From Clinical Criteria to AI-Based Classification of Advanced Parkinson's Disease: A Data-Driven Approach Using Structured PPMI Data

Iñigo Gabilondo
igabilon@gmail.com

Biobizkaia Health Research Institute

Angela Sáenz
Biobizkaia Health Research Institute

Sandra Seijo
Ibermática (Spain)

Alvaro Ochoa
Ibermática (Spain)

Unai Zalabarria
Ibermática (Spain)

Itziar Cuenca
Ibermática (Spain)

Beatriz Tijero
Hospital de Cruces

Tamara Fernández-Valle
Hospital de Cruces

Marta Ruiz
Hospital de Cruces

Marian Acera
Biobizkaia Health Research Institute

Inês Sousa
Fraunhofer Portugal Research

Juan Carlos Gómez□Esteban
Biobizkaia Health Research Institute

Rocío Del Pino
Biobizkaia Health Research Institute

Article

Keywords: Advanced Parkinson's Disease, Artificial Intelligence, Machine Learning, PPMI, Predictive Modelling, CDEPA criteria

Posted Date: October 13th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7745559/v1

Additional Declarations: No competing interests reported.

From Clinical Criteria to AI-Based Classification of Advanced Parkinson's Disease: A Data-Driven Approach Using Structured PPMI Data

Authors: Iñigo Gabilondo MD PhD*[1,2,3], Angela Sáenz[1], Sandra Seijo PhD[4], Alvaro Ochoa[4], Unai Zalabarria PhD[4], Itziar Cuenca PhD[4], Beatriz Tijero MD PhD[1,3], Tamara Fernández-Valle MD[1,3,5], Marta Ruiz MD PhD[1,3], Marian Acera[1], Inês Sousa PhD [5], Juan Carlos Gómez-Esteban MD PhD[1,3,5], Rocío Del Pino PhD[1,2]*

[1]Neurodegenerative Diseases Group, Biobizkaia Health Research Institute, Barakaldo, Bizkaia, Spain; [2]Ikerbasque: The Basque Foundation for Science, Bilbao, Spain; [3] Department of Neurology, Cruces University Hospital, Bizkaia, Spain; [4]Ibermática Innovation Institute-i3B, Derio, Bizkaia, Spain; [5]Department of Neurosciences, Faculty of Medicine and Nursing, University of the Basque Country (UPV/EHU), Leioa, Spain; [6]Fraunhofer Portugal AICOS, Porto, Portugal.

*** Corresponding authors:**

Iñigo Gabilondo, Neurodegenerative Diseases Group, Biobizkaia Health Research Institute. Email: igabilon@gmail.com

Rocio del Pino, Neurodegenerative Diseases Group, Biobizkaia Health Research Institute. Email: delpinorocio@gmail.com

Word count: Abstract: 148, Main text: 4509, No. of references: 32, Figures: 4, Tables: 5**,** Supplementary Material: Suppl Table 1 and Suppl Table 2

Running head:

AI-Based Classification of Advanced Parkinson's Disease

Abbreviations (list): PD – Parkinson's Disease**;** APD – Advanced Parkinson's Disease**;** CDEPA – Clinical Diagnostic Criteria for Advanced Parkinson's Disease**;** PPMI – Parkinson's Progression Markers Initiative**;** ML – Machine Learning**;** AI – Artificial Intelligence; SVM – Support Vector Machine; AUC-ROC – Area Under the Receiver Operating Characteristic Curve; XGBoost – Extreme Gradient Boosting; HRI – Health Research Institute; DBS – Deep Brain Stimulation; HIFU – High-Intensity Focused Ultrasound; MLP–Multilayer Perceptron; RIPPER – Repeated Incremental Pruning to Produce Error Reduction; PART – Projective Adaptive Resonance Theory; LMT – Logistic Model Tree.

Keywords: Advanced Parkinson's Disease, Artificial Intelligence, Machine Learning, PPMI, Predictive Modelling, CDEPA criteria

ABSTRACT

Advanced Parkinson's disease (APD) involves severe motor and non-motor complications and requires early identification, yet lacks a standardized quantitative definition. This study translated expert-defined CDEPA criteria into 216 structured variables from the Parkinson's Progression Markers Initiative (PPMI) dataset to train machine learning models for early APD classification. A 1,302 patients cohort was followed up for 13 years. A label-rescuing strategy addressed longitudinal incompleteness. Supervised models trained on baseline data predicted future APD status. Binary classifiers outperformed multiclass approaches; the best-performing model (XGBoost, Year 9) achieved AUC 0.881, balanced accuracy 0.824, and F1 score 0.819. Top predictors included genetic mutation status, age, MDS-UPDRS I–II, REM sleep behavior disorder, and tremor severity. Non-motor symptoms—especially autonomic dysfunction and sleep disturbances—were more informative than motor signs, comprising 57.1% of top features. These findings support the feasibility of early APD classification and propose a scalable, data-driven framework for APD prediction.

INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by a wide range of motor and non-motor symptoms that significantly impair quality of life[1]. As the disease progresses, patients often experience disabling complications such as motor fluctuations, dyskinesias, freezing of gait, falls, cognitive decline, hallucinations, and dysautonomia[1]. These manifestations increase morbidity and mortality and often prompt the need for second-line treatments or device-aided therapies, such as deep brain stimulation (DBS), high-intensity focused ultrasound (HIFU), or intrajejunal levodopa infusion.

A major challenge in clinical practice is the timely identification of patients who have reached the advanced stage of PD (Advanced Parkinson's Disease, APD), when such interventions are most beneficial[2]. However, no universally accepted, quantitative definition of APD currently exists[3]. This lack of consensus contributes to heterogeneity in in clinical practice, delays in therapeutic escalation, and difficulties in patient stratification for clinical research.

Among the available tools, the CDEPA questionnaire (Cuestionario De Enfermedad de Parkinson Avanzada) [4] is one of the most widely recognized. It is based on 13 clinical items grouped into six domains—including motor and non-motor symptoms, disease duration, and functional disability—and classifies patients into four diagnostic categories: definitive, probable, possible, or non-APD. Although validated in clinical settings[5], the CDEPA remains a qualitative instrument. Other tools, such as the 5-2-1 criteria [6] or the MANAGE-PD

algorithm[7], also support clinical decision-making but similarly rely on subjective assessments and lack a standardized, data-driven framework for defining APD.

This reflects a broader unmet need in PD management: despite the availability of several screening tools, there is no objective methodology to identify patients at risk of advanced disease and in need of therapeutic escalation. Recent reviews have emphasized key limitations, including the absence of standardized diagnostic paradigms, limited multimodal data integration, and underuse of longitudinal data for predicting disease trajectories[8]

Integrating expert-driven clinical criteria with structured, real-world clinical data could improve diagnostic precision and enable earlier identification of APD. In this context, the Parkinson's Progression Markers Initiative (PPMI) offers a uniquely valuable resource: a large, longitudinal, and publicly available dataset including comprehensive clinical, neuropsychiatric, and biomarker data across all stages of PD [9]. In parallel, machine learning (ML) provides powerful tools to model high-dimensional data and enhance early detection of disease progression.[10].

In this study, we developed interpretable ML models that integrate structured baseline data from the PPMI cohort with expert-defined CDEPA criteria to predict long-term progression to APD. This approach combines clinical expertise with data-driven modelling to support early patient stratification and timely therapeutic decision-making.

SUBJECTS AND METHODS

Study Overview and Data Sources

We conducted a retrospective, data-driven analysis integrating the Clinical Diagnostic Criteria for Advanced Parkinson's Disease (CDEPA) with quantitative

and semi-quantitative structured data from the Parkinson's Progression Markers Initiative (PPMI). Only participants diagnosed with PD were included, and variables covered demographic, clinical, and genetic information to model progression and identify predictors of advanced stages. This work was part of AISym4Med project, a Horizon Europe-funded project involving 15 institutions from eight countries, aimed at promoting the application of AI and machine learning in medical research through interoperable, privacy-compliant datasets[11], with our specific focus within the project on developing AI tools for the early detection of APD. PPMI is a multicenter, observational cohort initiated by The Michael J. Fox Foundation in 2010 to characterize the natural history of PD[9]. It includes 4,823 individuals, spanning prodromal to moderate PD stages, as well as healthy controls. Data collection follows standardized protocols and includes motor and non-motor clinical assessments, neuropsychiatric and autonomic evaluations, imaging (DaTSCAN, MRI), and biospecimens such as blood, cerebrospinal fluid, DNA, and RNA. For this analysis, we selected a subset of 1,302 individuals from the "PPMI Clinical" database, enrolled in the PD cohort, and with recorded status as "Enrolled," "Withdrew," or "Complete." Healthy controls and subjects with non-PD diagnoses were excluded (**Error! Reference source not found.**).

Definition of Advanced Parkinson's Disease

CDEPA is a clinician-administered screening instrument validated for detecting advanced stages of PD. It comprises 13 items grouped into six domains: (1) general characteristics (e.g., disease duration); (2) disability in activities of daily living; (3) treatment-related motor complications (e.g., motor fluctuations); (4)

disease-related motor symptoms (e.g., dysphagia, falls, freezing of gait, dysarthria, posture/equilibrium); (5) disease-related non-motor symptoms (e.g., dysautonomia/orthostatic hypotension, somnolence); and (6) neuropsychiatric/cognitive symptoms (e.g., dementia or significant cognitive impairment, hallucinations/delirium, apathy). Based on item presence and severity across domains, patients are categorized as Definitive, Probable, Possible, or Not APD. Definitive APD typically requires at least one unequivocal indicator of advanced disease (e.g., severe functional dependence, marked dysphagia, recurrent falls, or dementia). Probable APD requires qualifying features across ≥2 domains; Possible APD includes clinically relevant combinations that do not meet higher thresholds. CDEPA has shown high sensitivity, moderate specificity, and good inter-rater/test–retest reliability, supporting its use in clinical and research settings.

