgamating predictions from multiple weak models (decision trees) to construct a robust and interpretable predictive model. Moreover, CatBoost employs various techniques including regularization and early stopping to further alleviate overfitting, thereby enhancing its ability to generalize effectively to unseen data. This capability is crucial when dealing with medical datasets characterized by limited size.

Thirdly, CatBoost is moderately robust to multicollinearity and can handle missing values in the dataset without needing imputation, allowing users to work with data as is (cf. Section 2.3). This means users can analyze and process their data without the potentially heavy task of filling in missing values artificially. CatBoost employs an advanced algorithm that effectively integrates missing values into its decision-making process during training. By treating missing values as a separate category, CatBoost ensures that no valuable information is lost due to data pre-processing steps. The CatBoost approach of treating missing values does not need additional parameter settings. It is particularly beneficial in heterogeneous tabular data, such as clinical datasets where each patient may have unique characteristics, and little variations can impact diagnosis outcomes. This method is advantageous because it preserves the original distribution and characteristics of the data without introducing biases from imputation. Trusting CatBoost for handling missing values simplifies the data pre-processing pipeline, saving time and effort typically spent on imputation methods. Moreover, by not relying on imputation, CatBoost reduces the risk of introducing bias or distortion into the data, which can be crucial for accurate model training and prediction. CatBoost, being a Gradient Boosted Decision Tree model, remains highly interpretable. It natively offers scores linked to measures of information gain or loss, which allows easy sorting of features according to their importance for classification. The model ability to divide the feature space accurately, even in the presence of missing values, maintains the clarity of feature importance and decision paths.

Finally, CatBoost is known for its high prediction accuracy, which is essential for uncovering subtle differences and patterns in medical data, especially when studying sex differences in PD where the effects might be nuanced. All these aspects make CatBoost an flexibility and performance tool to capture complex relationships within the dataset, delivering robust and interpretable predictions [33].

The CatBoostClassifier model from the Python package was utilized, employing CrossEntropy as the loss function. To determine the optimal hyperparameters, a Grid Search leveraging the built-in function of CatBoost was conducted. This function inherently incorporates a stratified K-Fold method, which proves beneficial for addressing unbalanced datasets. Three folds were opted for. Each best-performing model underwent testing on the validation set. Subsequently, the best model was validated before evaluating its performance against the test set. The Grid Search algorithm was employed with the following parameter grid to identify the optimal hyperparameters: 'iterations': [5, 10, 100, 1000]; 'depth': [2, 3, 5]; 'learning_rate': [0.01, 0.05, 0.1]; 'l2_leaf_reg': [1, 3, 5]; 'early_stopping_rounds': [5]. The primary aim was to achieve results close to optimal while considering the robustness and stability inherent in CatBoost. Unlike deep neural networks that often demand meticulous fine-tuning, CatBoost resilience allows for a more pragmatic approach, balancing optimization efforts with achievable gains. Moreover, an initial exploration with alternative hyperparameters revealed that the optimal performance typically resided within the above specified ranges. To address the risk of overfitting, the inherent robustness of CatBoost acts as an initial safeguard, supplemented by the deliberate choice of moderate 'depth' and the implementation of cross-validation during the Grid Search process. The 'early_stopping_rounds' hyperparameter not only serves to prevent performance degradation but also acts as a guard against overfitting, ensuring that the model doesn't persist in training when further improvements are marginal or non-existent. This comprehensive approach helped maintain a balance between model complexity and generalization, ultimately leading to a more reliable and effective predictive model.

2.4.2. Model evaluation

The classification performance was assessed by calculating the accuracy, which is determined by the ratio of correct predictions to the total number of predictions made by the model. Additionally, the AUC (Area Under the Curve) metric was employed to gauge the overall effectiveness of the binary classification model. This involved computing the area under the ROC (Receiver Operating Characteristic) curve, which graphically illustrates the true positive rate (sensitivity) against the false positive rate (1 - specificity) across various thresholds of the binary classification model. The best Grid Search model achieved an accuracy of 0.89 and an AUC of 0.86 on the test set for females. This

Table 2
Dataset partitioning at the end of the dataset pre-processing.

	Males partition	Females partition
Training set size	261	139
Validation set size	61	33
Test set size	81	44
Total	403	216

Table 3
Dataset partitioning in terms of appropriate diagnosis (APPRDX), "0" for "Healthy Control" and "1" for "PD Participant".

Sex	APPRDX	Count	Percent
F	0	95	0.24
M	0	181	0.45
F	1	44	0.11
M	1	80	0.2

model is characterized by a depth of 2, a learning rate of 0.1, and an L2 leaf regulation parameter of 3, with 10 iterations (10 estimators). For males, the accuracy was 0.90, with an AUC of 0.88 on the test set. This model is also characterized by a depth of 2, a learning rate of 0.1, and an L2 leaf regulation parameter of 1, with 10 iterations (10 estimators).

2.4.3. Feature importance and statistical analysis

SHAP (SHapley Additive exPlanations) analysis techniques with the "permutation importance" method were utilized to further evaluate the significance of features within the CatBoost model [34]. This method, grounded in game theory, evaluates the importance of each feature by examining all possible feature combinations within the model. It aids in comprehending why the ML algorithm generated a particular prediction for an individual by analyzing each feature (such as age, APOE, SDMTOTAL, etc.) and determining its contribution to the final prediction. More in detail, SHAP initiates by exploring all conceivable combinations of features for an individual and observes how each combination impacts the prediction. Subsequently, it evaluates how the addition or removal of one feature influences the prediction. For instance, it may be observed that including "APOE" heightens the likelihood of predicting PD, whereas adding "SDMTOTAL" diminishes that likelihood. Through these various scenarios, the algorithm computes the significance of each feature, revealing its contribution to the final prediction. Therefore, if the CatBoost algorithm predicts that an individual has PD, SHAP might elucidate, for instance, that it's due to factors such as advanced age, specific medical history, etc. Ultimately, SHAP aids in understanding the rationale behind a particular prediction by breaking down the importance of each feature in the prediction process.

The SHAP approach offers several advantages over alternative methods for evaluating feature importance. SHAP values furnish an intuitive explanation of feature significance by quantifying the influence of each feature on model predictions. This interpretive capability is pivotal for comprehending the elements contributing to sex differentials in PD and can aid clinicians and researchers in pinpointing pertinent biomarkers or risk factors. In addition, the algorithm satisfies several desirable attributes such as local accuracy, adaptability to missing data, and consistency, ensuring that the assessments of feature importance remain robust and dependable across diverse subsets of the data. SHAP is model-agnostic, thus adaptable to any machine learning model, including CatBoost. This adaptability empowers researchers to study feature importance across an array of algorithms without being confined to a specific modeling methodology.

The combined power of CatBoost and SHAP offers a robust framework to analyze complex data, pinpoint crucial features, and gain insights into the underlying mechanisms influencing disease development and progression in both sexes. CatBoost and SHAP capacity to handle categorical variables, mitigate overfitting, and offer interpretable feature importance metrics make them particularly well-suited for this task compared to other ML algorithms and feature importance techniques.

A Mann-Whitney U test was conducted to statistically compare SHAP values between males (M) and females (F). The significance level was set at 0.05. Each test included the calculation of 95% confidence intervals for the sampled differences between the datasets, achieved through bootstrapping. Notably, the Mann-Whitney test does not mandate that the compared samples have equal sizes. One of its assumptions is that instances or records in the dataset are independent of each other. Since our dataset is anonymous, verifying this assumption directly is not feasible, and any patient relationships remain unknown. Nevertheless, precautions were implemented, such as employing a single record for each patient and partitioning the data based on patient identifiers (PATNO). All analyses were performed using the Python programming language and the Pandas, Matplotlib, Seaborn, CatBoost, scikit-learn, and SHAP libraries.

2.4.4. Feature interaction analysis

Feature interaction analysis stands as a robust method for unraveling the relationships among predictor variables, ultimately bolstering model interpretability and predictive efficacy. It facilitates the identification and measurement of interactions between features throughout the model training process. In datasets, such as ours, characterized by complex dependencies among predictor variables, feature interaction analysis proves invaluable.

At its essence, feature interaction analysis involves examining how the predictive performance of the model is affected by the combined effects of multiple features. Specifically, the algorithm assesses the feature interaction strength for each pair of features by examining all feature splits within the resulting ensemble of trees. For each pair of features, CatBoost looks at all the splits in the trees where these features are used. If splits of both features are present in the same tree, CatBoost calculates how much the leaf value (the final prediction) changes when these splits have the same value and when they have opposite values. The larger the change, the stronger the interaction between the two features. Thus, through the exploration of feature interactions, it is possible to uncover synergistic or antagonistic relationships between predictors, thereby shedding light on important patterns and dependencies within the data.

