Parkinsonism and Related Disorders 126 (2024) 107054

Contents lists available at ScienceDirect

Parkinsonism and Related Disorders

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

Levodopa-induced dyskinesia in Parkinson's disease: Insights from cross-cohort prognostic analysis using machine learning

Rebecca Ting Jiin Loo [a], Olena Tsurkalenko [b,c,d,e,f], Jochen Klucken [d,e,f], Graziella Mangone [g], Fouad Khoury [g], Marie Vidailhet [g], Jean-Christophe Corvol [g], Rejko Krüger [b,c,h], Enrico Glaab [a,*], on behalf ofthe NCER-PD Consortium

[a] Biomedical Data Science, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg
[b] Translational Neuroscience, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg
[c] Transversal Translational Medicine, Luxembourg Institute of Health (LIH), Strassen, Luxembourg
[d] Digital Medicine Group, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg
[e] Digital Medicine Group, Department of Precision Health, Luxembourg Institute of Health (LIH), Strassen, Luxembourg
[f] Digital Medicine Group, Centre Hospitalier de Luxembourg (CHL), Luxembourg
[g] Sorbonne Université, Paris Brain Institute - ICM, Inserm, CNRS, Assistance Publique Hôpitaux de Paris, Pitié-Salpêtrière Hospital, Department of Neurology, Paris, 75013, France
[h] Department of Neurology, Centre Hospitalier de Luxembourg (CHL), Luxembourg

ARTICLE INFO

Keywords:
Levodopa-induced dyskinesia
Longitudinal cohorts
Prognosis
Cross-cohort analysis
Machine learning
Predictive modeling

ABSTRACT

Background: Prolonged levodopa treatment in Parkinson's disease (PD) often leads to motor complications, including levodopa-induced dyskinesia (LID). Despite continuous levodopa treatment, some patients do not develop LID symptoms, even in later stages of the disease.
Objective: This study explores machine learning (ML) methods using baseline clinical characteristics to predict the development of LID in PD patients over four years, across multiple cohorts.
Methods: Using interpretable ML approaches, we analyzed clinical data from three independent longitudinal PD cohorts (LuxPARK, n = 356; PPMI, n = 484; ICEBERG, n = 113) to develop cross-cohort prognostic models and identify potential predictors for the development of LID. We examined cohort-specific and shared predictive factors, assessing model performance and stability through cross-validation analyses.
Results: Consistent cross-validation results for single and multiple cohort analyses highlighted the effectiveness of the ML models and identified baseline clinical characteristics with significant predictive value for the LID prognosis in PD. Predictors positively correlated with LID include axial symptoms, freezing of gait, and rigidity in the lower extremities. Conversely, the risk of developing LID was inversely associated with the occurrence of resting tremors, higher body weight, later onset of PD, and visuospatial abilities.
Conclusions: This study presents interpretable ML models for dyskinesia prognosis with significant predictive power in cross-cohort analyses. The models may pave the way for proactive interventions against dyskinesia in PD by optimizing levodopa dosing regimens and adjunct treatments with dopamine agonists or MAO-B inhibitors, and by employing non-pharmacological interventions such as dietary adjustments affecting levodopa absorption for high-risk LID patients.

1. Introduction

Levodopa is a drug commonly prescribed for the treatment of Parkinson's disease (PD) [1]. It relieves motor symptoms [2], but prolonged use can lead to motor complications, including levodopa-induced dyskinesia (LID) [3]. These hyperkinetic movements can be attributed to the development of abnormal pre- and post-synaptic plasticity in the basal ganglia network induced by levodopa in the context of degeneration of nigral neurons [4]. Although delaying levodopa treatment has been suggested to reduce the risk of LID [5], evidence supporting this strategy remains inconclusive.

LID affects approximately 30% of PD patients within the initial five

* Corresponding author.
E-mail address: enrico.glaab@uni.lu (E. Glaab).

https://doi.org/10.1016/j.parkreldis.2024.107054
Received 28 March 2024; Received in revised form 29 June 2024; Accepted 2 July 2024
Available online 4 July 2024

to six years of levodopa treatment [2]. In a smaller subgroup, however, 20–30% of treated patients will experience LID symptoms within an average of 20.5 months [6]. After ten years of continuous treatment, the majority of individuals report experiencing adverse effects associated with LID [4]. The impact of LID on quality of life is significant, affecting daily activities, communication and increasing the risk of falls [6]. However, the factors that determine why the delay in the onset of LID varies widely among PD patients on similar treatment regimens remain elusive. Identifying risk factors that predispose patients to develop LID could pave the way for the design of personalized medicine approaches to improve prevention and early treatment efficacy by individually adjusting pharmacological and non-pharmacological interventions (e.g., preferred use of controlled release formulations of levodopa for patients at high risk of developing LID [2], early consideration of Deep Brain Stimulation as an alternative treatment option [1], or guiding patients on dietary adjustments that can affect levodopa absorption and metabolism [7]).

While previous studies have improved our understanding of the factors that influence LID development in PD, they have primarily focused on single-cohort analyses, which may have potential cohort-specific biases. Furthermore, most prior studies did not optimize the predictive models for sparsity and interpretability, and no testing of the robustness and reproducibility in a cross-study setting was performed. To help address these gaps, our study aims to construct prognostic models for LID risk using baseline clinical characteristics from three distinct PD cohorts, focusing on sparse and interpretable modeling approaches to complement the previous investigations and increase model robustness and explainability. The main goal is to obtain prognostic models that are biologically plausible, independently confirmed in different cohorts, and suitable for future cross-study applications.

The application of machine learning (ML) techniques enables the creation of accurate, multivariate prognostic risk models that can identify multiple interrelated risk factors [8]. We also address challenges such as systematic biases in cross-cohort studies by using cross-cohort normalization methods to ensure the reliability and robustness of the results [8]. This work establishes the foundation for future studies on early preventive and therapeutic interventions against LID in PD.

2. Methods

2.1. Inclusion criteria

PD patients were evaluated using the Movement Disorders Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS), which includes assessing motor complications and identifying dyskinesia symptoms during clinical visits. LID occurrence was defined as a score $\geq$1 on MDS-UPDRS Part IV, item 4.1 (time spent with dyskinesias), item 4.2 (functional impact of dyskinesias), or by the presence of dyskinesia identified in the clinical motor examination. If any relevant assessments indicated the presence of dyskinesia during a clinical visit, the corresponding patient was defined to have PD with LID. As the evaluation of LID was integrated into the MDS-UPDRS Part IV score, we excluded the total MDS-UPDRS Part IV score from any further analysis. Importantly, the LuxPARK cohort participants were assessed solely during the ON state. Due to this limitation, all analyses focused only on data obtained during the ON stage, and assessments in the OFF state from other cohorts were excluded.

To evaluate potential factors influencing LID symptom occurrence, we extracted baseline clinical features and longitudinal LID status records for PD patients from all three cohorts, covering up to four years from baseline. We analyzed the data over this four-year period for classification analysis, ensuring consistency and comparability across all cohorts, in alignment with the four-year follow-up available in the ICEBERG cohort. This timeframe corresponds with clinical practice, where significant changes in treatment response and complications often occur within the first few years of therapy. However, focusing solely on this period may limit the study by not capturing long-term trends and outcomes beyond four years. To mitigate this limitation, we conducted a time-to-event analysis, which assesses events and outcomes over varying time frames. This approach provides a more comprehensive understanding of the data, despite the uniform follow-up period used for the classification analysis. Specifically, in the time-to-event analysis, the duration of follow-up varied among patients depending on their total time of participation within each cohort study. The outcome was defined as the duration until the LID event occurred, or until the last follow-up, if the event was censored.

The inclusion criteria were: (1) a diagnosis of PD according to the current criteria by the International Movement Disorder Society; (2) patients with LID, defined as presenting with LID symptoms within four years of the baseline clinical visit (LID+), and patients without LID, defined as having recorded assessments showing no LID symptoms during these four years (LID-). Importantly, patients with LID at baseline were excluded in the test set evaluations for ML. This uniform approach for all cohorts helped to ensure that the considered subjects were fully aligned regarding the criteria for PD diagnosis and LID assessment. Further information on the cohort characteristics is provided in the Suppl. Material (section 1).

The number of PD patients who met the two inclusion criteria for the classification and survival analyses are displayed in Table 1. The 'Events' columns indicate the total number and percentage of subjects who developed LID during the considered timeframe (up to 4 years follow-up for LID classification and up to the last available follow-up visit for each patient for the time-to-LID analysis). A comprehensive overview of the demographic and baseline clinical characteristics for all subjects covered in the cross-cohort analysis is shown in Table 2. Additional cohort-specific descriptive statistics for the LuxPARK, PPMI, and ICEBERG cohorts are provided in Suppl. Tables 2–4. It is important to recognize that these baseline characteristics reflect characteristics during the first clinical visit, which generally does not correspond to the onset of the disease. As a result, patients may have different disease durations at baseline, and we therefore consider disease duration as a key variable in our analyses.

2.2. Machine learning analysis for LID classification

Using the clinical data from the three patient cohorts, we first applied pre-processing and normalization as described in the Suppl. Material (section 2) and then performed supervised ML methods, focusing on interpretable, rule-based approaches, to predict the occurrence of LID in PD within the next 4 years. In total, nine tree-based ML algorithms for classification were evaluated, including Adaptive Boosting (AdaBoost) [9], Classification and Regression Trees (CART), Category Boosting (CatBoost), C4.5 trees [10], Fast Interpretable Greedy-Tree Sums (FIGS) [11], Fast-Sparse Decision Tree (GOSDT-GUESSES) [12], Gradient Boosting (GBoost) [13], Hierarchical Shrinkage (HS) [14], and Extreme Gradient Boosting (XGBoost). These classification algorithms were used to build models for LID prognosis from the baseline clinical information of PD patients and to assess the most predictive features as candidate risk

Table 1
Number of patients who met the inclusion criteria and distribution of events in the LuxPARK, PPMI, and ICEBERG cohorts.