PPMI Data Labelling Strategy

To apply CDEPA within PPMI, each of the 13 items was mapped to objective, cohort-available variables via consensus among five neurologists and two neuropsychologists with PD expertise. This enabled assignment of APD certainty levels using structured measures (Table 1). Examples of mappings included: disease duration ≥10 years as supportive of Probable APD; functional disability defined by a Modified Schwab & England score ≤60% or the presence of marked complications (e.g., severe OFF time, dysphagia, or recurrent falls) indicating Definitive APD; Montreal Cognitive Assessment (MoCA) ≤15 as evidence of significant cognitive impairment; and SCOPA-AUT thresholds (e.g., ≥6) to capture autonomic dysfunction.

Longitudinal completeness varied markedly across visits and variables. Some key variables exhibited >85% missingness at certain timepoints, and cohort attrition produced progressive data loss. To improve temporal coherence and mitigate label instability caused by missing assessments, we implemented a label-rescuing strategy: (1) once assigned, Definitive APD labels were propagated forward; (2) the most recent label was carried through missing intervals; and (3) individuals never meeting APD criteria were consistently labeled as Not APD. This strategy increased the number of usable labeled visits while preserving conservative estimates of severity. Because label distributions were highly unbalanced at early and late timepoints, we identified Year 7 as a key analytic window: it offered the most balanced representation of Not, Possible, Probable, and Definitive APD categories and therefore served as a central reference for training and evaluation.

Patient retention across the 13-year follow-up period (comprising 14 scheduled visits, from Baseline to Year 13) was monitored using PPMI metadata. Although participant dropout was expected over time, a substantial proportion of withdrawals lacked a registered cause in the dataset, potentially reflecting missing-not-at-random patterns. Supplementary Table 2 summarizes the number, percentage, and documented reasons for participant withdrawal at each visit. This information was considered essential for assessing data completeness and informing label-rescuing strategies, especially in the context of temporally uneven missingness and potential attrition bias.

Data Curation and Feature Selection

The raw extraction encompassed 7,091 variables across 13 PPMI folders spanning up to 13 years of follow-up. To avoid circularity with the labeling

approach and to maximize external applicability, we restricted predictors to baseline and clinically derived variables available at initial assessment. We then executed a two-stage variable reduction driven by clinical relevance and data quality. First, the expert panel selected 216 variables most representative of the CDEPA domains, drawn from four folders: Subject Characteristics (n=13), Medical History (n=78), Motor Assessments (n=113), and Non-Motor Assessments (n=12) (**Error! Reference source not found.**). Second, variables with excessive missingness (often >85%) or low information content were excluded from modeling.

Missing numerical values were imputed using the feature mean, and categorical variables with the mode. Categorical predictors were one-hot encoded; numerical variables were standardized to comparable scales. Identifiers (e.g., PATNO), enrollment status, constant features, and date fields were removed. After curation, 83 baseline variables remained; one-hot encoding expanded the feature space to 186 model-ready predictors. This pipeline prioritized reproducibility and interpretability while retaining clinically meaningful breadth across motor, non-motor, functional, and demographic domains.

Machine Learning-Based Classification Models

We formulated two complementary prediction tasks. The primary task was a binary classification distinguishing Definitive APD from Not Definitive (pooling Not, Possible, and Probable APD), which yields a clinically actionable dichotomy for early identification of individuals at high risk of advanced disease. A secondary multiclass task preserved label granularity by predicting Not, Possible, Probable, or Definitive APD. Because label imbalance and phenotypic overlap complicate

multiclass learning, we targeted evaluation windows with more favorable class distributions: Years 7–11 for binary models and Years 5–9 for multiclass models.

Predictors comprised all baseline features with <15% missingness, alongside demographic (e.g., sex, years of education) and genetic mutation status. For each target year, we trained an independent model to predict APD status at that future timepoint using only baseline inputs. We compared seven algorithms covering kernel-based, neural, tree/boosting, and rule-based families: Support Vector Machines (SVM)(radial-basis and polynomial kernels)[12], multilayer perceptron (MLP)[13], eXtreme Gradient Boosting (XGBoost) [14], Repeated Incremental Pruning to Produce Error Reduction (RIPPER)[15], Partial decision trees (PART)[16] and Logistic Model Tree (LMT)[17]. Implementations used scikit-learn for SVM and MLP, the xgboost library for XGBoost, and the python-weka-wrapper for rule-based and LMT models [18],

Data were split into training (75%) and test (25%) sets. To reduce sampling variance and enhance reliability, we trained over 100 bootstrap iterations (resampling with replacement) on the training set and computed the mean AUC on out-of-bag or held-out samples across iterations. Hyperparameters were tuned within each modeling family, and a fixed random seed (2024) ensured reproducibility. Performance was summarized with area under the receiver operating characteristic curve (AUC-ROC), F1 score[19], and balanced accuracy—the average of sensitivity and specificity— [20] to account for class imbalance. These metrics were used for binary classification. For multiclass problems, macro-averaged metrics were computed across the four classes.