A pivotal aspect of feature interaction analysis in CatBoost lies in its ability to automatically detect and exploit interactions between categorical and numerical features, obviating the need for explicit feature engineering. By considering interactions across different feature types, CatBoost can capture more nuanced representations of the underlying data landscape, consequently enhancing predictive accuracy and model performance.

3. Results

3.1. Key features to predict PD in females and males

Fig. 1 presents two plots that compare the nine most important features, as determined by SHAP values obtained from the ML algorithm, separately for females and males. Specifically, each plot illustrates the SHAP values for the features utilized in CatBoost model training across all examples (records) sampled from the test set. These values indicate the extent to which a feature contributes to altering the CatBoost prediction in comparison to the expected value (refer to Section 2.4.3). SHAP values less than zero lean towards a PD outcome, whereas SHAP values greater than zero lean towards a healthy controls outcome. The SHAP algorithm assesses both motor and non-motor features. For both females and males, the SHAP values associated with certain features appear to be spread across the x-axis, while those linked to other features show less dispersion. In the latter scenario, these features not only prove to be significant but also provide a robust indication of the likelihood for females or males to belong to either a healthy controls or PD group with greater accuracy compared to instances with wide dispersion of SHAP values.

Fig. 2 provides a concise representation of the most influential features identified through SHAP analysis for both females and males, based on the average of the absolute SHAP values. Table 4 summarizes the result of the Mann-Whitney test to statistically compare SHAP values between females and males.

3.2. Interactions among key PD predictors for females and males

Fig. 3 summarizes the results of the features interaction analysis (cf., 2.4.4). This analysis shows how the predictive performance of the ML model is influenced by the collective effects of multiple features. In datasets characterized by complex dependencies among predictor variables, such as ours, this kind of analysis proves invaluable.

The interaction of features inherently operates asymmetrically, taking into account outcome alterations when feature B is either present or

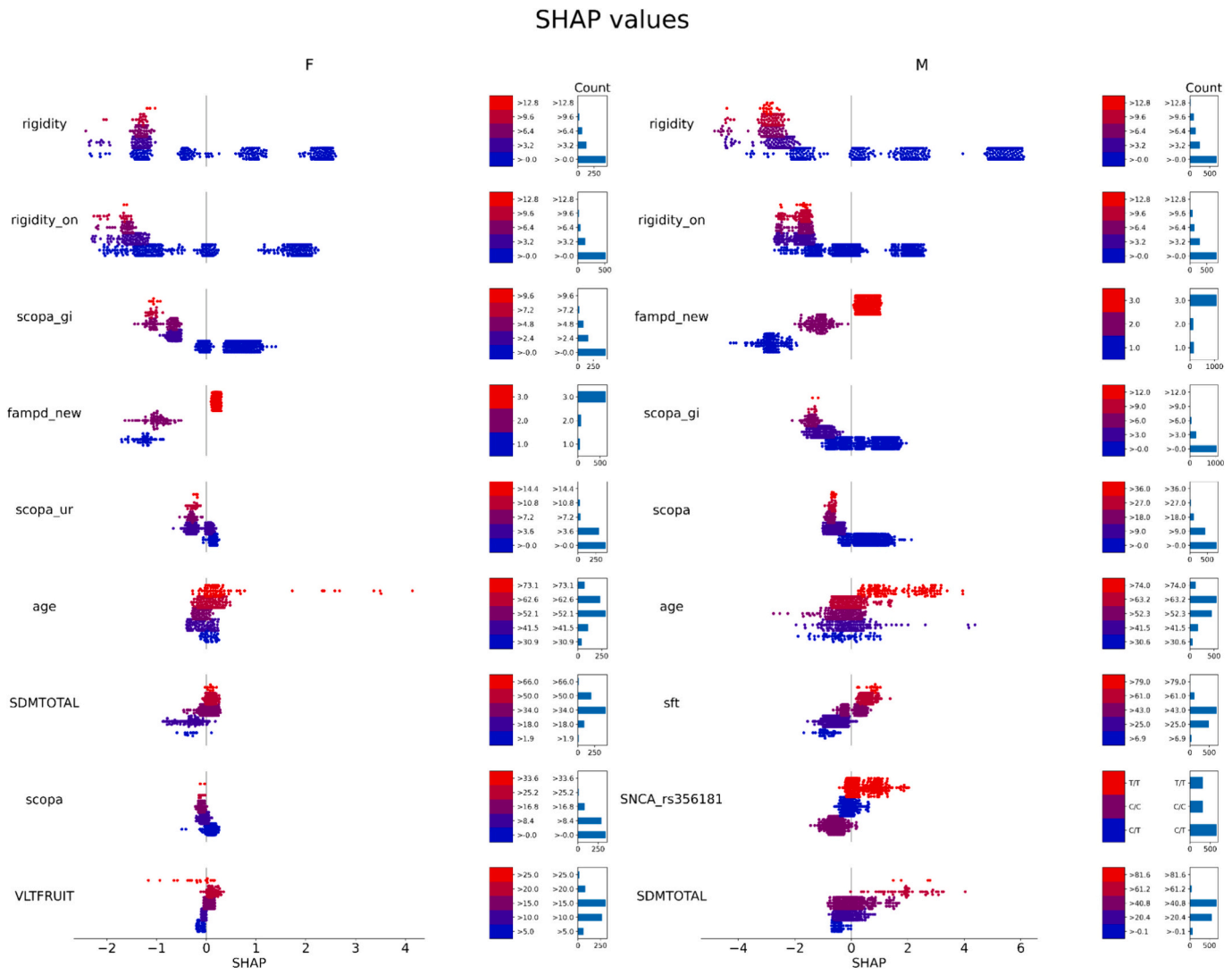

Fig. 1. SHAP beeswarm plot from females (F) and males (M) individuals. Each data point corresponds to a specific record within the test set. The vertical axis represents training features, sorted by their respective feature importance. Feature names remain consistent with their original source from the PPMI dataset. On the horizontal axis, each point signifies an individual record and expresses the SHAP value attributed to that particular feature during prediction. The color of each point reflects the true value of the feature in the corresponding record used during prediction. To enhance figure readability, categorical features were represented using three strips, each corresponding to a category. Similarly, data points associated with continuous features were binned into five strips, obtained by using five ranges representing the values the feature can encompass. The "Count" column summarizes the number of records for each strip.

absent in splits, with feature A serving as a reference point. Consequently, each split entails not only features A and B but also additional variables that may differ across various splits, leading to observed asymmetries. While CatBoost consideration of feature presence or absence in splits may be asymmetric, it effectively identifies and leverages interactions that improve predictive accuracy, maintaining its ability to capture complex relationships between features and outcomes. The asymmetry could be interesting as it underscores how certain features might exert a more pronounced influence on the target than others. Consequently, they might engage in interactions with other features, whereas the latter may not wield the same level of impact or relevance to the same extent.

4. Discussion

4.1. Model results framed within the current literature on sex difference in PD

The observation that PD manifests differently in males and females could imply the involvement of partially distinct neurophysiological mechanisms. It could be crucial to categorize sex variations using cutting-edge tools such as ML, which can aid clinicians in comprehending the impact of these differences on the etiology of PD. Upon comparing our ML model findings with existing literature on sex differences in PD, several key aspects of the ML model output that align with or corroborate previous research were identified. In our analyses, the deliberate choice was made not to focus on the classic couple of symptoms associated with the disease at the time of diagnosis (tremor, and bradykinesia), as these are now recognized as established clinical biomarkers and the most predictive features [35]. However, we decided to incorporate the total score of rigidity into the model, encompassing both the period before and after diagnosis, and considering the follow-up of both healthy controls and PD patients. This decision stems from the fact that while current clinical instrumentation offers greater recognition and objectivity for tremor and bradykinesia, the pathophysiological basis of rigidity remains unknown, and its experimental measurement is notably inadequate [36]. Aside from rigidity, we trained the ML model to predict the differentiation between healthy controls

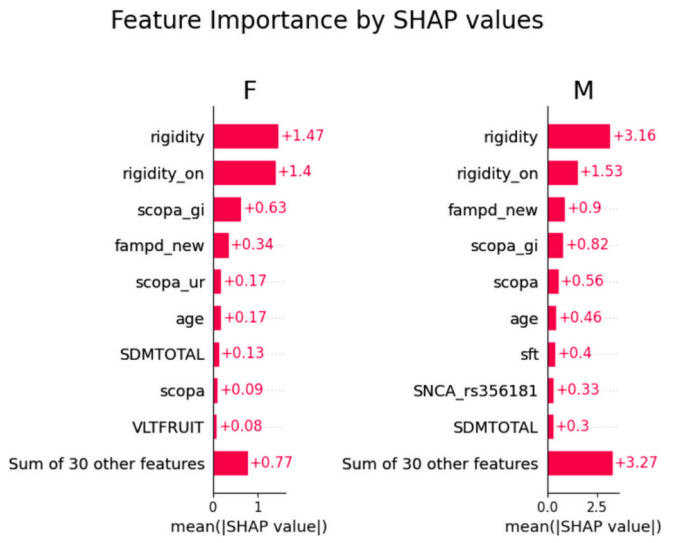

Fig. 2. Critical features to predict PD identified through SHAP analysis. Summary bar chart depicts the importance of features based on their SHAP values, with the mean SHAP value displayed. Each bar represents the expected value of SHAP values in absolute terms.