Cohort	Inclusion criteria (1)	Inclusion criteria (2)	Events (LID Classification)	Events (Time-to-LID)
LuxPARK	706	356	210 (59.0 %)	222 (62.4 %)
PPMI	796	484	173 (35.7 %)	348 (71.9 %)
ICEBERG	162	113	36 (31.9 %)	36 (31.9 %)
Total	1664	953	419 (44.0 %)	606 (63.6 %)

Table 2
Overview of the demographic and baseline clinical characteristics for all subjects covered in the cross-cohort analysis during a 4-year follow-up period. Subjects who developed LID are indicated by a "+" sign (LID+) and subjects who did not develop LID by a "-" sign (LID-). P-values for the significance of differences between the LID- and LID + groups for individual features are shown in the last column.

Baseline feature	Statistics	All samples	LID -	LID +	P-values
Number of patients	N	953	534 (56.0 %)	419 (44.0 %)	
Demographic and general information:					
Age of onset	n	947 (99.4 %)	532 (55.8 %)	415 (43.5 %)	7.72E-12
	Mean (SD)	58.2 (10.80)	60.4 (10.16)	55.5 (11.00)	
Disease duration	n	947 (99.4 %)	532 (55.8 %)	415 (43.5 %)	7.66E-38
	Mean (SD)	4.5 (5.56)	2.8 (4.28)	6.7 (6.24)	
Levodopa treatment	n	952 (99.9 %)	533 (55.9 %)	419 (44.0 %)	2.57E-38
	No	493 (51.7 %)	374 (39.2 %)	119 (12.5 %)	
	Yes	459 (48.2 %)	159 (16.7 %)	300 (31.5 %)	
Weight (kg)	n	947 (99.4 %)	533 (55.9 %)	414 (43.4 %)	5.38E-03
	Mean (SD)	78.2 (16.29)	79.5 (16.39)	76.5 (16.01)	
Height (cm)	n	947 (99.4 %)	532 (55.8 %)	415 (43.5 %)	4.29E-05
	Mean (SD)	170.4 (9.95)	171.6 (9.79)	168.9 (9.95)	
BMI (kg/m2)	n	946 (99.3 %)	532 (55.8 %)	414 (43.4 %)	0.440
	Mean (SD)	26.8 (4.59)	26.9 (4.51)	26.7 (4.68)	
Neuropsychological assessments:					
Benton Judgement of Line Orientation	n	828 (86.9 %)	455 (47.7 %)	373 (39.1 %)	1.66E-10
	Mean (SD)	11.4 (8.24)	13.0 (7.25)	9.5 (8.95)	
Non-motor, motor and disability:					
Hoehn & Yahr stage	n	949 (99.6 %)	531 (55.7 %)	418 (43.9 %)	5.00E-04
	Stage 0	4 (0.4 %)	3 (0.3 %)	1 (0.1 %)	
	Stage 1	223 (23.4 %)	176 (18.5 %)	47 (4.9 %)	
	Stage 2	643 (67.5 %)	327 (34.3 %)	316 (33.2 %)	
	Stage 3	61 (6.4 %)	21 (2.2 %)	40 (4.2 %)	
	Stage 4	13 (1.4 %)	2 (0.2 %)	11 (1.2 %)	
	Stage 5	5 (0.5 %)	2 (0.2 %)	3 (0.3 %)	
Axial symptoms	n	951 (99.8 %)	533 (55.9 %)	418 (43.9 %)	3.99E-30
	Mean (SD)	3.5 (3.90)	2.3 (3.03)	5.0 (4.35)	
Selective axial symptoms	n	951 (99.8 %)	533 (55.9 %)	418 (43.9 %)	1.38E-20
	Mean (SD)	2.5 (2.72)	1.8 (2.41)	3.3 (2.87)	
Motor fluctuation composite	n	647 (67.9 %)	293 (30.7 %)	354 (37.1 %)	1.14E-24
	Mean (SD)	1.4 (2.36)	0.4 (1.23)	2.2 (2.74)	
Freezing of gait	n	951 (99.8 %)	533 (55.9 %)	418 (43.9 %)	2.07E-22
	Mean (SD)	0.4 (1.11)	0.2 (0.69)	0.8 (1.41)	
Resting tremor	n	579 (60.8 %)	247 (25.9 %)	332 (34.8 %)	0.016
	Mean (SD)	3.6 (3.37)	3.9 (3.27)	3.4 (3.42)	
Tremor	n	579 (60.8 %)	247 (25.9 %)	332 (34.8 %)	2.10E-03
	Mean (SD)	5.5 (4.39)	6.0 (4.20)	5.2 (4.50)	
Bradykinesia	n	579 (60.8 %)	247 (25.9 %)	332 (34.8 %)	4.03E-03
	Mean (SD)	14.8 (8.34)	13.5 (7.45)	15.8 (8.83)	
Rigidity lower extremities	n	578 (60.7 %)	247 (25.9 %)	331 (34.7 %)	1.65E-04
	Mean (SD)	1.7 (1.68)	1.4 (1.67)	1.9 (1.66)	
Rigidity upper extremities	n	578 (60.7 %)	247 (25.9 %)	331 (34.7 %)	0.710
	Mean (SD)	2.5 (1.55)	2.4 (1.44)	2.5 (1.64)	
Initial motor symptom - Rigidity or bradykinesia	n	949 (99.6 %)	532 (55.8 %)	417 (43.8 %)	0.489
	No	225 (23.6 %)	131 (13.7 %)	94 (9.9 %)	
	Yes	724 (76.0 %)	401 (42.1 %)	323 (33.9 %)	
Modified Schwab & England ADL	n	594 (62.3 %)	386 (40.5 %)	208 (21.8 %)	2.10E-08
	Mean (SD)	92.0 (7.44)	93.4 (5.93)	89.4 (9.09)	
MDS-UPDRS Part I score	n	939 (98.5 %)	527 (55.3 %)	412 (43.2 %)	7.79E-19
	Mean (SD)	6.7 (6.52)	5.0 (5.15)	8.9 (7.35)	
MDS-UPDRS Part II score	n	941 (98.7 %)	528 (55.4 %)	413 (43.3 %)	1.10E-28
	Mean (SD)	8.7 (6.66)	6.6 (5.06)	11.4 (7.43)	
MDS-UPDRS Part III score (ON)	n	572 (60.0 %)	243 (25.5 %)	329 (34.5 %)	0.014
	Mean (SD)	28.5 (14.69)	26.6 (13.27)	29.9 (15.53)	
MDS-UPDRS I - Apathy	n	944 (99.1 %)	529 (55.5 %)	415 (43.5 %)	5.00E-04
	Normal	714 (74.9 %)	436 (45.8 %)	278 (29.2 %)	

(*continued on next page*)

Table 2 (*continued*)

Baseline feature	Statistics	All samples	LID -	LID +	P-values
	Slight	149 (15.6 %)	66 (6.9 %)	83 (8.7 %)	
	Mild	63 (6.6 %)	22 (2.3 %)	41 (4.3 %)	
	Moderate	13 (1.4 %)	4 (0.4 %)	9 (0.9 %)	
	Severe	5 (0.5 %)	1 (0.1 %)	4 (0.4 %)	
MDS-UPDRS I - Depressed mood	n	945 (99.2 %)	529 (55.5 %)	416 (43.7 %)	5.00E-04
	Normal	602 (63.2 %)	364 (38.2 %)	238 (25.0 %)	
	Slight	247 (25.9 %)	134 (14.1 %)	113 (11.9 %)	
	Mild	71 (7.5 %)	25 (2.6 %)	46 (4.8 %)	
	Moderate	18 (1.9 %)	6 (0.6 %)	12 (1.3 %)	
	Severe	7 (0.7 %)	0 (0.0 %)	7 (0.7 %)	
MDS-UPDRS I - Sleep problems (night)	n	944 (99.1 %)	530 (55.6 %)	414 (43.4 %)	5.00E-04
	Normal	341 (35.8 %)	223 (23.4 %)	118 (12.4 %)	
	Slight	237 (24.9 %)	143 (15.0 %)	94 (9.9 %)	
	Mild	178 (18.7 %)	101 (10.6 %)	77 (8.1 %)	
	Moderate	141 (14.8 %)	53 (5.6 %)	88 (9.2 %)	
	Severe	47 (4.9 %)	10 (1.0 %)	37 (3.9 %)	
MDS-UPDRS I - Urinary problems	n	944 (99.1 %)	530 (55.6 %)	414 (43.4 %)	6.00E-03
	Normal	415 (43.5 %)	250 (26.2 %)	165 (17.3 %)	
	Slight	337 (35.4 %)	191 (20.0 %)	146 (15.3 %)	
	Mild	125 (13.1 %)	65 (6.8 %)	60 (6.3 %)	
	Moderate	54 (5.7 %)	18 (1.9 %)	36 (3.8 %)	
	Severe	13 (1.4 %)	6 (0.6 %)	7 (0.7 %)	
MDS-UPDRS II - Freezing	n	944 (99.1 %)	530 (55.6 %)	414 (43.4 %)	5.00E-04
	Normal	759 (79.6 %)	483 (50.7 %)	276 (29.0 %)	
	Slight	101 (10.6 %)	38 (4.0 %)	63 (6.6 %)	
	Mild	41 (4.3 %)	4 (0.4 %)	37 (3.9 %)	
	Moderate	35 (3.7 %)	4 (0.4 %)	31 (3.3 %)	
	Severe	8 (0.8 %)	1 (0.1 %)	7 (0.7 %)	
MDS-UPDRS II - Hygiene	n	943 (99.0 %)	530 (55.6 %)	413 (43.3 %)	5.00E-04
	Normal	583 (61.2 %)	366 (38.4 %)	217 (22.8 %)	
	Slight	312 (32.7 %)	151 (15.8 %)	161 (16.9 %)	
	Mild	41 (4.3 %)	13 (1.4 %)	28 (2.9 %)	
	Moderate	5 (0.5 %)	0 (0.0 %)	5 (0.5 %)	
	Severe	2 (0.2 %)	0 (0.0 %)	2 (0.2 %)	
MDS-UPDRS II - Saliva and drooling	n	944 (99.1 %)	530 (55.6 %)	414 (43.4 %)	5.00E-04
	Normal	527 (55.3 %)	329 (34.5 %)	198 (20.8 %)	
	Slight	126 (13.2 %)	69 (7.2 %)	57 (6.0 %)	
	Mild	190 (19.9 %)	86 (9.0 %)	104 (10.9 %)	
	Moderate	85 (8.9 %)	41 (4.3 %)	44 (4.6 %)	
	Severe	16 (1.7 %)	5 (0.5 %)	11 (1.2 %)	
MDS-UPDRS II - Tremor	n	944 (99.1 %)	530 (55.6 %)	414 (43.4 %)	5.00E-04
	Normal	211 (22.1 %)	101 (10.6 %)	110 (11.5 %)	
	Slight	536 (56.2 %)	335 (35.2 %)	201 (21.1 %)	
	Mild	161 (16.9 %)	83 (8.7 %)	78 (8.2 %)	
	Moderate	28 (2.9 %)	7 (0.7 %)	21 (2.2 %)	
	Severe	8 (0.8 %)	4 (0.4 %)	4 (0.4 %)	
Autonomic function: SCOPA-AUT Gastrointestinal (GI)	n	939 (98.5 %)	528 (55.4 %)	411 (43.1 %)	1.15E-18
	Mean (SD)	3.3 (2.79)	2.6 (2.43)	4.2 (2.96)	
SCOPA-AUT Thermoregulatory	n	939 (98.5 %)	529 (55.5 %)	410 (43.0 %)	6.67E-14
	Mean (SD)	2.0 (2.09)	1.5 (1.63)	2.6 (2.42)	
SCOPA-AUT Urinary	n	947 (99.4 %)	532 (55.8 %)	415 (43.5 %)	1.51E-04
	Mean (SD)	4.7 (3.30)	4.3 (3.09)	5.2 (3.50)	
Gene mutation: *GBA* mutation	n	842 (88.4 %)	458 (48.1 %)	384 (40.3 %)	0.110
	No	741 (77.8 %)	411 (43.1 %)	330 (34.6 %)	
	Yes	101 (10.6 %)	47 (4.9 %)	54 (5.7 %)	