To improve interpretability, we examined feature importance for the best models at each year using SHAP values[21] for tree-based algorithms (e.g., XGBoost). When SHAP was not available (e.g., WEKA-based models), we used Information Gain —equivalent to Mutual Information[22] — as a global importance metric. Together, these procedures provided complementary global insights into the contribution of motor, non-motor, functional, demographic, and genetic features to APD risk stratification.

RESULTS

Cohort Characteristics and Baseline Data

A total of 1,302 individuals with Parkinson's disease (PD) from the PPMI database were included in the study. Clinical visits were scheduled every six months over a 13-year period. Participant retention declined progressively due to attrition, with an average follow-up of 3.9 ± 4.2 years. By year 7, 396 participants remained, decreasing to 224 by year 13 (Figure 2).

The highest dropout rate occurred between baseline and Year 1 (40.5%), while Year 6 showed the lowest attrition (4.4%). The percentage of losses with a registered cause was highly variable across visits, with an average of 52.8% and a range between 26.3% and 84.6%. Missing reasons were more frequent during early visits, potentially reflecting missing-not-at-random patterns. In contrast, from Year 7 onwards, causes of withdrawal were more consistently recorded. The most frequent cause of discontinuation across all visits was death (n = 92), followed by general disinterest (n = 53), Parkinson's-related disability (n = 46), and burden of study procedures (n = 36). The prevalence of each cause varied over time: in early visits (Years 1–4), logistical and motivational reasons

dominated, with burden of study procedures and general disinterest as leading causes. In later visits (Years 10–13), death predominated, followed by disability, while logistical factors became less relevant. These trends are detailed in Supplementary Table 2.

For all analyses, patients were stratified into four groups based on APD certainty levels from the CDEPA criteria: Definitive APD (Def), Probable APD (Pro), Possible APD (Pos), and Not APD (Not). At baseline, the distribution was: Definitive APD (n = 5, 0.4%), Probable APD (n = 88, 6.8%), Possible APD (n = 229, 17.6%), and Not APD (n = 980, 75.3%) (Table 2).

Among the 1,302 patients, 323 (24.8%) had genetically confirmed PD. The most frequent mutations were LRRK2 (n = 173), GBA (n = 100), followed by SNCA (n = 29), PARKIN (n = 12), PINK1 (n = 1), and combined LRRK2 and GBA mutations (n = 8). GBA and SNCA mutations—associated with faster progression and cognitive decline—accounted for nearly 40% of all genetically confirmed cases.

Overall, the cohort's mean age at baseline was 62.8 ± 9.7 years, with a male predominance (61.4%). The average disease duration was 2.2 years, and the median Hoehn & Yahr (H&Y) stage was 2. This profile indicates a predominantly early-stage population, with mild impairments in MDS-UPDRS Part I and II scores (mean 5.0 and 6.6, respectively), and preserved cognition (MoCA score 26.8 ± 2.7). Although only a small fraction of participants met APD criteria at baseline, a substantial proportion progressed to more advanced stages over time. Baseline comparisons revealed a severity gradient across APD groups. Monitoring time and disease duration were significantly longer in Def vs. other groups ($p<0.001$), indicating a latency time bias. Functional disability and non-motor symptoms were significantly greater in Definitive ($p<0.001$), while cognitive performance

was slightly but significantly lower ($p<0.001$). Notably, motor examination scores (MDS-UPDRS Part III) did not differ significantly between groups at baseline, underscoring the greater sensitivity of functional and non-motor assessments for early APD detection. L-dopa equivalent daily dose (LEDD) values did not significantly differ across APD certainty levels ($p = 0.12$), with slightly higher mean doses observed in Definitive and Probable APD groups (Table 2). These findings are consistent with early disease stages, where dopaminergic therapy is generally moderate and does not yet reflect advanced treatment needs

Longitudinal Evolution of APD Labels

Figure 2 shows the evolution of APD classifications. From Year 1 onwards, a shift toward more advanced APD stages emerged. Not APD cases fell below 50% by Year 2 and 15% by Year 7. Probable APD peaked at 39% in Years 3–4, Possible APD at 32% in Years 5–6. Definitive APD rose steadily after Year 4, reaching >50% by Year 10 and 99% by Year 13. Year 7 was a critical window with balanced APD groups: Definitive (27%), Pro (36.1%), Pos (19.9%), Not (17.4%) (Supplementary Table 1). At this point, 102 participants (25.8%) had genetically confirmed PD. GBA and SNCA mutations were overrepresented among those classified as Definitive APD by Year 7. These proportions were consistent with baseline distributions, indicating stable genetic representation across disease evolution.