Table 4
Mann-Whitney *U* test comparing SHAP value distributions between females (F) and males (M). A *p*-value <0.05 (bold) indicates a statistically significant difference. Abbreviations: MM: male median; FM: female median; DL95: lower 95% confidence interval boundary for the difference between male and female medians; DU95: upper 95% confidence interval boundary.

	p	MM	FM	DL95	DU95
rigidity	**0.0**	−1.87	−0.71	−0.15	0.35
rigidity_on	**0.026**	−0.66	−0.90	−0.32	−0.005
fampd_new	**0.0**	0.40	0.18	−0.12	0.05
scopa_gi	0.431	−0.14	−0.02	−0.08	0.07
scopa	0.578	0.09	−0.002	−0.04	0.04
age	**0.00**	−0.11	−0.06	−0.03	0.07
sft	**0.011**	0.20	0.004	−0.03	0.02
SNCA_rs356181	**0.0**	−0.08	0.0	−0.10	−0.04
SDMTOTAL	**0.0**	−0.09	0.03	−0.04	0.02
scopa_ur	**0.0**	0.01	0.09	−0.02	0.02
VLTFRUIT	**0.001**	0.02	−0.03	−0.01	0.03

and individuals with PD by using heterogeneous data encompassing demographic, genetic, biological, and clinical information related to both motor and non-motor features (cf. Section 2.3).

Figs. 1, 2 show that the most predictive feature for both females and males is the total rigidity score during OFF phase treatment, followed by the total rigidity score during ON phase treatment. Muscle rigidity is a prevalent symptom among PD patients, persisting from the disease onset through its progression [37,38]. Research indicates that males often exhibit more pronounced rigidity, while females tend to display increased postural instability and susceptibility to levodopa-induced dyskinesia [39,15,40,41]. However, these findings do not diminish the importance of rigidity in predicting PD in females; they merely underscore its higher prevalence in males. The existing literature does not explicitly investigate whether rigidity significantly contributes to PD prediction when stratified by sex. Our model addresses this gap. Our ML analysis suggests, indeed, that rigidity is a significant predictor for males (Figs. 1, 2). In addition, the SHAP value for males is significantly higher than that for females (Table 4), suggesting that the rigidity and rigidity_on may be more predictive factors in determining the risk of PD in males than in females. Overall these results align with the literature emphasis on rigidity prominence in males. Nonetheless, the ML analysis also underscores rigidity importance in predicting PD in females,

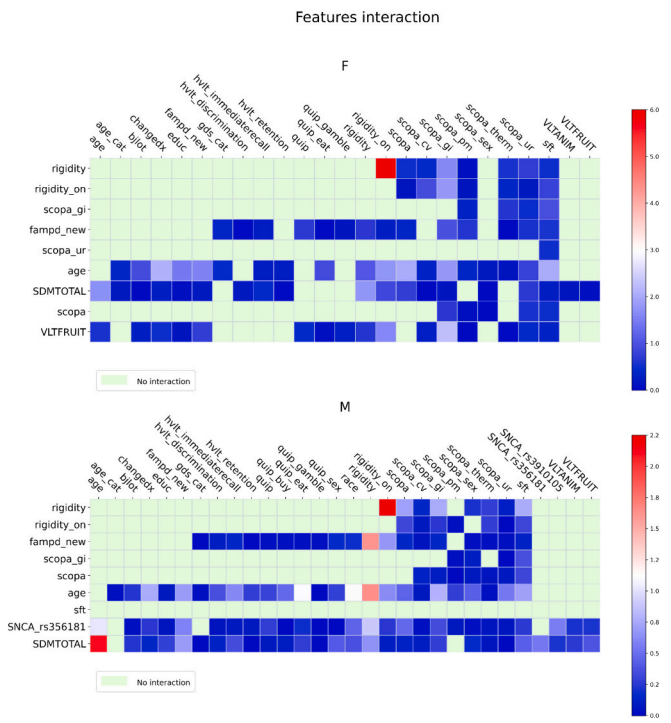

Fig. 3. Interactions among the key features and all other features.

representing a novel insight with respect to current literature.

In terms of SHAP analysis (Fig. 1), rigidity consistently demonstrates a polarized distribution for both sexes, with low values indicating a healthy condition and high values suggesting disease. This observation corroborates existing literature [39,41] and underscores rigidity clinical significance as a pivotal motor symptom, making it a valuable diagnostic marker for both males and females.

The ML model also identified the tests for autonomic dysfunctions (scopa scores in Table 1), as highly significant for predicting PD progression in both females and males, albeit with notable differences (see Figs. 1, 2). Specifically, the score associated with gastrointestinal symptoms (scopa_gi) emerges as influential for both females and males. The comprehensive score evaluating gastrointestinal, urinary, cardiovascular, thermoregulatory, pupillomotor, and sexual dysfunctions (scopa) holds greater importance for males. Additionally, the scores related to urinary dysfunctions (scopa_ur) are crucial for females but not as much for males. The difference in the SHAP values for scopa_gi and scopa is not statistically significant between the two groups, whereas is statistically significant for scopa_ur (Table 4).

Although explicit studies measuring the importance of these features in predicting PD with sex stratification are lacking, contextualizing the findings of our ML analysis within existing literature exploring sex differences in the scoring of these tests is possible. According to existing research, females tend to exhibit higher autonomic dysfunction scores [42], whereas males demonstrate higher rates of gastrointestinal symptoms [43] and urinary issues [44]. For instance, in terms of gastrointestinal symptoms, studies have shown that stool samples from female PD patients exhibit elevated levels of various immunological and angiogenesis mediators associated with the illness, while those from male patients do not significantly differ from controls [45]. Additionally, while male PD patients are more prone to severe drooling, female patients are more likely to develop significant dysphagia [46,47].

These findings do not undermine the importance of gastrointestinal and urinary deficits in predicting PD in females; rather, they highlight the higher prevalence of these symptoms in males. Despite scoring lower compared to males, the ML analysis underlines the significance of these features as predictors for females. This outcome presents a novel insight

compared to existing literature and suggests a need to be more cautious in linking higher scoring values with greater feature importance. Likely, various factors contribute to this perspective, recognizing the complexity of sex-specific manifestations in PD and the necessity for nuanced interpretations of autonomic dysfunction markers in both sexes. Our examination of these diverse autonomic features contributes to a more nuanced comprehension of the sex-specific manifestations of PD, illuminating the intricate interplay between autonomic dysfunction and disease outcomes in both males and females.

According to the feature importance algorithm, the presence of a family history of disease (fampd_new) tends to increase the likelihood of developing PD for both males and females (refer to Fig. 1). The model indicates that the effect of family history is significantly higher for males compared to females (refer to Fig. 2 and Table 4). This finding appears to diverge from the literature, which has not identified significant differences between the sexes concerning this feature [48]. Therefore, it would be beneficial to conduct further analyses to explore this inconsistency between the ML model results and those documented in the literature, aiming to gain deeper insights into the role of family history of disease in the context of PD.