and protective factors for LID.

2.3. Time-to-event machine learning analysis

We conducted time-to-event analyses as a complementary method to identify predictive clinical features associated with the risk of LID development. This approach is commonly used in biomedical research to examine the effect of clinical factors on events monitored over time.

The data is subject to censoring, meaning that some patients may not have experienced the event of interest (LID development) by the end of the study period, resulting in incomplete observations [15]. To address this, we applied the following methods for time-to-event analysis: component-wise Gradient Boosting (CW-GBoost) [16], Survival Trees [17], Extra Survival Trees [18], Survival Gradient Boosting (Survival GBoost) [16], Survival Linear Support Vector Machine (LSVM) [15], Naive LSVM (NLSVM), penalized Cox regression [15], and Random Survival Forests (Survival RF) [17].

2.4. Model evaluation

To quantitatively assess different ML approaches and how clinical predictors relate to the risk and timing of LID symptoms following the initial clinical visit, we used a cross-validation (CV) workflow (see Suppl. Material section 3 and Suppl. Fig. 1), involving the learning algorithms detailed in sections 2.2 and 2.3. For the integrative analyses of multiple cohorts, we focused on those clinical features as candidate predictors shared between the LuxPARK, PPMI, and ICEBERG datasets.

Two distinct prediction models were developed, here referred to as comprehensive and refined models. The comprehensive model included all baseline clinical features shared across the cohorts and was trained without prior feature selection. By contrast, the refined model consisted of a subset of clinical features obtained by excluding baseline dyskinesia and levodopa medication. The motivation for building this refined model was to better identify and assess clinical factors strongly associated with dyskinesia development that are independent of levodopa use and of interest as potential risk factors or protective factors for LID development. Both models were assessed in identical training and testing workflows. To interpret the LID prediction models, we performed a SHAP (SHapley Additive exPlanations) value analysis [19]. This model-agnostic procedure allows *post hoc* interpretation, regardless of the underlying model [19]. It quantifies the predictive value of individual features and their influence on outcome predictions.

We evaluated the predictive ability of ML models for LID classification and time-to-LID risk modeling using the Area Under the Curve (AUC) [20] and the Concordance index (C-index) [21] as performance measures for both CV analyses (averaging the results across the CV cycles) and independent testing. Furthermore, to statistically assess and compare the performance scores of the optimized comprehensive and refined models, DeLong's test [20], along with its adaptation by Kang et al. [21], known as the one-shot nonparametric approach, was applied to the hold-out test set. The performance scores between optimized unnormalized and normalized models for all normalization methods were also compared. The obtained p-values from the comparisons within the same cohorts were adjusted using the Benjamini-Hochberg method to address multiple hypothesis testing. To quantify the stability of the model predictions, the standard deviations (SD) of the performance metrics across the CV cycles were computed. Next, we performed decision curve analyses (DCA) to evaluate the clinical utility of the predictive models for LID classification and time-to-LID analysis, assessing the net benefit in clinical decision-making [22]. In addition, calibration analyses were performed to measure the slope and mean squared error (MSE) of the models, quantifying calibration performance by comparing predicted probabilities with actual outcomes for the LID classification model and comparing the predicted survival probabilities with the observed survival probabilities for the time-to-LID model. Overall, the combination of these analyses allowed us to evaluate the prediction models' generalization performance, robustness, and utility for clinical decision-making.

2.5. Comparison of selected features across different cohorts

To compare the feature selection results across the single-cohort analyses, we calculated statistics to identify features consistently chosen as LID predictors across different cohorts and methods. Before feature selection, categorical variables were one-hot encoded into binary features. If multiple features derived from the same original categorical variable were selected within a CV fold, they were treated as a single feature to prevent duplication. First, the percentage of times each candidate feature was selected in each of the 5 CV folds for each cohort was computed. Next, the average percentage was calculated for each cohort, providing a consolidated measure of the predictive utility of each feature. These average percentages were compared across the three cohorts for the optimized comprehensive and refined models for LID classification and time-to-LID analysis, aiming to identify consistent LID predictors across different methodologies and cohorts. This provides insight into their generalizability and potential applicability as biomarkers for cross-cohort LID prediction. This feature selection process was performed separately for the classification and time-to-event analyses.

2.6. Statistical analyses

We applied univariate hypothesis tests to explore potential statistical associations between clinical measurements at baseline and the occurrence of LID within the 4-year visit. When the normality assumption was not met, the Mann-Whitney *U* test was applied to compare continuous variables between independent groups, and the two-sample *t*-test was used for normally distributed variables. Categorical variables were analyzed using Fisher's exact test. To assess the differences between continuous variables across the three cohorts, ANOVA was used when the normality assumption was met; otherwise, the Kruskal-Wallis test was applied. Spearman's rank correlation coefficient, a nonparametric correlation measure, was used to assess monotonic relationships between variables. Findings were considered statistically significant if the associated p-value was below 0.05.

3. Results

3.1. Individual cohort analyses

When evaluating the predictive performance of ML models for LID prognostic classification, we observed significant differences between the results for individual cohorts. In the 5-fold CV on the training data, using the modeling approaches described in sections 2.2 and 2.3, the average AUC values for the optimized comprehensive model ranged from 0.595 (SD 0.144) in ICEBERG to 0.735 (SD 0.091) in the LuxPARK study, and similarly, for the hold-out data, from 0.533 in ICEBERG to 0.678 in LuxPARK (see Table 3). Comparable trends across the cohorts were also observed for the time-to-LID analysis, with an average cross-validated C-index ranging from 0.576 (SD 0.115) in ICEBERG to 0.714 (SD 0.027) in LuxPARK, and a hold-out C-index ranging from 0.451 for ICEBERG to 0.557 for LuxPARK (see Table 4). The performance scores for the PPMI study fell between those of the ICEBERG and LuxPARK studies. In contrast to the CV results, a lower hold-out AUC was observed for the optimized comprehensive prognostic model in the ICEBERG study. However, for most other combinations of modeling approaches and cohorts, the CV and hold-out test set results were similar, indicating the overall robustness and consistency of the performance estimates.

LuxPARK consistently achieved the highest predictive performance in the classification and time-to-LID analyses, with PPMI and ICEBERG following in second and third place, respectively. The refined model achieved significantly higher predictive performance in the LID classification for PPMI and time-to-LID analysis for LuxPARK compared to the comprehensive model (see Table 5). Finally, when assessing the stability of the comprehensive and refined models for LID prognosis and time-to-LID analyses, it was comparable for LuxPARK and PPMI (see Suppl. Tables 7 and 8), but lower for ICEBERG, indicating more consistent and stable performance for the cohorts with a larger sample size. Predictors with a high selection frequency across the CV cycles for the different optimized models are presented in Table 6.

Table 3
Overview of predictive performance statistics for comprehensive LID prognostic classification, including cross-validated and hold-out AUC values for single and multi-cohort analyses and the corresponding number of features used in each model. The models with the highest average AUC scores in the cross-validation of the cohort analyses are highlighted in bold. For the column "Number of features", the number in brackets represents the count of candidate features selected during cross-validation, whereas the number in front of the brackets indicates the number of selected features with predictive influence, as determined through permutation importance analysis.