Clinical comparisons at Year 7 revealed persistent and significant differences across groups. Monitoring time remained longer in Definitive ($p<0.001$), supporting the persistence of a Latency Time Bias. Disease duration followed the same gradient (Definitive > Probable > Possible > Not APD, $p<0.001$). Functional impairment, as measured by MDS-UPDRS Part I (EDL) and Part II, was significantly greater in Definitive

compared to Pro and Pos ($p<0.001$). Non-motor symptoms displayed the same progressive pattern ($p<0.001$). Importantly, motor examination scores (MDS-UPDRS Part III), which did not differentiate groups at baseline, now exhibited clear differences with higher scores in Definitive ($p<0.01$), indicating that motor severity emerges as a key marker of advanced disease at mid-follow-up. Cognitive scores (MoCA) remained lower in Definitive compared to Probable and Possible ($p<0.001$), reflecting cumulative cognitive deterioration. LEDD increased progressively over time, particularly in individuals who transitioned to Definitive APD. While baseline values were comparable across groups, by Year 7, patients classified as Definitive APD exhibited substantially higher LEDD (mean ± SD: 724.8 ± 456.2 mg) compared to other categories (Supplementary Table 1). This dose escalation reflects the increasing complexity of symptom management in advanced stages. By Year 13, 99% of the remaining 224 participants were classified as Definitive APD, underscoring the relentless progression of PD.

Progression of Disability Markers in Future Definitive APD Cases

Figure 3 illustrates the longitudinal trajectories of four clinical variables—MDS-UPDRS Part III (motor severity), MoCA (cognitive performance), Hoehn & Yahr (H&Y) stage (global functional status) and LEDD (dopaminergic treatment burden)— comparing patients who fulfilled criteria for Definitive APD by Year 7 versus those who did not. These measures, while not top predictors in our models, are essential for tracking disease progression. Across all three variables, individuals who eventually transitioned to APD exhibited consistently worse scores. MDS-UPDRS Part III scores were similar at baseline but diverged over time: non-APD participants' scores stabilized, while future APD cases showed a steady increase, reflecting progressive motor deterioration. Similarly, H&Y stage was already higher at baseline in future APD patients and continued to rise. MoCA

scores were lower from baseline in individuals who developed APD and declined more sharply over time. LEDD trajectories diverged progressively, with future Definitive APD cases requiring higher dopaminergic doses, particularly after Year 4, indicating increased treatment needs despite advancing disability. These intergroup differences remained statistically significant throughout the observation period, highlighting the value of these variables for tracking progression. The widening confidence intervals in later years reflected progressive data sparsity due to participant attrition (Figure 3).

Performance of Predictive Classification Models for Definitive APD

While clinical markers delineate the trajectories of disease progression, they exhibit limited standalone predictive power. Therefore, we developed ML models integrating multidimensional baseline data to enhance early identification of Definitive APD. Supervised classification models were developed to predict the future onset of Definitive APD at various follow-up years (Years 5 to 11), using only baseline clinical data (Table 3). Binary classification models, which distinguished Definitive APD from all other categories, consistently outperformed multiclass models. At Year 9, the best-performing binary model (XGBoost) achieved an AUC of 0.88, balanced accuracy of 0.82, and F1 score of 0.82—indicating that it correctly identified approximately 8 out of 10 patients who progressed to APD, while accurately excluding about 9 out of 10 who did not. These results reflect a strong balance between sensitivity and specificity, with a low rate of misclassification. In clinical terms, this performance suggests that the system could reliably flag patients at high risk of advancing to APD up to nine years before onset, enabling earlier therapeutic decisions and referral to

specialized care. In contrast, multiclass models had lower and more variable performance, with AUC values ranging from 0.66 to 0.70 and F1 scores below 0.5 (Table 3).

A detailed evaluation of the best-performing models from Year 9, binary XGBoost and multiclass RBF-SVM, confirmed this. The binary XGBoost classifier demonstrated robust performance (AUC = 0.881, balanced accuracy = 0.824, recall = 0.791, and F1 score = 0.819) with high agreement metrics (Cohen's Kappa = 0.647, MCC = 0.649) (Table 4). This robust performance reflects a favourable balance between sensitivity and precision—crucial features when anticipating the need for device-aided therapies. Its confusion matrix showed low false negatives and acceptable false positives, indicating reliable discrimination (Table 5). Conversely, the multiclass RBF-SVM exhibited lower discriminative capacity (AUC = 0.808, balanced accuracy = 0.620, recall = 0.378, F1 score = 0.357). Its confusion matrix revealed substantial misclassification, especially for early-stage categories, which were never correctly identified (Tables 4 and 5). These findings emphasize the superior performance and reliability of binary classifiers for predicting future Definitive APD in the context of imbalanced and evolving phenotypic distributions.