The model identifies age as a significant predictor for PD in both males and females. Specifically, advancing age is correlated with a higher probability of developing the disease (see Fig. 1). However, the SHAP values for males are significantly higher than those for females, indicating that older age may exert a stronger predictive influence on PD risk in males than in females (see Fig. 2 and Table 4). These findings partially align with the existing literature. Some studies suggest no sex disparities among PD patients concerning age [48]. However, alternative evidence indicates that females tend to develop the disease at a later age, with longer disease duration observed at the time of their initial UPDRS-III assessment [30]. Females typically experience the illness later in life, display longer-lasting symptoms, and undergo their initial PD assessment at an older age [20,49,21,22]. Hence, the SHAP values derived from our model corroborate literature findings, confirming that age is a significant predictor of disease development. However, a discrepancy arises between the ML model results and those of the literature regarding sex differences. While the literature emphasizes that females with PD usually have a higher age, the ML model indicates a more pronounced impact of age in males compared to females. This contrast suggests that age may hold varying importance in the context of PD, depending on factors like the studied population or patients' level of education. Therefore, given these factors and the fact that some studies do not emphasize significant differences between males and females with PD concerning age, further exploration of the significance of this feature in PD is necessary. This exploration should consider sex differences and include a detailed examination of the interactions among various factors influencing the disease development and progression.

Our analysis further suggests that scores obtained from the Semantic Verbal Fluency Test (sft) could be crucial in predicting PD, particularly among males. In contrast, for the female cohort, only the test evaluating fruit semantics (VLTFRUIT) seems to have some significance, albeit with a relatively minor impact (see Figs. 1 and 2). These findings partially align with the literature emphasis on the prominence of fluency tests in females [50,51]. However, our ML analysis also underscores the importance of sft in predicting PD in males, representing a novel insight compared to the current literature. Further analyses would be beneficial to gain deeper understanding into the role of these tests in the context of PD stratified by sex.

The analysis of feature importance conducted through the ML model indicates that cognitive assessments (SDMTOTAL), influence the diagnosis of PD in both males and females (Fig. 1). The effect of SDMTOTAL is significantly higher for males compared to females (Fig. 2 and Table 4). This result agrees with literature [15,52] and, in general, aligns with data on cognitive distinctions between males and females, as discussed below (Section 4.2).

Finally, our model indicates that the specific variant SNCA-rs356181 of the alpha-synuclein gene SNCA may hold greater significance in characterizing PD among males compared to females. This result represents a model prediction since, although there is evidence suggesting that certain genetic variants may interact with sex-related factors to influence PD risk and progression [53,54], comprehensive investigations into the sex-specific effects of SNCA-rs356181 are lacking. It's essential to recognize that the influence of SNCA-rs356181 on PD manifestation may vary depending on factors such as the demographics of the study population, research methodologies employed, and other genetic or environmental variables. Therefore, further research is indispensable to ascertain whether SNCA-rs356181 indeed plays a more prominent role in characterizing PD in males compared to females. In broader terms, future studies elucidating the intricate interplay between genetic components, sex-specific disparities, and various PD phenotypes are pivotal for enhancing our comprehension of the disease and devising personalized treatment strategies. Such endeavors may entail large-scale genetic analyses, functional assessments, and in-depth explorations into the molecular mechanisms underlying sex-specific discrepancies in PD [13,37,55].

Features interaction (Fig. 3) shows that for females, the level of rigidity in the OFF condition (rigidity) strongly interacts with the level of rigidity in the ON condition (rigidity_on). Additionally, there are other interactions, albeit of lesser intensity but still relevant, between rigidity and rigidity_on and the scopa_gi. Age interacts with various tests for autonomic dysfunctions, as well as with rigidity and certain cognitive assessment tests. Furthermore, there is some interaction observed between VLTFRUIT, rigidity_on, and scopa_gi.

The level of rigidity strongly interacts with the level of rigidity_on, even among males. Additionally, there are other interactions, though of lesser intensity yet still relevant, between rigidity and autonomic dysfunctions tests, as well as with sft. Familiarity also interacts with rigidity. Like for females, even for males age interacts with various tests for autonomic dysfunctions, with rigidity, and with certain cognitive assessment tests. However, in males, age also interacts with family history and with tests investigating compulsive behavior and with race. Furthermore, SNCA-rs356181 interacts with age, family history, and rigidity.

These interactions could be partially explained by existing literature. For instance, the notable correlation between age and certain characteristics in males is consistent with established data showing that males are more prone to developing PD at an earlier age, whereas females typically experience the onset of the illness later in life, endure prolonged symptoms, and undergo their initial PD assessment at an advanced age [20,49,21,22]. However, there is still a lack of literature investigating the complex interplay of various features in defining PD based on sex. Consequently, the model results suggest the need for further exploration.

Interestingly, the analysis of feature interaction reveals a notable distinction: there exists more interactions among features in males as opposed to females, as illustrated in Fig. 3. The female group exhibits a decreased number of interactions between features; however, the interactions present are notably stronger compared to those in males (refer to the heatmap ranges for both groups). This discovery aligns with existing literature that underscores the neurobiological differences in brain structure, function, and neurotransmitter systems between the sexes [56,57]. Fig. 3 provides additional clarity by highlighting the increased significance of studying these interactions in males for predicting PD. The heightened feature interactions observed in males may suggest unique patterns of disease-related biomarkers or alterations in neural circuitry specific to Parkinsonian males when compared to females. These disparities could offer valuable insights into the pathophysiology of PD and potentially guide the development of sex-specific diagnostic and therapeutic approaches.

4.2. Literature on sex difference in PD

Table 5 presents a summary of the main scientific literature investigating sex differences in PD.

In recent years, investigations have brought to light various sex differences in the PD characteristics [15,17]. Males generally show a higher likelihood of developing PD, with females experiencing the illness later in life, exhibiting longer-lasting symptoms, and undergoing their initial PD assessment at an older age [20,49,21,22]. The influence of estrogen on dopaminergic neurons is suggested as a potential cause for the clinical and cognitive differences between sex in PD [63,23,64,65]. Differences for serum UA levels report males having higher levels associated with slower disease progression [66]. Recent research, however, indicates that high UA levels reduce PD risk in both sexes [67]. Males also tend to have higher UA levels, suggesting a potential protective effect.

Some studies emphasize comparable age-related differences between females and males with PD. Middle-aged individuals with PD exhibit lower motor scores despite having a disease duration similar to those with an onset at an older age. However, additional evidence indicates that females tend to experience a later onset of the disease and present with higher age and disease duration at the initial assessment [62,68,69,21].

PD patients exhibit distinct clinical phenotypes based on their sex [15,35]. Several studies have investigated the variation in frequency and severity of motor symptoms across sex in relation to PD [15]. While extensive evidence suggests a potentially milder and slower progression of PD in females, other studies indicate that females are more prone than males to experience postural instability [15] and motor issues, including levodopa-induced dyskinesia [39,61,70,71] and the "wearing-off" phenomenon [40,35]. Tremor often serves as the initial presenting symptom in females [39,41,38,25], while other motor signs tend to manifest later [37,40]. Additionally, recent observations indicate that males are more prone than females to encounter issues such as drooling, sexual dysfunction, urinary problems, and excessive daytime drowsiness [72,14].

Several studies have investigated the correlation between sex and the prevalence of non-motor symptoms in individuals with PD [73,38,35]. Females with PD show pronounced fatigue, sadness, restless legs, constipation, discomfort, loss of taste or smell, weight fluctuations, and increased perspiration compared to their male counterparts [15,74]. Furthermore, they are more prone to experiencing anxiety and depression [75,76,77]. Additionally, pain emerges as one of the most prevalent non-motor symptoms of PD, with females consistently reporting a higher frequency of this symptom than males [77,78,79]. Contrary to this pattern, other literature presents conflicting findings regarding cognitive impairment. While the majority of studies reveal no significant distinctions between males and females in this aspect [75,80,38,74,81,82,83], some studies suggest that males are more likely than females to experience cognitive impairment [77,48,35,78]. Conversely, other research argues in favor of females exhibiting cognitive impairment [76]. Supporting these findings, male PD patients have been observed to undergo more significant brain atrophy and structural connection disruption than their female counterparts, indicating that males may experience more severe neurodegeneration at the onset of the illness [72,84]. Female patients with PD demonstrate diminished visuospatial function but obtain good scores in verbal fluency tests (SDMTOTAL), and the Montreal Cognitive Assessment scale (MOCA) compared to male patients with PD [15,52]. Notably, females exhibit superior cognitive functioning, evident through their exceptional performance on assessments such as the HVLT [42,60] and the Scales for Outcomes in Parkinson's Disease-Cognition (SCOPA-COG) [59]. Conversely, males often display poorer performances on tests measuring verbal, letter, and category fluency [60], Hopkins Verbal Learning Test - Discrimination (hvlt_discrimination), Hopkins Verbal Learning Test - Delayed Recall (HVLTRDLY), and the Semantic Fluency Task (sft), where female PD patients typically achieve higher scores [50,51]. However, results from the Benton Judgment of Line Orientation Test (bjlot) indicate significantly better performance among males compared to females [61,51], aligning with the observed weaker visuospatial abilities of females in the Hopkins Verbal Learning Tests-Discrimination (hvlt_discrimination). Regarding mental health assessment, no significant sex disparities were found in the Geriatric Depression Scale (GDS) scores [48]. Family history, education, age, and race did not exhibit significant sex differences in PD according to recent literature [48].