Single-cohort analyses:									
Algorithm	LuxPARK			PPMI			ICEBERG		
	Mean (SD)	Hold-out AUC	Number of features	Mean (SD)	Hold-out AUC	Number of features	Mean (SD)	Hold-out AUC	Number of features
AdaBoost	0.647 (0.079)	0.532	4 (6)	0.623 (0.059)	0.582	2 (6)	0.481 (0.122)	0.512	1 (2)
CART	0.610 (0.059)	0.561	2 (3)	0.607 (0.043)	0.555	9 (10)	0.509 (0.069)	0.512	1 (2)
CatBoost	0.664 (0.068)	0.577	5 (6)	0.626 (0.060)	0.656	14 (24)	0.546 (0.078)	0.652	8 (16)
C4.5	0.607 (0.069)	0.559	2 (7)	**0.629 (0.071)**	**0.570**	**5 (11)**	0.540 (0.097)	0.390	1 (6)
FIGS	0.584 (0.121)	0.539	7 (11)	0.602 (0.070)	0.557	3 (9)	0.482 (0.092)	0.512	1 (2)
GOSDT-GUESSES	0.612 (0.132)	0.562	10 (10)	0.592 (0.057)	0.521	21 (36)	0.521 (0.197)	0.448	8 (10)
GBoost	0.650 (0.089)	0.641	16 (24)	0.628 (0.032)	0.621	17 (24)	0.533 (0.073)	0.607	6 (7)
HS	0.620 (0.061)	0.530	5 (9)	0.597 (0.077)	0.557	3 (9)	0.482 (0.092)	0.512	1 (2)
XGBoost	**0.735 (0.091)**	**0.678**	**37 (59)**	0.602 (0.052)	0.621	9 (9)	**0.595 (0.144)**	**0.533**	**13 (14)**
Multi-cohort analyses:									
Algorithm	**CROSS-COHORT**			**LEAVE-ICEBERG-OUT**			**LEAVE-PPMI-OUT**		
	Mean (SD)	**Hold-out AUC**	**Number of features**	**Mean (SD)**	**Hold-out AUC**	**Number of features**	**Mean (SD)**	**Hold-out AUC**	**Number of features**
AdaBoost	**0.682 (0.080)**	**0.646**	**8 (14)**	0.691 (0.052)	0.664	4 (7)	0.667 (0.039)	0.528	4 (5)
CART	0.662 (0.059)	0.647	4 (7)	0.671 (0.016)	0.535	2 (4)	0.648 (0.057)	0.518	15 (21)
CatBoost	0.675 (0.062)	0.702	9 (17)	0.697 (0.045)	0.600	18 (27)	0.675 (0.099)	0.626	8 (18)
C4.5	0.631 (0.111)	0.664	5 (9)	0.669 (0.026)	0.535	2 (4)	0.628 (0.034)	0.627	1 (3)
FIGS	0.659 (0.059)	0.673	3 (5)	0.694 (0.040)	0.606	9 (20)	0.659 (0.051)	0.594	9 (20)
GOSDT-GUESSES	0.622 (0.064)	0.613	42 (59)	0.644 (0.018)	0.547	38 (53)	0.623 (0.070)	0.542	27 (43)
GBoost	0.656 (0.054)	0.672	5 (5)	**0.699 (0.016)**	**0.513**	**15 (26)**	0.670 (0.042)	0.619	14 (21)
HS	0.660 (0.033)	0.664	2 (3)	0.692 (0.039)	0.606	9 (20)	0.659 (0.051)	0.594	9 (20)
XGBoost	0.654 (0.032)	0.632	52 (66)	0.690 (0.036)	0.631	35 (64)	**0.690 (0.040)**	**0.582**	**58 (74)**

3.2. Cross-cohort analyses

In the integrated ML analysis of all three cohorts, the optimized comprehensive model achieved an average cross-validated AUC of 0.682 (SD 0.080) for LID prognostic classification and a C-index of 0.718 (SD 0.052) for time-to-LID analysis (see bottom of Tables 3 and 4). Despite the additional challenges of integrating information from diverse cohorts with potential study-specific biases, the cross-cohort model achieved comparable performance statistics to the single-cohort analyses, while the integrated model has the added value of being applicable across diverse cohorts with a higher expected generalization capability across distinct populations. Similarly, the refined cross-cohort model achieved performance statistics for the average cross-validated AUC of 0.688 (SD 0.043) and the hold-out test set AUC of 0.639 (see Suppl. Table 5), with no significant difference in hold-out predictive performance compared to the comprehensive LID classification model (see Table 5). However, a significant difference in predictive performance for the hold-out test set was observed between the comprehensive and refined time-to-LID models (see Table 5). The refined time-to-LID models had a significantly ($p = 0.019$) higher hold-out C-index of 0.685 (average cross-validated C-index of 0.715; SD 0.054), compared to the comprehensive model with a hold-out C-index of 0.627 (average cross-validated C-index of 0.718; SD 0.052) (see Table 4 and Suppl. Table 6).

Interestingly, among the multi-cohort analyses, the cross-cohort models, which were trained on subsets of the data for all three cohorts, provided superior predictive performances compared to the leave-one-cohort-out analyses, trained on two cohorts (see Tables 3–4 and Suppl. Tables 5–6). Furthermore, considering the stability of prediction results, the cross-cohort model also provided superior results compared to the single-cohort models, highlighting the benefits of training predictive models for LID prognosis on data from multiple diverse cohorts (see Suppl. Tables 7 and 8).

3.3. Comparative evaluation of cross-study integration methods

The cross-cohort comprehensive and refined LID prognostic models showed no significant difference in hold-out test set predictive performance between normalized and unnormalized versions (see Table 5 and Suppl. Table 9). However, the comprehensive classification and refined

Table 4
Overview of predictive performance statistics for the comprehensive time-to-LID models, including cross-validated and hold-out C-indices for single and multi-cohort analysis, and the corresponding number of features used in each model. The models with the highest average C-indices in the cross-validation analyses of each cohort are highlighted in bold. In the column "Number of features" the number in brackets represents the count of candidate features selected during cross-validation, whereas the number in front of the brackets indicates the number of selected features with predictive influence, as determined through permutation importance analysis.

Single-cohort analyses:									
Algorithm	LuxPARK			PPMI			ICEBERG		
	Mean (SD)	Hold-out C-index	Number of features	Mean (SD)	Hold-out C-index	Number of features	Mean (SD)	Hold-out C-index	Number of features
CW-GBoost	0.667 (0.055)	0.651	13 (17)	0.696 (0.033)	0.640	14 (21)	**0.576 (0.115)**	**0.451**	**4 (11)**
Extra Survival	0.699 (0.101)	0.610	11 (11)	0.694 (0.054)	0.651	70 (97)	0.548 (0.088)	0.542	9 (9)
Survival GBoost	0.648 (0.069)	0.647	119 (120)	0.669 (0.074)	0.643	17 (22)	0.570 (0.132)	0.585	49 (58)
LSVM	0.628 (0.043)	0.614	15 (15)	0.670 (0.040)	0.652	35 (35)	0.531 (0.116)	0.557	103 (104)
NLSVM	0.642 (0.017)	0.618	20 (20)	0.673 (0.045)	0.650	29 (29)	0.561 (0.142)	0.452	10 (10)
Penalized Cox	0.685 (0.046)	0.532	1 (5)	**0.697 (0.041)**	**0.663**	**28 (51)**	0.551 (0.092)	0.517	4 (4)
Survival RF	**0.714 (0.027)**	**0.577**	**11 (11)**	0.672 (0.061)	0.650	46 (72)	0.543 (0.114)	0.549	10 (33)
Survival Trees	0.613 (0.052)	0.582	4 (7)	0.644 (0.098)	0.622	8 (12)	0.564 (0.087)	0.498	1 (3)
Multi-cohort analyses:									
Algorithm	**CROSS-COHORT**			**LEAVE-ICEBERG-OUT**			**LEAVE-PPMI-OUT**		
	Mean (SD)	**Hold-out C-index**	**Number of features**	**Mean (SD)**	**Hold-out C-index**	**Number of features**	**Mean (SD)**	**Hold-out C-index**	**Number of features**
CW-GBoost	0.712 (0.045)	0.673	14 (25)	0.696 (0.049)	0.639	13 (23)	0.702 (0.022)	0.613	10 (16)
Extra Survival	0.704 (0.045)	0.667	160 (161)	0.698 (0.050)	0.605	160 (161)	0.696 (0.049)	0.651	12 (12)
Survival GBoost	0.704 (0.061)	0.661	24 (49)	0.686 (0.046)	0.684	11 (21)	**0.719 (0.029)**	**0.655**	**13 (22)**
LSVM	**0.718 (0.052)**	**0.627**	**36 (36)**	0.681 (0.048)	0.547	52 (52)	0.701 (0.025)	0.639	147 (147)
NLSVM	0.701 (0.068)	0.652	52 (52)	0.681 (0.052)	0.681	56 (56)	0.705 (0.016)	0.669	36 (36)
Penalized Cox	0.692 (0.055)	0.666	28 (28)	**0.701 (0.056)**	**0.531**	**57 (96)**	0.691 (0.077)	0.551	1 (32)
Survival RF	0.705 (0.055)	0.682	134 (139)	0.701 (0.048)	0.612	104 (136)	0.689 (0.036)	0.663	116 (122)
Survival Trees	0.649 (0.078)	0.646	8 (13)	0.661 (0.042)	0.491	13 (16)	0.654 (0.048)	0.512	11 (17)

Table 5
Overview of the statistical significance of the differences between the hold-out predictive performance metrics for the optimized comprehensive and refined models across cohorts and the optimized cross-study normalized and unnormalized models. The normalization method used is indicated in the column "Normalization". The p-values for the significance of the difference were calculated using DeLong's test for LID classification (at the top) and the one-shot nonparametric test for time-to-LID analysis (bottom). The comprehensive model outperforms the refined model in terms of predictive performance, with a statistically significant difference in the hold-out AUC/C-index.

		Comprehensive model		Refined model	
Cohort	Comprehensive vs Refined	Normalized vs Unnormalized	Normalization	Normalized vs Unnormalized	Normalization
LID classification:					
LuxPARK	0.925	–	–	–	–
PPMI	0.036	–	–	–	–
ICEBERG	0.776	–	–	–	–
CROSS-COHORT	0.092	0.912	Mean-centering	0.683	Quantile
LEAVE-ICEBERG-OUT	0.399	0.549	M-ComBat	0.981	Mean-centering
LEAVE-PPMI-OUT	0.954	0.965	Ratio-A	0.719	Ratio-A
Time-to-LID:					
LuxPARK	0.020	–	–	–	–
PPMI	0.908	–	–	–	–
ICEBERG	0.367	–	–	–	–
CROSS-COHORT	0.019	0.096	Standardize	0.527	Mean-centering
LEAVE-ICEBERG-OUT	1.000	0.298	Quantile	0.867	Quantile
LEAVE-PPMI-OUT	0.235	0.019	Ratio-A	0.955	Mean-centering

Table 6
Overview of statistics on the average percentage of times clinical features were selected in 5-fold cross-validation analyses. The table compares the statistics for the comprehensive and refined models for LID classification and time-to-LID analyses in LuxPARK, PPMI, and ICEBERG cohorts. The different columns contain the following information: Average in CV (%): The average percentage of times each feature was selected in the 5-fold CV in single-cohort analyses in LuxPARK, PPMI, and ICEBERG for both LID and time-to-LID analyses. Average (%): The mean of the 'Average in CV (%)' for LID and time-to-LID analyses across the cohorts. The list of features was arranged in descending order according to the overall average selection percentages for both comprehensive and refined models in LID and time-to-LID analyses, and the top 15 features with the highest average percentages are shown.