Variable Importance Across Models

Interpretable models are essential for translating AI predictions into actionable clinical insights. We assessed the relative importance of predictive features across the five best-performing models (from Years 7 to 11), using SHAP values and Information Gain. Figure 4 illustrates the normalized importance of the variables included in the top 10 of each model, ordered by their appearance across models. In total, 28 distinct variables were represented, with Genetic

Mutation Status consistently emerging as the most influential predictor, ranking highest across multiple models and timepoints. This finding supports previous evidence that monogenic forms of PD may follow more aggressive progression trajectories. Other top-ranking variables included age at baseline, tremor severity, REM sleep behaviour disorder (RBD), the total score of MDS-UPDRS Part I, and various measures of autonomic and functional status. The top-ranked variables revealed a relatively balanced contribution from different clinical domains: 12 (42.9%) were non-motor, 11 (39.3%) were motor, and 5 (17.9%) were other variables (genetic, demographic, and functional). This trend became more pronounced at the top of the ranking: non-motor variables accounted for 57.1% of the top 7, suggesting they hold greater predictive value for early APD identification than classic motor items. Within the non-motor category, the most represented subdomains were autonomic dysfunction (e.g., SCOPA-AUT) and sleep disturbances (e.g., REM sleep behaviour disorder, Epworth Sleepiness Scale, and daytime sleepiness). These findings highlight the clinical relevance of dysautonomia and sleep-related dysfunction as early indicators of advanced progression. While several motor variables from MDS-UPDRS Part II and III appeared among the top 28, they were generally distributed in the mid-to-lower part of the ranking. No items from Part IV were selected. Other influential variables included the Hoehn & Yahr stage and the Schwab & England Scale. These findings support a multidimensional view of APD. The consistent prominence of non-motor and systemic features—particularly autonomic function and sleep—reinforces their importance in predictive modelling and patient stratification, moving beyond classic motor-centric frameworks toward more holistic disease staging models.

DISCUSSION

This study presents a novel methodological framework to quantitatively define and predict progression to Advanced Parkinson's Disease by operationalizing expert-driven clinical criteria—specifically the CDEPA questionnaire—through structured clinical data and ML models. Rather than focusing on the development of a single predictive model, the aim was to demonstrate that qualitative, consensus-based definitions can be translated into quantitative, data-driven labels that enable interpretable ML models for early APD classification in longitudinal cohorts. This approach aligns with emerging trends in precision neurology, where expert knowledge is systematically integrated with multimodal datasets to refine disease staging and prognosis [23,24].

Using clinical variables from the PPMI database mapped to CDEPA domains, the models accurately predicted future Definitive APD status up to seven years in advance—addressing a key unmet need in PD management: the lack of standardized, scalable tools to identify patients likely to require therapeutic escalation. While screening tools such as the 5-2-1 criteria [6] or MANAGE-PD[7] are clinically useful, they remain constrained by fixed, qualitative definitions. In contrast, the present framework enables the dynamic calibration and continuous refinement of APD classification, adaptable across cohorts, timepoints, and healthcare systems.

Binary classifiers trained on baseline data alone achieved high discriminative performance (AUC > 0.88), underscoring the prognostic value of early-stage clinical features. As expected, multiclass classification was more challenging, owing to phenotypic overlap in intermediate categories and marked class

imbalance—limitations often encountered in longitudinal PD studies. Feature importance analyses underscored the multidimensional nature of APD. Among the top 28 predictors, 43% were non-motor symptoms, 39% motor symptoms, and 18% functional or staging measures. Non-motor features, particularly autonomic dysfunction and sleep disturbances such as SCOPA-AUT, REM sleep behavior disorder, and Epworth Sleepiness Scale scores, were especially prominent among the top predictors. These findings are consistent with previous evidence suggesting that non-motor manifestations frequently precede overt motor decline and may provide superior prognostic insight[25-27]. Functional and global indicators such as Hoehn & Yahr stage, Schwab & England Scale, and genetic mutation status also contributed, whereas MDS-UPDRS Part IV (motor complications) items were absent, likely reflecting their rarity in early disease. Genetic mutation status emerged as the most influential predictor across all models, reinforcing its relevance in early APD stratification. Beyond its statistical weight, the clinical implications are significant: GBA and SNCA mutations are known to confer more aggressive phenotypes, including earlier cognitive decline, more rapid motor progression, and poorer treatment response. Their overrepresentation among patients who developed Definitive APD by Year 7 highlights their potential utility as early flags for therapeutic escalation. These findings support the integration of genetic profiling into routine risk assessment frameworks, particularly in cohorts with available genotyping. On the other hand, Part I (non-motor) and Part II (motor ADL) items appeared more frequently, while Part III motor examination scores ranked lower—reinforcing earlier observations that traditional motor scales may plateau in prognostic value in more advanced stages. Consistent with this progression, patients classified as Definitive APD by Year 7 also showed a marked increase in LEDD requirements over time, despite

similar baseline values, highlighting the growing pharmacological burden that accompanies clinical deterioration.

Beyond its direct clinical and research implications, this work builds upon and extends a growing body of literature on AI and ML in PD diagnosis, subtyping, and progression prediction. Recent reviews have emphasized the value of integrating multimodal datasets with longitudinal validation to achieve clinically actionable models [24,28,29]. This study shares objectives with previous works that leveraged large-scale PPMI data and ML to transform complex clinical information into predictive models, but differs by focusing specifically on translating CDEPA criteria into quantitative operational definitions from baseline data. Similarly, Gerraty et al. underscored the importance of standardized analytical pipelines and multimodal integration for improving patient classification and prognosis[24]. Our work builds on these principles by translating the CDEPA criteria into structured, model-ready variables, enabling early identification of APD phenotypes within a prospective cohort. The approach also fits squarely within the paradigm described by Abumalloh et al., who reviewed the growing role of AI in PD monitoring through multimodal clinical, sensor-based, and imaging data[29], demonstrating that structured clinical variables mapped to expert definitions can yield interpretable and scalable predictive tools.