The literature on sex differences in PD does not explicitly explore an analysis of feature importance in the manner discussed in this article. Here we underline that it is crucial to differentiate between scoring differences and the actual importance of features. For example, if research indicates that a certain feature scores higher in males compared to females, it does not necessarily imply that the feature lacks importance in characterizing PD in females. This finding does not diminish the significance of the feature in predicting PD in females; it simply emphasizes its lower prevalence in this group.

As of our current understanding, there is a lack of direct investigation into the most crucial features for characterizing PD diagnosis based on sex differences, with the meaning discussed in this article. Several factors may hide the importance of a feature for a particular sex. For instance, different scoring scales may be necessary for males and females for the same test, or the interaction between different features may need consideration. Our ML analysis allows us to clarify these underlying factors, emphasizing the importance of a feature not only through scoring but also by considering other aspects, such as interactions among features. This insight holds considerable potential for

Table 5
Main literature supporting sex-based differences in PD. None of the articles explicitly outline a feature importance analysis as addressed in this article. The most important predictive features of PD in sex stratification identified by our ML algorithm are highlighted in bold.

Feature	Sex-based difference	Reference
urate	M higher serum UA levels are associated with a slower disease progression of PD	[19]
hvlt_discrimination	M worse in verbal, letter, and category fluency tests, F worse in visuospatial abilities	[58]
scopa_gi	M higher scores	[43]
scopa_ur	M higher scores	[44]
scopa_cog	F better performance	[59]
scopa	F higher scores	[42]
HVLTFPRL	F higher performances in memory, executive functions (Semantic fluency)	[42]
rigidity	M more advanced rigidity, and F more instability and a tendency towards increased levodopa-induced dyskinesia	[39,15,40,41]
SDMTOTAL	M more rapid progression of cognitive decline in processing speed	[15,52]
HVLTRDLY	M worse performance than F in delayed recall	[42,51,60]
HVLTREC	No sex differences	[51,60]
bjlot	M better performance	[61,51]
gds_cat	No sex differences	[48]
sft	F higher scores	[50,51]
fampd_new	No sex differences	[48]
educ	No sex differences	[48]
age	Some studies find no sex disparities in PD patients. However, other evidence suggests F have later disease onset and higher age and disease duration when first assessed with UPDRS-III	[20,62,30,49,48,21,22]
race	No sex differences	[48]

personalized medicine, suggesting the need to tailor data collection methods for males and females to enhance diagnostic and therapeutic outcomes. Despite this, Table 5 and the accompanying discussion on sex differences in literature presented in this subsection, serve as valuable reference points. They enable us to contextualize the findings of our model within the established body of knowledge. Indeed, as discussed earlier (Section 4.1), we found a certain alignment between the analysis of feature importance conducted by our model and the literature summarized in Table 5.

4.3. The route to integrate sex-based biomarkers into current clinical practice

Integrating sex-specific biomarkers for PD diagnosis into current clinical practice is essential for improving patient care and advancing personalized medicine in neurology. The study highlights significant differences in PD development and manifestation between males and females, underscoring the importance of tailoring diagnostic approaches accordingly. To begin with, educating healthcare professionals about these sex-specific biomarkers is paramount. This action involves raising awareness about the differences in PD presentation and progression between sexes and emphasizing the role of biomarkers in accurately diagnosing and treating the disease.

Updating diagnostic guidelines is another crucial step. Collaborating with medical associations and organizations to incorporate sex-specific biomarkers into these guidelines ensures that healthcare providers have clear guidance on which assessments to prioritize and how to interpret their results in the context of sex differences. Incorporating biomarker testing into routine PD diagnostic protocols is a practical approach. By including specific sex-based assessments for muscle rigidity, autonomic function, cognitive abilities, and family history, as identified in this article, clinicians can gather more comprehensive data to inform their diagnostic decisions. Tailoring treatment plans based on sex-specific biomarker results is essential for optimizing patient care. For instance, if some biomarkers indicate a higher risk or different progression pattern in males, treatment strategies can be tuned accordingly to address these needs.

Enhancing patient education is also crucial in this process. Educating patients about the significance of sex-specific biomarkers empowers them to advocate for comprehensive assessments and personalized care based on their characteristics, ultimately improving their treatment outcomes. Furthermore, fostering ongoing research and development in this area is essential. The exploration of sex-specific biomarkers for PD and refinement of existing ones will contribute to improving diagnostic accuracy and informing tailored treatment approaches in the future.

5. Conclusions and future works

This study sheds light on crucial sex-specific variations in PD, unraveling distinctive clinical phenotypes and manifestation patterns. Four main aspects distinguish our ML-based approach. Firstly, results obtained through the ML model reveal that muscle rigidity symptoms, autonomic features, cognitive assessments, and family history emerge as pivotal contributors to PD diagnosis, with acknowledged variations between sexes. Moreover, the specific genetic variant SNCA-rs356181 may hold greater significance in characterizing PD among males compared to females. These results largely align with existing literature on sex differences in PD, underscoring the necessity for deeper empirical investigations, especially in cases with partial contrasts with existing literature. Validation of literature data emphasizes the clinical relevance of the model results, offering direction for future research and personalized diagnostic strategies.

Secondly, literature on sex differences in PD does not explicitly delve into the importance of features as proposed by our ML approach. It is crucial to differentiate between scoring differences and the actual importance of features: if a certain feature scores lower in, for example, males compared to females, it does not necessarily imply that the feature lacks importance in characterizing PD in males. It simply emphasizes its lower prevalence in this group. Indeed, several factors may hide the importance of a feature for a particular sex. For instance, different scoring scales may be necessary for males and females for the same test, or the interaction between different features may need consideration. Our ML analysis elucidates these underlying factors, emphasizing the significance of a feature while considering its score as one aspect among others in determining its impact. This insight holds considerable potential for personalized medicine, suggesting the need to tailor data collection methods for males and females to enhance diagnostic and therapeutic outcomes.

Thirdly, we adopt a system-level approach, integrating heterogeneous data from clinical, imaging, genetics, and demographics, and investigating on their interactions to better understand the complexity of PD and potentially identify novel biomarkers related to both motor and non-motor symptoms for diagnosis and prognosis [7,13].

Finally, we utilize an explainable ML approach to uncover patterns and relationships within the data, shedding light on the distinct features defining PD in females and males. An explainable ML algorithm provides clear and understandable explanations about how it reached a result [29,30]. This transparency helps humans, especially non-ML experts, grasp the reasoning behind the model decisions, fostering trust and facilitating interpretation of complex ML outcomes. In this way, it could impact usability, clarity, and real translational research.

Although previous studies have reviewed the use of ML in the diagnosis and assessment of PD, they were limited to the analysis of motor symptoms (tremor, rigidity, and bradykinesia), and only a few studies focused on non-motor symptoms [85,86,87]. By recognizing and addressing sex disparities through explainable ML techniques, healthcare professionals can tailor interventions to facilitate early diagnosis and effective disease management, thereby improving outcomes for both male and female patients.

Despite these results, our proposed approach stands to benefit from further refinement through future research endeavors. One area that requires attention is the need to enhance the sex balance within our datasets by increasing the number of both male and female participants.

Additionally, it's important to note that the analyzed PPMI dataset comprises both idiopathic and genetic forms of PD. Justifying our results becomes challenging considering the division of patients into genetic and non-genetic forms. The inclusion of a mixed population could introduce additional variability in the data, potentially affecting the generalizability of the findings. While the study aims to identify sex-specific patterns within this heterogeneous dataset, it's essential to acknowledge that different subtypes may display distinct clinical characteristics. Further investigations are necessary to analyze these distinctions to utilize our proposed method for studying different PD cohorts. Achieving a more balanced representation entails expanding beyond the confines of the PPMI cohort, which predominantly comprises individuals from North America. It is imperative to incorporate cohorts from diverse regions such as Europe and Asia to foster a broader and more inclusive dataset.