Description	Comprehensive model			Refined model		
	LID Classification	Time-to-LID	Overall	LID Classification	Time-to-LID	Overall
	Average in CV (%)	Average in CV (%)	Average (%)	Average in CV (%)	Average in CV (%)	Average (%)
Disease duration	86.7	46.7	66.7	80.0	80.0	80.0
Age of onset	73.3	53.3	63.3	80.0	80.0	80.0
MDS-UPDRS I - Urinary problems	53.3	46.7	50.0	46.7	93.3	70.0
MDS-UPDRS I - Sleep problems (night)	66.7	53.3	60.0	46.7	73.3	60.0
BMI (kg/m2)	66.7	33.3	50.0	60.0	66.7	63.3
MDS-UPDRS Part II score	60.0	46.7	53.3	60.0	60.0	60.0
Benton Judgment of Line Orientation score	46.7	46.7	46.7	53.3	66.7	60.0
MDS-UPDRS Part I score	53.3	46.7	50.0	40.0	73.3	56.7
SCOPA-AUT Gastrointestinal (GI) score	66.7	40.0	53.3	26.7	73.3	50.0
Axial symptoms score	33.3	46.7	40.0	46.7	80.0	63.3
Weight (kg)	46.7	40.0	43.3	33.3	80.0	56.7
Height (cm)	46.7	33.3	40.0	46.7	66.7	56.7
MDS-UPDRS II - Saliva and drooling	40.0	53.3	46.7	33.3	66.7	50.0
SCOPA-AUT Urinary score	60.0	26.7	43.3	33.3	73.3	53.3
SCOPA-AUT Thermoregulatory score	60.0	26.7	43.3	33.3	73.3	53.3

time-to-LID model achieved superior hold-out performance when using the mean-centering normalization approach as compared to no normalization. These results indicate that significant predictive performance can be achieved in a cross-cohort setting even without applying a cross-study normalization but adding dedicated batch adjustments has the potential to further improve the performance.

3.4. Assessment of clinical utility and calibration for the cross-cohort analysis

We assessed the clinical utility and calibration of the developed predictive models for the comprehensive cross-cohort analysis. A decision curve analysis (DCA) was used to evaluate the net benefit of the models in clinical decision-making, while a calibration analysis assessed the models' accuracy in predicting actual outcomes. Among all considered modeling approaches, CatBoost demonstrated the highest overall net benefit in the DCA for LID classification (see Suppl. Fig. 3), whereas Extra Survival Trees performed favorably in the time-to-LID analysis (see Suppl. Fig. 4). In contrast, AdaBoost and the NLSVM model showed the lowest net benefits for classification and time-to-LID analysis, respectively. The calibration analysis provided further insights into the models' ability to accurately predict LID outcomes. In line with the positive results in the DCA, CatBoost exhibited high calibration performance, with slopes close to 1, indicating a high level of agreement between predicted probabilities and actual outcomes (see Suppl. Table 10). Furthermore, this model also achieved lower mean squared error (MSE) values than other methods. For time-to-LID analysis, the NLSVM model is well-calibrated but has lower clinical utility than other models. Moreover, comprehensive models trained with CW-GBoost and Extra Survival Trees, as well as refined models trained with LSVM and Survival RF achieved high net benefits and reliable calibration. However, significant variation between the performance of different methods was observed, and GBoost, AdaBoost, Survival Trees, and Survival GBoost performed lowest in terms of calibration (see Suppl. Table 10).

3.5. Associations between clinical features and dyskinesia status

When comparing baseline clinical features between patients with LID+ and LID-within four years across three cohorts, as expected, LID + patients had a longer disease duration and younger age at PD onset, consistent with the negative correlation between age at PD onset and disease duration (cross-cohort Spearman correlation coefficient: −0.28, p = 5.5E-19). The cohorts cover a broad distribution for both the disease duration and age at onset, with significantly lower ages at onset (p = 7.3E-06) and significantly longer disease durations (p = 3.9E-87) in LuxPARK compared to the other two cohorts. LID + patients also displayed higher disease severity scores, including the total scores for MDS-UPDRS Part I, II, and III (ON), and the Modified Schwab & England ADL scale. LID + patients also scored significantly lower on Benton's Judgment of Line Orientation (JLO) test (p = 1.7E-10) and had higher motor fluctuation severity (p = 1.1E-24) and more pronounced rigidity in the lower extremities (p = 1.7E-04, see Table 2). SHAP value analysis of the ML models highlighted the variables contributing most significantly to the model's predictions (see Fig. 1 and Suppl. Figs. 5–7), including age of onset, tremor-related characteristics, axial symptoms, gastrointestinal dysfunction, *GBA* mutations, body weight, and various impairment assessments from the MDS-UPDRS and Autonomic Dysfunction Scales for Outcomes in PD (SCOPA-AUT). An overview of the top 10 most informative predictors for the cross-cohort analysis according to the average feature ranking scores is shown in Suppl. Table 11, including brief feature descriptions and references to relevant studies from the literature (see also the Discussion in section 4.2 for interpretations of the relevant predictive features).

4. Discussion

Predicting the onset of LID in PD patients involves several challenges due to both inter-individual heterogeneity and the many factors that influence disease progression in an individual. Recent advancements in modeling PD progression have already highlighted the importance of accounting for both intra-individual and inter-individual variability. For example, Severson et al. (2021) developed a statistical progression model of PD that used a contrastive latent variable model followed by a personalized input-output hidden Markov model [23]. This approach was designed to define disease states and assess their clinical significance across multiple key motor and cognitive outcomes. In contrast, our study differs not only by its specific focus on LID prediction and cross-study validation, but also by using different approaches to disease state modeling and feature selection. We did not rely on a particular modeling approach but compared different multivariable machine learning models for both classification and time-to-event analysis, using nested cross-validation to robustly determine feature importance and

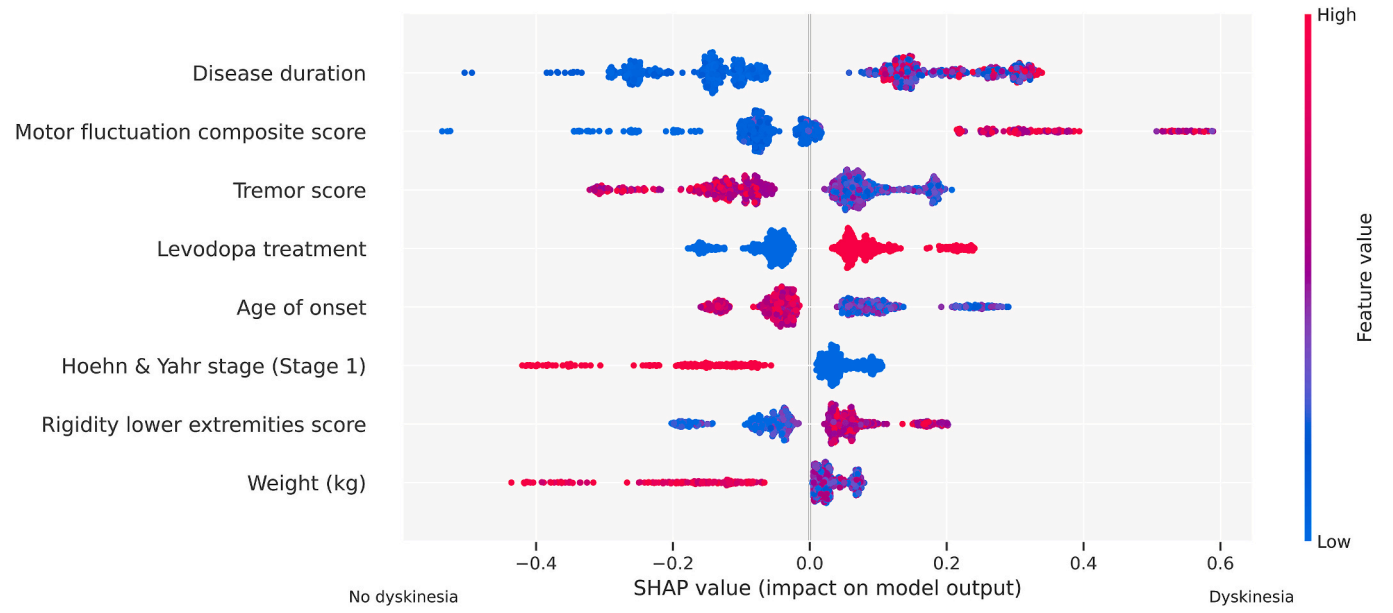

Fig. 1. SHAP value plot of the top predictors for the optimized comprehensive model for cross-cohort LID prognostic classification, displaying the magnitude and direction (positive or negative) of each feature's influence on the LID prognostic status as output.

optimize the models' generalizability. Both approaches have their merits, and the choice of methodology should be guided by the specific research objectives and the nature of the available data. For studies aiming specifically to capture the complex, overlapping trajectories of PD progression, Severson et al.'s method offers a dedicated and sophisticated framework. Conversely, for applications focused on predictive accuracy and generalizability across diverse cohorts, our approach provides a robust and broadly applicable methodology.

Our method also differs from previous studies dedicated specifically to LID prognosis, which have focused on investigating single cohorts, restricting the applicability of resulting prediction models to a specific patient population. To address this limitation, we compared multivariable ML models for both single-cohort and cross-cohort settings, covering both LID classification within a four-years timeframe and time-to-LID analysis. The cross-cohort model achieved competitive prediction results and improved model robustness, ensuring broad applicability. To guide the reader in finding the most appropriate prediction approaches for new datasets, our study statistically compared several methods for model building, feature selection, and cross-study normalization using a rigorous two-level CV approach.

4.1. Comparative evaluation of predictive models

When comparing the results across different cohorts, significant differences in predictive performance were observed. Superior cross-validated and hold-out test set performance for both LID classification and time-to-LID analysis was consistently observed for LuxPARK, likely due to its large sample size and more diverse patient population. By contrast, the sample size was smaller for ICEBERG, and PPMI and ICEBERG both covered shorter average disease durations.

In order to identify new predictive features associated with LID development, we also conducted feature selection analyses for the refined models, excluding baseline dyskinesia and levodopa medication use. While the hold-out AUCs of the refined models varied from those of the comprehensive models, they achieved significant predictive performance, suggesting that the refined modeling method may capture additional, more subtle relationships for data interpretation.