Comparisons with more targeted studies further illustrate the novelty of this work. Freire-Álvarez et al. developed a CatBoost model to identify candidates for device-aided therapies, outperforming the 5-2-1 criteria[30]; however, their approach addressed cross-sectional device-aided therapies eligibility, whereas the present study focuses on longitudinal prediction years before APD onset. Hani et al. proposed the PPMI-Benchmark framework for standardizing

preprocessing and synthetic data generation[31]; integrating such methods could enhance the robustness of clinically grounded classification frameworks like ours. Salmanpour et al. combined clinical and SPECT-based radiomics to derive progression subtypes and predict trajectories via unsupervised clustering [32], in contrast, the present work operationalizes an existing, expert-defined construct into a scalable classification tool optimized for long-term binary prediction. Together, these represent complementary strategies—some deriving new phenotypic subgroups from multimodal patterns, others translating consensus-based definitions into reproducible tools for precision prognostication

This study has limitations. The PPMI cohort, although comprehensive, exhibits class imbalance and patient attrition, which affected multiclass model performance and may limit generalizability. Patterns of participant dropout, as detailed in Supplementary Table 2, revealed both random and non-random withdrawal dynamics—with early visits more affected by undocumented reasons and later visits dominated by mortality and disability. These patterns could influence the generalizability and temporal robustness of predictive models. Some CDEPA-aligned variables were missing or inconsistently recorded, requiring imputation and label-rescuing strategies; while methods such as MICE were applied, missing-not-at-random patterns could still bias results. The analysis was restricted to structured clinical data, and future integration of neuroimaging, digital biomarkers, or fluid-based measures may improve accuracy and applicability. Furthermore, APD status was assigned by a single expert, which, while ensuring consistency, may introduce labeling bias.

In conclusion, this study provides proof of concept that expert-defined clinical constructs like APD can be operationalized into objective, longitudinally

consistent labels using structured data, enabling interpretable ML models that support early classification and personalized care. Future work should validate this framework in independent cohorts, integrate multimodal data, and explore advanced modeling strategies such as time-to-event analyses and temporal deep learning. Such developments could help move from static disease definitions to dynamic, data-driven staging systems that better capture the heterogeneous trajectories of PD.

DATA AVAILABILITY STATEMENT

The original dataset analyzed in this study is publicly available from the Parkinson's Progression Markers Initiative (PPMI) at https://www.ppmi-info.org. The derived data and source code used for the analyses presented in this work are available from the corresponding author upon reasonable request and following publication.

AUTHOR CONTRIBUTIONS

Types of authors' roles:

1) Research project: A. Conception, B. Organization, C. Execution;

2) Statistical Analysis: A. Design, B. Execution, C. Review and Critique;

3) Manuscript: A. Writing of the first draft, B. Review and Critique.

I.G. contributed to 1A, 1B, 1C, 2A, 2B, 2C, 3A and 3B

A.S. contributed to 1C, 2B, 2C, 3A and 3B

S.S. contributed to 2A, 2B, 2C, 3A and 3B

A.O. contributed to 2A, 2B, 2C, 3A and 3B

U.Z. contributed to 2A, 2B, 2C, 3A and 3B

I.C. contributed to 2A, 2B, 2C, 3A and 3B

B.T. contributed to 1C, 3A and 3B

T.F.V. contributed to 1C, 3A and 3B

M.R. contributed to 1C, 3A and 3B

M.A. contributed to 3A and 3B

I.S. contributed to 2A, 2B, 2C, 3B

J.C.G.E. contributed to 1A, 1B, 1C, 3A and 3B

R.D.P. contributed to 1A, 1B, 1C, 2A, 2B, 2C, 3A and 3B

COMPETING INTERESTS

All authors declare that they do not have any competing interests and/or personal financial interests regardless of relationship to current manuscript.

FUNDING

This work was supported by the European Commission Horizon Europe Programme under the project AISym4MED: Synthetic and scalable data platform for medical empowered AI (Grant Agreement ID: 101095387, DOI: 10.3030/101095387). I.G. acknowledges funding from Ikerbasque, the Basque Foundation for Science. R.D.P. is supported by the Spanish Ministry of Science and Innovation (grant RYC2024-049664-I) and by Ikerbasque, the Basque Foundation for Science.

REFERENCES

1 Tanner, C. M. & Ostrem, J. L. Parkinson's Disease. *N Engl J Med* **391**, 442-452, doi:10.1056/NEJMra2401857 (2024).

2 Moes, H. R. *et al.* Tools and criteria to select patients with advanced Parkinson's disease for device-aided therapies: a narrative review. *J Neural Transm (Vienna)* **130**, 1359-1377, doi:10.1007/s00702-023-02656-z (2023).