Another limitation of our study may lie in our model ability to only differentiate between healthy controls and PD patients. Future research could benefit from a more targeted exploration of specific subtypes to enhance the precision and applicability of sex-based analyses in PD that consider the unique characteristics of each individual. Future work should also consider the potential influence of lifestyle factors and disease progression over time on the manifestation and diagnosis of PD [13]. Incorporating longitudinal data and lifestyle variables such as diet, physical activity, and environmental exposures into our ML model might enhance its predictive accuracy and clinical relevance. By tracking changes in these variables over time, researchers can better understand how lifestyle factors interact with genetic predispositions and other risk factors to influence disease onset and progression in females and males [88]. This approach could also help to identify critical periods where

interventions might be most effective, leading to more personalized and proactive disease management strategies.

All these actions could allow us to capture a richer spectrum of populations encompassing different geographical regions, ethnicities, and socio-cultural backgrounds. This diversification not only improves the robustness and reliability of our findings but also increases their applicability across a wide array of real-world scenarios and populations.

CRediT authorship contribution statement

Gianfrancesco Angelini: Methodology, Software, Formal analysis, Investigation, Data curation, Visualization. **Antonio Malvaso:** Methodology, Validation, Writing – review & editing. **Aurelia Schirripa:** Validation, Visualization, Writing – review & editing. **Francesca Campione:** Validation, Writing – review & editing. **Sebastian Luca D'Addario:** Methodology, Validation, Resources. **Nicola Toschi:** Validation, Writing – review & editing. **Daniele Caligiore:** Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Acknowledgements

Gianfrancesco Angelini is a PhD student enrolled in the National PhD in Artificial Intelligence, XXXVIII cycle, course on Health and life sciences, organized by Università Campus Bio-Medico di Roma.

References

[1] D. Caligiore, F. Mannella, G. Baldassarre, Different dopaminergic dysfunctions underlying Parkinsonian akinesia and tremor, Front. Neurosci. 13 (2019) 550.
[2] A. Kouli, K.M. Torsney, W.L. Kuan, Parkinson's disease: etiology, neuropathology, and pathogenesis, in: T.B. Stoker, J.C. Greenland (Eds.), Parkinson's Disease: Pathogenesis and Clinical Aspects [Internet], Codon Publications, Brisbane (AU), 2018, https://doi.org/10.15586/codonpublications.parkinsonsdisease.2018.ch1, 2018 Dec 21. Chapter 1.
[3] J.A. Obeso, M.C. Rodriguez-Oroz, C.G. Goetz, C. Marin, J.H. Kordower, M. Rodriguez, G. Halliday, Missing pieces in the Parkinson's disease puzzle, Nat. Med. 16 (6) (2010) 653–661.
[4] X. Li, C. Chen, T. Pan, X. Zhou, X. Sun, Z. Zhang, X. Chen, Trends and hotspots in non-motor symptoms of Parkinson's disease: a 10-year bibliometric analysis, Front. Aging Neurosci. 16 (2024).
[5] R.F. Pfeiffer, Non-motor symptoms in Parkinson's disease, Parkinsonism Relat. Disord. 22 (2016) S119–S122.
[6] O.B. Tysnes, A. Storstein, Epidemiology of Parkinson's disease, J. Neural Transm. (Vienna) 124 (8) (2017 Aug) 901–905. Epub 2017 Feb 1. PMID: 28150045, https://pubmed.ncbi.nlm.nih.gov/28150045/, https://doi.org/10.1007/s00702-017-1686-y.
[7] D. Caligiore, R.C. Helmich, M. Hallett, A.A. Moustafa, L. Timmermann, I. Toni, G. Baldassarre, Parkinson's disease as a system-level disorder, NPJ Parkinson's Dis. 2 (2016) 16025, https://doi.org/10.1038/npjparkd.2016.25.
[8] D. Caligiore, G. Pezzulo, G. Baldassarre, A.C. Bostan, P.L. Strick, K. Doya, I. Herreros, Consensus paper: towards a systems-level view of cerebellar function: the interplay between cerebellum, basal ganglia, and cortex, Cerebellum 16 (2017) 203–229.
[9] R.C. Helmich, The cerebral basis of parkinsonian tremor: a network perspective, Mov. Disord. 33 (2) (2018) 219–231.
[10] J.L. Lanciego, N. Luquin, J.A. Obeso, Functional neuroanatomy of the basal ganglia, Cold Spring Harb. Perspect. Med. 2 (12) (2012) a009621.
[11] C. Marras, C.G. Canning, S.M. Goldman, Environment, lifestyle, and Parkinson's disease: implications for prevention in the next decade, Mov. Disord. 34 (6) (2019) 801–811.
[12] A. Parashar, M. Udayabanu, Gut microbiota: implications in Parkinson's disease, Parkinsonism Relat. Disord. 38 (2017) 1–7.
[13] D. Caligiore, F. Giocondo, M. Silvetti, The Neurodegenerative Elderly Syndrome (NES) Hypothesis: Alzheimer and Parkinson are Two Faces of the Same Disease, IBRO Neuroscience Reports, 2022.
[14] S. Meoni, A. Macerollo, E. Moro, Sex differences in movement disorders, Nat. Rev. Neurol. 16 (2) (2020) 84–96.
[15] S. Cerri, L. Mus, F. Blandini, Parkinson's disease in women and men: What's the difference? J. Parkinsons Dis. 9 (3) (2019) 501–515, https://doi.org/10.3233/JPD-191683.
[16] L. Hirsch, N. Jette, A. Frolkis, T. Steeves, T. Pringsheim, The incidence of Parkinson's disease: a systematic review and meta-analysis, Neuroepidemiology 46 (4) (2016) 292–300.
[17] R. Patel, K. Kompoliti, Sex and gender differences in Parkinson's disease, Neurol. Clin. 41 (2) (2023 May) 371–379, https://doi.org/10.1016/j.ncl.2022.12.001. Epub 2023 Feb 19. PMID: 37030964.
[18] T. Pringsheim, N. Jette, A. Frolkis, T.D. Steeves, The prevalence of Parkinson's disease: a systematic review and meta-analysis, Mov. Disord. 29 (13) (2014) 1583–1590.