Finally, when comparing cross-cohort integrative analyses with single-cohort analyses, the cross-cohort models provided competitive predictive performance for LID prognosis and time-to-LID analysis, increasing the stability of the prediction results across the CV cycles. While the inclusion criteria and population characteristics must be carefully considered when training cross-cohort models, these integrative models provide more robust prediction results and are applicable across a broader range of patient populations.

4.2. Interpretation of models and predictors

The cross-cohort analyses revealed several clinical variables as informative predictors for LID prognosis, relevant not only for the practical purpose of forecasting a patient's future LID status but also for data interpretation in the context of the current knowledge on LID development in PD. Apart from the two most obvious predictors of future LID status, motor fluctuations and disease duration, several other clinical features contributed significantly to the model predictions. Previous studies have already identified levodopa equivalent daily dose (LEDD) as a significant predictor of levodopa-induced dyskinesia (LID) [2,24]. As a limitation, incomplete and inconsistent LEDD data for two of the cohorts prevented the inclusion of levodopa dosage in the predictive model. To address this, we performed a statistical analysis to examine the relationship between LEDD and LID specifically within the LuxPARK cohort. This analysis revealed a statistically significant difference between LID and LEDD with a p-value of 5.25E-06, confirming previous findings. In addition, a significant difference in time to LID was observed between PD patients with an LEDD of 400 mg or more and those with an LEDD of less than 400 mg (log-rank $p = 3.14E-03$).

Elevated MDS-UPDRS Part II and III scores were significant in forecasting LID, matching the observation that severity of motor symptoms correlates with LID risk. Furthermore, motor fluctuations, especially those emerging from prolonged levodopa therapy [4], were also highlighted as a relevant predictor by the SHAP value analysis. However, the correlations may reflect indirect relationships, and the complex associations between PD symptomatology and levodopa management challenges need to be considered.

Furthermore, axial impairments, including freezing of gait and heightened rigidity in the lower and upper extremities, were associated with LID development. However, this association may be explained by an indirect correlation, since both LID development and axial impairments are associated with disease duration (Spearman $p = 5.1E-65$ for axial impairments, and Mann-Whitney U test $p = 7.7E-38$ for LID).

Additionally, bradykinesia, i.e., slowness of movement, emerged as a significant LID predictor. In contrast, resting tremors displayed a negative association with LID risk, suggesting a reduced dyskinesia risk for tremor-dominant PD patients. This finding aligns with previous studies consistently reporting that PD patients with a tremor-dominant phenotype rather than an akinetic-rigid phenotype have a lower risk of LID [2]. Specifically, individuals with resting tremor as their initial motor symptom have a significantly reduced risk of developing LID [3] compared with other initial motor symptoms [24]. Overall, these results indicate potential nuanced interrelationships between specific motor symptoms and LID.

Assessment of non-motor symptoms of PD as possible LID predictors revealed gastrointestinal, urinary, and thermoregulatory symptoms associated with both PD severity and an increased LID risk. These symptoms may indirectly correlate with LID risk via PD severity [25]; however, these non-motor symptoms, in particular the gastrointestinal symptoms, have also been described to affect levodopa pharmacokinetics and treatment efficacy [26]. Furthermore, cognitive impairments, particularly visuospatial functions, and neuropsychiatric symptoms, such as depression and apathy, were significant LID predictors and also significantly associated with LID (Mann-Whitney *U* test p = 1.7E-10 for Benton JLO). While this may be explained by the significant correlation between visuospatial function (Benton JLO) and disease duration (Spearman p = 3.8E-10), further investigation is warranted regarding links between LID and cognition [27]. Previous research has pointed out that the deterioration seen in attention and executive function domains in PD, which is predictive of LID, might be due to disruptions in a shared cortical network that plays a role in both the emergence of LID and cognitive processes [28]. Additionally, research by Chung et al. demonstrated that specific baseline cognitive profiles, including visual memory/visuospatial functions, are associated with the progression of motor disability in PD, suggesting that these cognitive impairments can predict some aspects of motor prognosis, including LID development [29].

Genetic factors, including *GBA* mutations, influence PD symptoms and have previously been associated with an earlier LID development [30]. However, the association with future LID was not statistically significant in our meta-analysis (p = 0.11). Further molecular studies are needed to assess potential mechanistic interrelationships between mutations and PD complications.

Among the demographic variables, our analyses highlighted a significant negative correlation between older age at PD onset and LID risk (cross-cohort Spearman correlation: −0.22, p = 4.4E-12), consistent with previous studies [3]. While this correlation may be confounded by associations between age at PD onset and the follow-up duration in the cohort studies, the possibility that early-onset PD patients may require different management strategies to reduce LID risk may warrant further study. Understanding the distinct clinical manifestations of early- and late-onset PD patients may be necessary for effective prognosis and treatment of LID [31]. Our results also suggest body weight as a relevant covariate in LID prediction models, consistent with previous studies indicating that lower body weight in PD is a risk factor for dyskinesia [1]. Thus, apart from optimizing pharmaceutical therapies, non-pharmaceutical interventions to optimize nutrition and physical exercise may provide further means to mitigate LID risk.

Overall, our findings show that integrating clinical data from multiple cohorts can provide valuable information for LID prognosis across distinct patient populations. The analyses identified both established and novel predictors of LID, underlining the utility of interpretable ML approaches. Furthermore, the increased model robustness observed in the cross-study analyses highlights the effectiveness of combining data from multiple studies. Our evaluation of clinical usefulness and calibration for LID prognosis shows that different prediction models have different utility for clinical decision-making. Among these models, CatBoost and Extra Survival Trees demonstrate the most favorable results in terms of net benefit and calibration. These findings further support the applicability of some of the best-performing models in real-world clinical settings. Nevertheless, it is essential to further validate the models using independent data from more diverse cohorts to ensure their applicability in clinical settings across different geographic regions, populations, and clinical practices. Validation on larger datasets will also help to better address common limitations in the machine learning and cross-validation analysis on datasets with limited sample sizes, helping to reduce the variance in performance estimates. Finally, while our feature selection analyses have sought to increase model interpretability in how the models make predictions, for future clinical translation, further aspects beyond transparency need to be taken into account to ensure reliability and maintain trust. These include ethical and legal issues, particularly concerning the collection, processing, storage, and reuse of potentially sensitive patient data, while ensuring patient privacy and informed consent. The application of artificial intelligence algorithms on biomedical data in particular raises a variety of ethical questions, e.g., concerning potential biases, security and privacy issues, which have previously already been discussed in detail [32,33]. Addressing both the technical and validation challenges, and these ethical and legal considerations will be essential to not only improve clinical outcomes but also uphold the highest standards of data integrity and patient care.

For future research directions, researchers may want to explore the integration of multiple data types. For example, recent studies suggest that combining radiomic features with clinical data holds promise for improving LID prediction [34]. Such strategies to leverage the complementary information from different data types may lead to both more robust and more accurate prognostic models, and could help to pave the way for more individualized treatment strategies in PD.

5. Summary and conclusions

This study introduces three new aspects to researching LID in PD. Firstly, it develops cross-study prediction models for LID, offering tools for prognostic classification and accurate time-to-LID prediction from clinical data, aimed at early intervention and personalized PD management. Secondly, unlike traditional single-variable methods, these models use multivariable signatures incorporating complementary clinical descriptors for improved prediction robustness and accuracy. Thirdly, the study employs feature selection and SHAP value analysis for better model interpretability and more informed decision-making.

The cross-study ML models demonstrate significant predictive capabilities and robustness, outperforming single-cohort models in prediction stability.

The model evaluation analyses also highlighted the benefits of nested cross-validation and hyperparameter optimization, along with feature selection and cross-study normalization techniques for some of the considered cross-cohort analyses.

For the model interpretation, the statistical analyses and SHAP value analyses identified key clinical factors linked to LID risk, including well-known predictors such as levodopa use, PD progression markers, as well as new candidate predictors for further study.

Overall, the findings highlight the potential of ML for cross-study LID prognosis, facilitating precision medicine in PD by integrating information from distinct cohorts to enable personalized predictions. This approach has the potential to enhance the understanding of LID risk factors and support future clinical decision-making, enabling the design of new preemptive measures through tailored drug dosing protocols along with non-pharmacological interventions such as physical activity.

Code availability

The data processing, normalization and statistical analyses were performed using R (v4.2.1). Python-3.8.6-GCCcore-10.2.0 was used for efficient machine learning predictions. The open source code is accessible in the GitLab repository under the MIT license: https://gitlab.lcsb.

uni.lu/bds/ml_dyskinesia.

Data availability

The LuxPARK clinical dataset used in this study was obtained from the National Centre of Excellence in Research on Parkinson's Disease (NCER-PD). The dataset for this manuscript is not publicly available as it is linked to the Luxembourg Parkinson's Study and its internal regulations. Any requests for accessing the dataset can be directed to request.ncer-pd@uni.lu.

Data used in the preparation of this article were obtained on January 11, 2023, from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data, RRID:SCR 006431). For up-to-date information on the study, please visit the PPMI website (www.ppmi-info.org).

Data from the ICEBERG cohort analyzed during this study is available from the corresponding study group (jean-christophe.corvol@aphp.fr, marie.vidailhet@aphp.fr).

Funding

EG acknowledges support by the Luxembourg National Research Fund (FNR) for the project RECAST (INTER/22/17104370/RECAST) as part of the Joint Programme - Neurodegenerative Disease Research (JPND) and for the project PreDYT (INTER/EJP RD 22/PREDYT) as part of the EJP RD Joint Transnational Call 2022 (JTC 2022). OT and JK acknowledge support by the FNR for the project "Digital Healthcare Solutions: Patient Management e-Health Concepts" (dHealthPD 14146272). The National Centre of Excellence in Research on Parkinson's Disease (NCER-PD) received funding from the Luxembourg National Research Fund (FNR/NCER13/BM/11264123). PPMI - a public-private partnership - is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners. A list of names of all of the PPMI funding partners can be found at sponsors/www.ppmi-info.org/about-ppmi/who-we-are/study-sponsors/. The ICEBERG cohort received funding and support from the Agence Nationale de la Recherche (ANR) under grant agreements ANR-10-IAIHU-06 (IHU ICM), association France Parkinson, and the Fondation d'Entreprise EDF, and the Fondation Saint Michel, and Energipole.