[19] M.C. Russillo, V. Andreozzi, R. Erro, M. Picillo, M. Amboni, S. Cuoco, P. Barone, M. T. Pellecchia, Sex differences in Parkinson's Disease: from bench to bedside, Brain Sci. 12 (7) (2022) 917, https://doi.org/10.3390/brainsci12070917.
[20] L.M.L. De Lau, P.C.L.M. Giesbergen, M.C. De Rijk, A. Hofman, P.J. Koudstaal, M.M. B. Breteler, Incidence of Parkinsonism and Parkinson disease in a general population: the Rotterdam study, Neurology 63 (7) (2004) 1240–1244.
[21] S.K. Van Den Eeden, et al., Incidence of Parkinson's disease: variation by age, sex, and race/ethnicity, Am. J. Epidemiol. 157 (1) (2003) 1015–1022.
[22] G.F. Wooten, L.J. Currie, V.E. Bovbjerg, J.K. Lee, J. Patrie, Are men at greater risk for Parkinson's disease than women? J. Neurol. Neurosurg. Psychiatry 75 (4) (2004) 637–639.
[23] I.N. Miller, A. Cronin-Golomb, Sex differences in Parkinson's disease: clinical characteristics and cognition, Mov. Disord. 25 (16) (2010) 2695–2703.
[24] M. Picillo, R. Erro, M. Amboni, C. Vitale, A. Iavarone, M. Moccia, R. Allocca, G. Orefice, P. Barone, Gender differences in non-motor symptoms in early Parkinson's disease: a 2-years follow-up study on previously untreated patients, Parkinsonism Relat. Disord. 8 (2014) 850–854.
[25] M. Picillo, A. Nicoletti, V. Fetoni, B. Garavaglia, P. Barone, M.T. Pellecchia, The relevance of gender in Parkinson's disease: a review, J. Neurol. 264 (2017) 1583–1607.
[26] M. Abdar, M. Zomorodi-Moghadam, Impact of patients' gender on Parkinson's disease using classification algorithms, J. AI Data Min. 6 (2) (2018) 277–285, https://doi.org/10.22044/JADM.2017.4673.1555.
[27] A.J. Perrin, E. Nosova, K. Co, A. Book, O. Iu, V. Silva, S. Appel-Cresswell, Gender differences in Parkinson's disease depression, Parkinsonism Relat. Disord. 36 (2017) 93–97.
[28] Y. Yang, Y. Hu, Y. Chen, Y. Xie, S. Nie, Machine Learning Model Based on Sex Subgroup for Improving Parkinson's Disease Diagnosis, 2023. Available at SSRN 4356412.
[29] S. Dhanalakshmi, R.S. Maanasaa, R.S. Maalikaa, R. Senthil, A review of emergent intelligent systems for the detection of Parkinson's disease, Biomed. Eng. Lett. 13 (4) (2023) 591–612, https://doi.org/10.1007/s13534-023-00319-2.
[30] U. Gupta, H. Bansal, D. Joshi, An improved sex-specific and age-dependent classification model for Parkinson's diagnosis using handwriting measurement, Comput. Methods Prog. Biomed. 189 (2020) 105305, https://doi.org/10.1016/j.cmpb.2019.105305. Epub 2019 Dec 28. PMID: 31935580.
[31] Martine Visser, et al., Assessment of autonomic dysfunction in Parkinson's disease: the SCOPA-AUT, Mov. Disord. 19 (11) (2004) 1306–1312.
[32] A.V. Dorogush, V. Ershov, A. Gulin, CatBoost: Gradient Boosting with Categorical Features Support, arXiv preprint, 2018 arXiv:1810.11363, https://arxiv.org/abs/1810.11363.
[33] C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, A comparative analysis of gradient boosting algorithms, Artif. Intell. Rev. 54 (2021) 1937–1967.
[34] Scott M. Lundberg, Su-In Lee, A unified approach to interpreting model predictions, in: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 4768–4777.
[35] A. Nicoletti, R. Baschi, C.E. Cicero, S. Iacono, V.L. Re, A. Luca, R. Monastero, Sex and gender differences in Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis: a narrative review, Mech. Ageing Dev. 212 (2023) e111821.
[36] F. Asci, M. Falletti, A. Zampogna, M. Patera, M. Hallett, J. Rothwell, A. Suppa, Rigidity in Parkinson's disease: evidence from biomechanical and neurophysiological measures, Brain 146 (9) (2023) 3705–3718.
[37] W. Chen, X. Yan, H. Lv, Y. Liu, Z. He, X. Luo, Gender differences in prevalence of LRRK2-associated Parkinson disease: a meta-analysis of observational studies, Neurosci. Lett. 715 (2020) 134609.
[38] K.W. Kang, S.M. Choi, B.C. Kim, Gender differences in motor and non-motor symptoms in early Parkinson disease, Medicine 101 (3) (2022).
[39] Y. Baba, J.D. Putzke, N.R. Whaley, Z.K. Wszolek, R.J. Uitti, Gender and the Parkinson's disease phenotype, J. Neurol. 252 (10) (2005) 1201–1205, https://doi.org/10.1007/s00415-005-0835-7.
[40] D. Colombo, G. Abbruzzese, A. Antonini, P. Barone, G. Bellia, F. Franconi, F. Stocchi, The "gender factor" in wearing-off among patients with Parkinson's disease: a post hoc analysis of DEEP study, Sci. World 2015 (1) (2015) 1–10.
[41] C.A. Haaxma, B.R. Bloem, G.F. Borm, W.J. Oyen, K.L. Leenders, S. Eshuis, J. Booij, D.E. Dluzen, M.W. Horstink, Gender differences in Parkinson's disease, J. Neurol. Neurosurg. Psychiatry 78 (2007) 819–824.
[42] C. Boccalini, G. Carli, A. Pilotto, A. Padovani, D. Perani, Gender differences in dopaminergic system dysfunction in de novo Parkinson's disease clinical subtypes, Neurobiol. Dis. 167 (2022) 105668, https://doi.org/10.1016/j.nbd.2022.105668.
[43] Ö. Skogar, M. Nilsson, J. Lökk, Gender differences in diagnostic tools, medication, time to medication, and nonmotor symptoms in parkinsonian patients, Brain Circ. 8 (4) (2022) 192–199, https://doi.org/10.4103/bc.bc_33_22.
[44] N.R. Bansal, B.S. Paul, G. Paul, G. Singh, Gender differences and impact of autonomic disturbance on fatigue and quality of life in Parkinson's disease, Neurol. India 70 (1) (2022) 203–208, https://doi.org/10.4103/0028-3886.336334.
[45] M.C. Houser, J. Chang, S.A. Factor, E.S. Molho, C.P. Zabetian, E.M. Hill-Burns, M. G. Tansey, Stool immune profiles evince gastrointestinal inflammation in Parkinson's disease, Mov. Disord. 33 (5) (2018) 793–804.

[46] C.J. Mao, Y.T. Xiong, F. Wang, Y.P. Yang, W. Yuan, C. Zhu, J. Chen, C.F. Liu, Motor subtypes and other risk factors associated with drooling in Parkinson's disease patients, Acta Neurol. Scand. 137 (2018) 509–514.
[47] J.C. Nienstedt, M. Bihler, A. Niessen, R. Plaetke, M. Pötter-Nerger, C. Gerloff, C. Buhmann, C. Pflug, Predictive clinical factors for penetration and aspiration in Parkinson's disease, Neurogastroenterol. Motil. 3 (2019) e13524.
[48] R. Liu, D.M. Umbach, S.D. Peddada, Z. Xu, A.I. Tröster, X. Huang, H. Chen, Potential sex differences in nonmotor symptoms in early drug-naive Parkinson disease, Neurology 84 (21) (2015) 2107–2115, https://doi.org/10.1212/WNL.0000000000001609.
[49] K.E. Lyons, J.P. Hubble, A.I. Tröster, Sex differences in Parkinson's disease, Clin. Neuropharmacol. 21 (1998) 118–1121.
[50] Z. Chen, B. Wu, G. Li, L. Zhou, L. Zhang, J. Liu, Age and sex differentially shape brain networks in Parkinson's disease, CNS Neurosci. Ther. 29 (7) (2023) 1907–1922, https://doi.org/10.1111/cns.14149.
[51] J. Oltra, C. Uribe, A. Campabadal, A. Inguanzo, G.C. Monté-Rubio, M.J. Martí, Y. Compta, F. Valldeoriola, C. Junque, B. Segura, Sex differences in brain and cognition in *de novo* Parkinson's disease, Front. Aging Neurosci. 13 (2022) 791532, https://doi.org/10.3389/fnagi.2021.791532.
[52] T.H. Reekes, C.I. Higginson, C.R. Ledbetter, N. Sathivadivel, R.M. Zweig, E. A. Disbrow, Sex specific cognitive differences in Parkinson disease, NPJ Parkinson's Dis. 6 (2020) 7, https://doi.org/10.1038/s41531-020-0109-1.
[53] C. Blauwendraat, H. Iwaki, M.B. Makarious, S. Bandres-Ciga, H.L. Leonard, F. P. Grenn, C. Garcia, Investigation of autosomal genetic sex differences in Parkinson's disease, Ann. Neurol. 90 (1) (2021) 35–42.
[54] K. Markopoulou, B.A. Chase, A.P. Premkumar, B. Schoneburg, N. Kartha, J. Wei, D. Maraganore, Variable effects of PD-risk associated SNPs and variants in parkinsonism-associated genes on disease phenotype in a community-based cohort, Front. Neurol. 12 (2021) 662278.
[55] Y. Li, Y. Xia, S. Yin, F. Wan, J. Hu, L. Kou, T. Wang, Targeting microglial α-synuclein/TLRs/NF-kappaB/NLRP3 inflammasome axis in Parkinson's disease, Front. Immunol. 12 (2021) 719807.
[56] H. Loke, V. Harley, J. Lee, Biological factors underlying sex differences in neurological disorders, Int. J. Biochem. Cell Biol. 65 (2015) 139–150.
[57] M.F. Ullah, A. Ahmad, S.H. Bhat, F.M. Abu-Duhier, G.E. Barreto, G.M. Ashraf, Impact of sex differences and gender specificity on behavioral characteristics and pathophysiology of neurodegenerative disorders, Neurosci. Biobehav. Rev. 102 (2019) 95–105.
[58] Y. Hou, Q. Wei, R. Ou, J. Yang, W. Song, Q. Gong, H. Shang, Impaired topographic organization in cognitively unimpaired drug-naïve patients with rigidity-dominant Parkinson's disease, Parkinsonism Relat. Disord. 56 (2018) 52–57, https://doi.org/10.1016/j.parkreldis.2018.06.021.
[59] E.F. Augustine, A. Pérez, R. Dhall, C.C. Umeh, A. Videnovic, F. Cambi, A.M. Wills, J.J. Elm, R.M. Zweig, L.M. Shulman, M.A. Nance, J. Bainbridge, O. Suchowersky, Sex differences in clinical features of early, treated Parkinson's disease, PLoS One 10 (7) (2015) e0133002, https://doi.org/10.1371/journal.pone.0133002.
[60] K. Yang, B. Shen, D.K. Li, Y. Wang, J. Zhao, J. Zhao, W.B. Yu, Z.Y. Liu, Y.L. Tang, F. T. Liu, H. Yu, J. Wang, Q.H. Guo, J.J. Wu, Cognitive characteristics in Chinese non-demented PD patients based on gender difference, Transl. Neurodegener. 7 (2018) 16, https://doi.org/10.1186/s40035-018-0120-1.
[61] H. Iwaki, C. Blauwendraat, H.L. Leonard, M.B. Makarious, J.J. Kim, G. Liu, M. A. Nalls, Differences in the presentation and progression of Parkinson's disease by sex, Mov. Disord. 36 (1) (2021) 106–117.
[62] N.J. Diederich, C.G. Moore, S.E. Leurgans, T.A. Chmura, C.G. Goetz, Parkinson disease with old-age onset: a comparative study with subjects with middle-age onset, Arch. Neurol. 60 (4) (2003) 529–533.
[63] Mozley L. Harper, R.C. Gur, R.D. Mozley, R.E. Gur, Striatal dopamine transporters and cognitive functioning in healthy men and women, Am. J. Psychiatry 158 (9) (2001) 1492–1499.
[64] J. Lavalaye, J. Booij, L. Reneman, J.B. Habraken, E.A. van Royen, Effect of age and sex on dopamine transporter imaging with [123I] FPCIT SPET in healthy volunteers, Eur. J. Nucl. Med. 27 (7) (2000) 867–869.
[65] C.A. Munro, M.E. McCaul, D.F. Wong, L.M. Oswald, Y. Zhou, J. Brasic, H. Kuwabara, A. Kumar, M. Alexander, W. Ye, G.S. Wand, Sex differences in striatal dopamine release in healthy adults, Biol. Psychiatry 59 (10) (2006) 966–974, https://doi.org/10.1016/j.biopsych.2006.01.008.
[66] M. Wen, B. Zhou, Y.H. Chen, Z.L. Ma, Y. Gou, C.L. Zhang, W.F. Yu, L. Jiao, Serum uric acid levels in patients with Parkinson's disease: a meta-analysis, PLoS One 12 (3) (2017) e0173731, https://doi.org/10.1371/journal.pone.0173731.
[67] H. Chang, B. Wang, Y. Shi, R. Zhu, Dose-response meta-analysis on urate, gout, and the risk for Parkinson's disease, NPJ Parkinson's Dis. 8 (1) (2022) 160.
[68] A. Friedman, Old-onset Parkinson's disease compared with young-onset disease: clinical differences and similarities, Acta Neurol. Scand. 89 (4) (1994) 258–261.
[69] A. Schrag, A. Hovris, D. Morley, N. Quinn, M. Jahanshahi, Young versus older-onset Parkinson's disease: impact of disease and psychosocial consequences, Mov. Disord. 18 (11) (2003) 1250–1256.
[70] M. Zappia, G. Annesi, G. Nicoletti, G. Arabia, F. Annesi, D. Messina, A. Quattrone, Sex differences in clinical and genetic determinants of levodopa peak-dose dyskinesias in Parkinson disease: an exploratory study, Arch. Neurol. 62 (4) (2005) 601–605.
[71] X. Zhou, Z. Liu, X. Zhou, Y. Xiang, Z. Zhou, Y. Zhao, Parkinson's Disease & Movement Disorders Multicenter Database and Collaborative Network in China (PD-MDCNC), The Chinese Parkinson's disease registry (CPDR): study design and baseline patient characteristics, Mov. Disord. 37 (7) (2022) 1335–1345.
[72] M. Bourque, M. Morissette, D. Soulet, T. Di Paolo, Impact of sex on neuroimmune contributions to Parkinson's disease, Brain Res. Bull. 199 (110668) (2023) 1–23.
[73] C. Baiano, P. Barone, L. Trojano, G. Santangelo, Prevalence and clinical aspects of mild cognitive impairment in Parkinson's disease: a meta-analysis, Mov. Disord. 35 (1) (2020) 45–54.
[74] P. Martinez-Martin, C. Falup Pecurariu, P. Odin, J.J. van Hilten, A. Antonini, J. M. Rojo-Abuin, V. Borges, C. Trenkwalder, D. Aarsland, D.J. Brooks, K. Ray Chaudhuri, Gender-related differences in the burden of non-motor symptoms in Parkinson's disease, J. Neurol. 259 (2012) 1639–1647.
[75] F. Farhadi, K. Vosoughi, G.A. Shahidi, A. Delbari, J. Lokk, S.M. Fereshtehnejad, Sexual dimorphism in Parkinson's disease: differences in clinical manifestations, quality of life and psychosocial functioning between males and females, Neuropsychiatr. Dis. Treat. 13 (2017) 329–338, https://doi.org/10.2147/NDT.S124984.
[76] T. Hu, R. Ou, H. Liu, Y. Hou, Q. Wei, W. Song, H. Shang, Gender and onset age related-differences of non-motor symptoms and quality of life in drug-naïve Parkinson's disease, Clin. Neurol. Neurosurg. 175 (2018) 124–129.
[77] Z. Kurlawala, P.H. Shadowen, J.D. McMillan, L.J. Beverly, R.P. Friedland, Progression of nonmotor symptoms in Parkinson's disease by sex and motor laterality, Parkinson's Dis. 2021 (2021).
[78] K. Szewczyk-Krolikowski, P. Tomlinson, K. Nithi, R. Wade-Martins, K. Talbot, Y. BenShlomo, M.T.M. Hu, The influence of age and gender on motor and nonmotor features of early Parkinson's disease: initial findings from the Oxford Parkinson Disease Center (OPDC) discovery cohort, Parkinsonism Relat. Disord. 20 (2014) 99–105, https://doi.org/10.1016/J.PARKRELDIS.2013.09.025.
[79] M.A.S. Zella, C. May, T. Müller, M. Ahrens, L. Tonges, R. Gold, K. Marcus, D. Woitalla, Landscape of pain in Parkinson's disease: impact of gender differences, Neurol. Res. 41 (2019) 87–97, https://doi.org/10.1080/01616412.2018.1531208.
[80] X. Guo, W. Song, K. Chen, X.P. Chen, Z. Zheng, B. Cao, R. Huang, B. Zhao, Y. Wu, H.F. Shang, Gender and onset age-related features of non-motor symptoms of patients with Parkinson's disease–a study from Southwest China, Parkinsonism Relat. Disord. 19 (2013) 961–965, https://doi.org/10.1016/J.PARKRELDIS.2013.06.009.
[81] M. Picillo, M. Amboni, R. Erro, K. Longo, C. Vitale, M. Moccia, M.T. Pellecchia, Gender differences in non-motor symptoms in early, drug naïve Parkinson's disease, J. Neurol. 260 (2013) 2849–2855.
[82] P. Solla, A. Cannas, F.C. Ibba, F. Loi, M. Corona, G. Orofino, M.G. Marrosu, F. Marrosu, Gender differences in motor and non-motor symptoms among Sardinian patients with Parkinson's disease, J. Neurol. Sci. 323 (2012) 33–39.
[83] J.E. Yoon, J.S. Kim, W. Jang, J. Park, E. Oh, J. Youn, S. Park, J.W. Cho, Gender differences of nonmotor symptoms affecting quality of life in Parkinson disease, Neurodegener. Dis. 17 (2017) 276–280, https://doi.org/10.1159/000479111.
[84] C. Tremblay, N. Abbasi, Y. Zeighami, Y. Yau, M. Dadar, S. Rahayel, A. Dagher, Sex effects on brain structure in de novo Parkinson's disease: a multimodal neuroimaging study, Brain 143 (10) (2020) 3052–3066.
[85] C. Ahlrichs, M. Lawo, Parkinson's Disease Motor Symptoms in Machine Learning: A Review, arXiv preprint, 2013 arXiv:1312.3825.
[86] M. Belić, V. Bobić, M. Badža, N. Šolaja, M. Đurić-Jovičić, V.S. Kostić, Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease—a review, Clin. Neurol. Neurosurg. 184 (2019) 105442.
[87] R.A. Ramdhani, A. Khojandi, O. Shylo, B.H. Kopell, Optimizing clinical assessments in Parkinson's disease through the use of wearable sensors and data driven modeling, Front. Comput. Neurosci. 12 (2018) 72.
[88] A. Metcalfe-Roach, A.C. Yu, E. Golz, M. Cirstea, K. Sundvick, D. Kliger, S. Appel-Cresswell, MIND and mediterranean diets associated with later onset of Parkinson's disease, Mov. Disord. 36 (4) (2021) 977–984.