CRediT authorship contribution statement

Jochen Klucken: Writing – review & editing. **Graziella Mangone:** Writing – review & editing. **Fouad Khoury:** Writing – review & editing. **Marie Vidailhet:** Writing – review & editing. **Jean-Christophe Corvol:** Writing – review & editing. **Rejko Krüger:** Writing – review & editing. **Enrico Glaab:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition. **Rebecca Ting Jiin Loo:** Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Olena Tsurkalenko:** Writing – review & editing, Validation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The machine learning predictions in this paper were partly performed using the HPC facilities of the University of Luxembourg (see http://hpc.uni.lu).

We would like to thank all participants of the Luxembourg Parkinson's Study for their essential support of our research. Furthermore, we acknowledge the joint effort of the National Centre of Excellence in Research on Parkinson's Disease (NCER-PD) Consortium members from the partner institutions Luxembourg Centre for Systems Biomedicine, Luxembourg Institute of Health, Centre Hospitalier de Luxembourg, and Laboratoire National de Santé generally contributed to the Luxembourg Parkinson's Study as listed below.

Geeta ACHARYA [2], Gloria AGUAYO [2], Myriam ALEXANDRE [2], Muhammad ALI [1], Wim AMMERLANN [2], Giuseppe ARENA [1], Michele BASSIS [1], Roxane BATUTU [3], Katy BEAUMONT [2], Sibylle BÉCHET [3], Guy BERCHEM [3], Alexandre BISDORFF [5], Ibrahim BOUSSAAD [1], David BOUVIER [4], Lorieza CASTILLO [2], Gessica CONTESOTTO [2], Nancy DE BREMAEKER [3], Brian DEWITT [2], Nico DIEDERICH [3], Rene DONDELINGER [5], Nancy E. RAMIA [1], Angelo Ferrari [2], Katrin FRAUENKNECHT [4], Joëlle FRITZ [2], Carlos GAMIO [2], Manon GANTENBEIN [2], Piotr GAWRON [1], Laura Georges [2], Soumyabrata GHOSH [1], Marijus GIRAITIS [2,3], Enrico GLAAB [1], Martine GOERGEN [3], Elisa GÓMEZ DE LOPE [1], Jérôme GRAAS [2], Mariella GRAZIANO [7], Valentin GROUES [1], Anne GRÜNEWALD [1], Gaël HAMMOT [2], Anne-Marie HANFF [2,10,11], Linda HANSEN [3], Michael HENEKA [1], Estelle HENRY [2], Margaux Henry [2], Sylvia HERBRINK [3], Sascha HERZINGER [1], Alexander HUNDT [2], Nadine JACOBY [8], Sonja JÓNSDÓTTIR [2,3], Jochen KLUCKEN [1,2,3], Olga KOFANOVA [2], Rejko KRÜGER [1,2,3], Pauline LAMBERT [2], Zied LANDOULSI [1], Roseline LENTZ [6], Laura LONGHINO [3], Ana Festas Lopes [2], Victoria LORENTZ [2], Tainá M. MARQUES [2], Guilherme MARQUES [2], Patricia MARTINS CONDE [1], Patrick MAY [1], Deborah MCINTYRE [2], Chouaib MEDIOUNI [2], Francoise MEISCH [1], Alexia MENDIBIDE [2], Myriam MENSTER [2], Maura MINELLI [2], Michel MITTELBRONN [1,2,4,10,12,13], Saïda MTIMET [2], Maeva Munsch [2], Romain NATI [3], Ulf NEHRBASS [2], Sarah NICKELS [1], Beatrice NICOLAI [3], Jean-Paul NICOLAY [9], Fozia NOOR [2], Clarissa P. C. GOMES [1], Sinthuja PACHCHEK [1], Claire PAULY [2,3], Laure PAULY [2,10], Lukas PAVELKA [2,3], Magali PERQUIN [2], Achilleas PEXARAS [2], Armin RAUSCHENBERGER [1], Rajesh RAWAL [1], Dheeraj REDDY BOBBILI [1], Lucie REMARK [2], Ilsé Richard [2], Olivia ROLAND [2], Kirsten ROOMP [1], Eduardo ROSALES [2], Stefano SAPIENZA [1], Venkata SATAGOPAM [1], Sabine SCHMITZ [1], Reinhard SCHNEIDER [1], Jens SCHWAMBORN [1], Raquel SEVERINO [2], Amir SHARIFY [2], Ruxandra SOARE [1], Ekaterina SOBOLEVA [1,3], Kate SOKOLOWSKA [2], Maud Theresine [2], Hermann THIEN [2], Elodie THIRY [3], Rebecca TING JIIN LOO [1], Johanna TROUET [2], Olena TSURKALENKO [2], Michel VAILLANT [2], Carlos VEGA [2], Liliana VILAS BOAS [3], Paul WILMES [1], Evi WOLLSCHEID-LENGELING [1], Gelani ZELIMKHANOV [2,3]

[1] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg.

[2] Luxembourg Institute of Health, Strassen, Luxembourg.

[3] Centre Hospitalier de Luxembourg, Strassen, Luxembourg.

[4] Laboratoire National de Santé, Dudelange, Luxembourg.

[5] Centre Hospitalier Emile Mayrisch, Esch-sur-Alzette, Luxembourg.

[6] Parkinson Luxembourg Association, Leudelange, Luxembourg.

[7] Association of Physiotherapists in Parkinson's Disease Europe, Esch-sur-Alzette, Luxembourg.

[8] Private practice, Ettelbruck, Luxembourg.

[9] Private practice, Luxembourg, Luxembourg.

[10] Faculty of Science, Technology and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg.

[11] Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Maastricht University Medical Centre, Maastricht, the Netherlands.

[12] Luxembourg Center of Neuropathology, Dudelange, Luxembourg.

[13] Department of Life Sciences and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg.

Furthermore, we extend our gratitude to the ICEBERG study group for their contribution. A list of ICEBERG members can be found below.

Steering committee: Marie Vidailhet, MD, PhD, (Pitié-Salpêtrière Hospital, Paris, principal investigator of ICEBERG), Jean-Christophe Corvol, MD, PhD (Pitié-Salpêtrière Hospital, Paris, scientific lead), Isabelle Arnulf, MD, PhD (Pitié-Salpêtrière Hospital, Paris, member of the steering committee), Stéphane Lehericy, MD, PhD (Pitié-Salpêtrière

Hospital, Paris, member of the steering committee).

Clinical data: Marie Vidailhet, MD, PhD, (Pitié-Salpêtrière Hospital, Paris, coordination), Graziella Mangone, MD, PhD (Pitié-Salpêtrière Hospital, Paris, co-coordination), Jean-Christophe Corvol, MD, PhD (Pitié-Salpêtrière Hospital, Paris), Isabelle Arnulf, MD, PhD (Pitié-Salpêtrière Hospital, Paris), Sara Sambin, MD (Pitié-Salpêtrière Hospital, Paris), Poornima Menon, MD (Pitié-Salpêtrière Hospital, Paris, Jonas Ihle, MD (Pitié-Salpêtrière Hospital, Paris), Caroline Weill, MD, (Pitié-Salpêtrière Hospital, Paris), David Grabli, MD, PhD (Pitié-Salpêtrière Hospital, Paris); Florence Cormier-Dequaire, MD (Pitié-Salpêtrière Hospital, Paris); Louise Laure Mariani, MD, PhD (Pitié-Salpêtrière Hospital, Paris), Bertrand Degos, MD, PhD (Avicenne Hospital, Bobigny).

Neuropsychological data: Richard Levy, MD (Pitié-Salpêtrière Hospital, Paris, coordination), Fanny Pineau, MS (Pitié-Salpêtrière Hospital, Paris, neuropsychologist), Julie Socha, MS (Pitié-Salpêtrière Hospital, Paris, neuropsychologist), Eve Benchetrit, MS (La Timone Hospital, Marseille, neuropsychologist), Virginie Czernecki, MS (Pitié-Salpêtrière Hospital, Paris, neuropsychologist), Marie-Alexandrine, MS (Pitié-Salpêtrière Hospital, Paris, neuropsychologist).

Eye movement: Sophie Rivaud-Pechoux, PhD (ICM, Paris, coordination); Elodie Hainque, MD, PhD (Pitié-Salpêtrière Hospital, Paris).

Sleep assessment: Isabelle Arnulf, MD, PhD (Pitié-Salpêtrière Hospital, Paris, coordination), Smaranda Leu Semenescu, MD (Pitié-Salpêtrière Hospital, Paris), Pauline Dodet, MD (Pitié-Salpêtrière Hospital, Paris).

Genetic data: Jean-Christophe Corvol, MD, PhD (Pitié-Salpêtrière Hospital, Paris, coordination), Graziella Mangone, MD, PhD (Pitié-Salpêtrière Hospital, Paris, co-coordination), Samir Bekadar, MS (Pitié-Salpêtrière Hospital, Paris, biostatistician), Alexis Brice, MD (ICM, Pitié-Salpêtrière Hospital, Paris), Suzanne Lesage, PhD (INSERM, ICM, Paris, genetic analyses).

Metabolomics: Fanny Mochel, MD, PhD (Pitié-Salpêtrière Hospital, Paris, coordination), Farid Ichou, PhD (ICAN, Pitié-Salpêtrière Hospital, Paris), Vincent Perlbarg, PhD, Pierre and Marie Curie University), Benoit Colsch, PhD (CEA, Saclay), Arthur Tenenhaus, PhD (Supelec, Gif-sur-Yvette, data integration).

Brain MRI data: Stéphane Lehericy, MD, PhD (Pitié-Salpêtrière Hospital, Paris, coordination), Rahul Gaurav, MS, (Pitié-Salpêtrière Hospital, Paris, data analysis), Nadya Pyatigorskaya, MD, PhD, (Pitié-Salpêtrière Hospital, Paris, data analysis); Lydia Yahia-Cherif, PhD (ICM, Paris, Biostatistics), Romain Valabregue, PhD (ICM, Paris, data analysis), Cécile Galléa, PhD (ICM, Paris).

Datscan imaging data: Marie-Odile Habert, MCU-PH (Pitié-Salpêtrière Hospital, Paris, coordination).

Voice recording: Dijana Petrovska, PhD (Telecom Sud Paris, Evry, coordination), Laetitia Jeancolas, MS (Telecom Sud Paris, Evry).

Study management: Alizé Chalançon (Pitié-Salpêtrière Hospital, Paris, Project manager), Carole Dongmo-Kenfack (Pitié-Salpêtrière Hospital, Paris, clinical research assistant); Christelle Laganot (Pitié-Salpêtrière Hospital, Paris, clinical research assistant), Valentine Maheo (Pitié-Salpêtrière Hospital, Paris, clinical research assistant), Manon Gomes (Pitié-Salpêtrière Hospital, Paris, clinical research assistant), Mickaël Lé (Pitié-Salpêtrière Hospital, Paris, clinical research assistant)

Study sponsoring**:** INSERM, Paris.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.parkreldis.2024.107054.

References

[1] D.K. Kwon, M. Kwatra, J. Wang, H.S. Ko, Levodopa-induced dyskinesia in Parkinson's disease: pathogenesis and emerging treatment strategies, Cells 11 (23) (2022) 3736, https://doi.org/10.3390/cells11233736.

[2] M. Hutny, J. Hofman, A. Klimkowicz-Mrowiec, A. Gorzkowska, Current knowledge on the background, pathophysiology and treatment of levodopa-induced dyskinesia - literature review, J. Clin. Med. 10 (19) (2021) 4377, https://doi.org/10.3390/jcm10194377.

[3] A. Tirozzi, N. Modugno, N.P. Palomba, R. Ferese, A. Lonbardi, E. Olivola, et al., Analysis of genetic and non-genetic predictors of levodopa induced dyskinesia in Parkinson's disease, Front. Pharmacol. 12 (2021), https://doi.org/10.3389/fphar.2021.640603.

[4] H. You, L.L. Mariani, G. Mangone, D. Le Febvre de Nailly, F. Charbonnier-Beaupel, J.C. Corvol, Molecular basis of dopamine replacement therapy and its side effects in Parkinson's disease, Cell Tissue Res. 373 (2018) 111–135, https://doi.org/10.1007/s00441-018-2813-2.

[5] R. Cilia, A. Akpalu, F.S. Sarfo, M. Cham, M. Amboni, E. Cereda, et al., The modern pre-levodopa era of Parkinson's disease: insights into motor complications from sub-Saharan Africa, Brain 137 (10) (2014) 2731–2742, https://doi.org/10.1093/brain/awu195.

[6] E. Bezard, J.M. Brotchie, C.E. Gross, Pathophysiology of levodopa-induced dyskinesia: potential for new therapies, Nat. Rev. Neurosci. 2 (2001) 577–588, https://doi.org/10.1038/35086062.

[7] J.T.B. Keun, I.A. Arnoldussen, C. Vriend, O. van de Rest, Dietary approaches to improve efficacy and control side effects of levodopa therapy in Parkinson's disease: a systematic review, Adv. Nutr. 12 (6) (2021) 2265–2287, https://doi.org/10.1093/advances/nmab060.

[8] R. Hornung, D. Causeur, C. Bernau, A.L. Boulesteix, Improving cross-study prediction through addon batch effect adjustment or addon normalization, Bioinformatics 33 (3) (2017) 397–404, https://doi.org/10.1093/bioinformatics/btw650.

[9] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55 (1) (1997) 119–139, https://doi.org/10.1006/jcss.1997.1504.

[10] J.R. Quinlan, Improved use of continuous attributes in C4.5, J. Artif. Intell. 4 (1996) 77–90, https://doi.org/10.1613/jair.279.

[11] Y.S. Tan, C. Singh, K. Nasseri, A. Agarwal, J. Duncan, O. Ronen, et al., Fast interpretable greedy-tree sums, Preprint at arXiv preprint, arXiv:2201.11931 (2022), https://doi.org/10.48550/arXiv.2201.11931.

[12] H. McTavish, C. Zhong, R. Achermann, I. Karimalis, J. Chen, C. Rudin, et al., Fast sparse decision tree optimization via reference ensembles, Preprint at arXiv preprint, arXiv:2112.00798, https://doi.org/10.48550/arXiv.2112.00798, 2022.

[13] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. 29 (5) (2001) 1189–1232. http://www.jstor.org/stable/2699986.

[14] A. Agarwal, Y.S. Tan, O. Ronen, C. Singh, B. Yu, Hierarchical Shrinkage: improving the accuracy and interpretability of tree-based methods, Preprint at arXiv preprint, arXiv:2202.00858 (2022), https://doi.org/10.48550/arXiv.2202.00858.

[15] D. Bertsimas, J. Dunn, E. Gibson, A. Orfanoudaki, Optimal survival trees, Mach. Learn. 111 (2022) 2951–3023, https://doi.org/10.1007/s10994-021-06117-0.

[16] K. He, Y. Li, J. Zhu, H. Liu, J.E. Lee, C.I. Amos, et al., Component-wise gradient boosting and false discovery control in survival analysis with high-dimensional covariates, Bioinformatics 32 (1) (2016) 50–57, https://doi.org/10.1093/bioinformatics/btv517.

[17] H. Ishwaran, U.B. Kogalur, E.H. Blackstone, M.S. Lauer, Random survival forests, Ann. Appl. Stat. 2 (3) (2008) 841–860, https://doi.org/10.1214/08-AOAS169.

[18] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach. Learn. 63 (2006) 3–42, https://doi.org/10.1007/s10994-006-6226-1.

[19] S.M. Lundberg, S.I. Lee, A unified approach to interpreting model predictions, in: Proceedings of the Advances in Neural Information Processing Systems, 2017, pp. 4765–4774.

[20] X. Sun, W. Xu, Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves, IEEE Signal Process. Lett. 21 (11) (2014) 1389–1393, https://doi.org/10.1109/LSP.2014.2337313.

[21] L. Kang, W. Chen, N.A. Petrick, B.D. Gallas, Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach, Stat. Med. 34 (4) (2015) 685–703, https://doi.org/10.1002/sim.6370.

[22] D. Piovani, R. Sokou, A.G. Tsantes, A.S. Vitello, S. Bonovas, Optimizing clinical decision making with decision curve analysis: insights for clinical investigators, Healthcare 11 (16) (2023) 2244, https://doi.org/10.3390/healthcare11162244.

[23] K.A. Severson, L.M. Chahine, L.A. Smolensky, M. Dhuliawala, M. Frasier, K. Ng, et al., Discovery of Parkinson's disease states and disease progression modelling: a longitudinal data study using machine learning, The Lancet Digital Health 3 (9) (2021) e555–e564, https://doi.org/10.1016/S2589-7500(21)00101-1.

[24] B.L. Santos-Lobato, A.F. Schumacher-Schuh, C.R.M. Rieder, M.H. Hutz, V. Borges, H.B. Ferraz, et al., Diagnostic prediction model for levodopa-induced dyskinesia in Parkinson's disease, Arq. Neuro. Psiquiatr. 78 (4) (2020) 206–216, https://doi.org/10.1590/0004-282x20190191.

[25] Z. Chen, G. Li, J. Liu, Autonomic dysfunction in Parkinson's disease: implications for pathophysiology, diagnosis, and treatment, Neurobiol. Dis. 134 (2020) 104700, https://doi.org/10.1016/j.nbd.2019.104700.

[26] A.H. Tan, K.H. Chuah, Y.Y. Beh, J.P. Schee, S. Mahadeva, S.Y. Lim, Gastrointestinal dysfunction in Parkinson's disease: neuro-gastroenterology perspectives on a multifaceted problem, Journal of Movement Disorders 16 (2) (2023) 138–151, https://doi.org/10.14802/jmd.22220.

[27] S. Navailles, P. De Deurwaerdère, Contribution of serotonergic transmission to the motor and cognitive effects of high-frequency stimulation of the subthalamic nucleus or levodopa in Parkinson's disease, Mol. Neurobiol. 45 (2012) 173–185, https://doi.org/10.1007/s12035-011-8230-0.

[28] A. Luca, R. Monastero, R. Baschi, C.E. Cicero, G. Mostile, M. Davi, et al., Cognitive impairment and levodopa induced dyskinesia in Parkinson's disease: a longitudinal study from the PACOS cohort, Sci. Rep. 11 (2021) 867, https://doi.org/10.1038/s41598-020-79110-7.

[29] S.J. Chung, H.S. Yoo, H.S. Lee, Y.H. Lee, K. Baik, J.H. Jung, et al., Baseline cognitive profile is closely associated with long-term motor prognosis in newly diagnosed Parkinson's disease, J. Neurol. 268 (2021) 4203–4212, https://doi.org/10.1007/s00415-021-10529-2.

[30] S. Thanprasertsuk, P. Phowthongkum, T. Hopetrungraung, C. Poorirerngpoom, T. Sathirapatya, P. Wichit, et al., Levodopa-induced dyskinesia in early-onset Parkinson's disease (EOPD) associates with glucocerebrosidase mutation: a next-generation sequencing study in EOPD patients in Thailand, PLoS One 18 (10) (2023) 0293516, https://doi.org/10.1371/journal.pone.0293516.

[31] Z. Zhang, G. Liu, D. Wang, H. Chen, D. Su, W. Kou, et al., Effect of onset age on the levodopa threshold dosage for dyskinesia in Parkinson's disease, Neurol. Sci. 43 (2022) 3165–3174, https://doi.org/10.1007/s10072-021-05694-1.

[32] H. Fröhlich, N. Bontridder, D. Petrovska-Delacréta, E. Glaab, F. Kluge, M. E. Yacoubi, et al., Leveraging the potential of digital technology for better individualized treatment of Parkinson's disease, Front. Neurol. 13 (2022) 788427, https://doi.org/10.3389/fneur.2022.788427.

[33] S. Gerke, T. Minssen, G. Cohen, Ethical and legal challenges of artificial intelligence-driven healthcare, Artificial Intelligence in Healthcare (2020) 295–336, https://doi.org/10.1016/B978-0-12-818438-7.00012-5.

[34] Y. Luo, H. Chen, M. Gui, Radiomics and hybrid models based on machine learning to predict levodopa-induced dyskinesia of Parkinson's disease in the first 6 years of levodopa treatment, Diagnostics 13 (15) (2023) 2511, https://doi.org/10.3390/diagnostics13152511.