3 Aslam, S., Manfredsson, F., Stokes, A. & Shill, H. "Advanced" Parkinson's disease: A review. *Parkinsonism Relat Disord* **123**, 106065, doi:10.1016/j.parkreldis.2024.106065 (2024).

4 Luquin, M. R., Kulisevsky, J., Martinez-Martin, P., Mir, P. & Tolosa, E. S. Consensus on the Definition of Advanced Parkinson's Disease: A Neurologists-Based Delphi Study (CEPA Study). *Parkinsons Dis* **2017**, 4047392, doi:10.1155/2017/4047392 (2017).

5 Martinez-Martin, P. *et al.* Validation of a simple screening tool for early diagnosis of advanced Parkinson's disease in daily practice: the CDEPA questionnaire. *NPJ Parkinsons Dis* **4**, 20, doi:10.1038/s41531-018-0056-2 (2018).

6 Santos-Garcia, D., de Deus Fonticoba, T., Suarez Castro, E., Aneiros Diaz, A. & McAfee, D. 5-2-1 Criteria: A Simple Screening Tool for Identifying Advanced PD Patients Who Need an Optimization of Parkinson's Treatment. *Parkinsons Dis* **2020**, 7537924, doi:10.1155/2020/7537924 (2020).

7 Antonini, A. *et al.* Validation and clinical value of the MANAGE-PD tool: A clinician-reported tool to identify Parkinson's disease patients

inadequately controlled on oral medications. *Parkinsonism Relat Disord* **92**, 59-66, doi:10.1016/j.parkreldis.2021.10.009 (2021).

8 Shokrpour, S. *et al.* Machine learning for Parkinson's disease: a comprehensive review of datasets, algorithms, and challenges. *NPJ Parkinsons Dis* **11**, 187, doi:10.1038/s41531-025-01025-9 (2025).

9 Marek, K. *et al.* The Parkinson's progression markers initiative (PPMI) - establishing a PD biomarker cohort. *Ann Clin Transl Neurol* **5**, 1460-1477, doi:10.1002/acn3.644 (2018).

10 Myszczynska, M. A. *et al.* Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat Rev Neurol* **16**, 440-456, doi:10.1038/s41582-020-0377-8 (2020).

11 Commission, E. AISym4MED: Synthetic and scalable data platform for medical empowered AI. (European Commission, Horizon Europe Programme, Brussels, Belgium, 2022).

12 Hastie, T., Tibshirani, R. & Friedman, J. in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (ed nd) 417-458 (Springer, 2009).

13 Bishop, C. M. in *Pattern Recognition and Machine Learning* 229-308 (Springer, 2006).

14 Chen, T. & Guestrin, C. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, San Francisco, California, USA, 2016).

15 Cohen, W. W. in *Machine Learning Proceedings 1995* (eds Armand Prieditis & Stuart Russell) 115-123 (Morgan Kaufmann, 1995).

16 Frank, E. & Witten, I. H. in *International Conference on Machine Learning.*

17 Landwehr, N., Hall, M. & Frank, E. Logistic Model Trees. *Machine Learning* **59**, 161-205, doi:10.1007/s10994-005-0466-3 (2005).

18 python-weka-wrapper3 (2025).

19 van Rijsbergen, C. J. *Information Retrieval*. (Butterworth-Heinemann, 1979).

20 Brodersen, K. H., Ong, C. S., Stephan, K. E. & Buhmann, J. M. in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR).* 3121-3124.

21 Shapley, L. S. in *Contributions to the Theory of Games, Vol. 2* (eds H. W. Kuhn & A. W. Tucker) 307-317 (Princeton University Press, 1953).

22 Rossi, F., Lendasse, A., François, D., Wertz, V. & Verleysen, M. Mutual information for the selection of relevant variables in spectrometric nonlinear modelling. *Chemometrics and intelligent laboratory systems* **80**, 215-226 (2006).

23 Strafella, C. *et al.* Application of Precision Medicine in Neurodegenerative Diseases. *Front Neurol* **9**, 701, doi:10.3389/fneur.2018.00701 (2018).

24 Gerraty, R. T. *et al.* Machine learning within the Parkinson's progression markers initiative: Review of the current state of affairs. *Front Aging Neurosci* **15**, 1076657, doi:10.3389/fnagi.2023.1076657 (2023).

25 Wu, Y. H., Lee, W. J., Chen, Y. H., Chang, M. H. & Lin, C. H. Premotor Symptoms as Predictors of Outcome in Parkinsons Disease: A Case-Control Study. *PLoS One* **11**, e0161271, doi:10.1371/journal.pone.0161271 (2016).

26 Martinez-Eguiluz, M. *et al.* Diagnostic classification of Parkinson's disease based on non-motor manifestations and machine learning strategies.

Neural Computing and Applications **35**, 5603-5617, doi:10.1007/s00521-022-07256-8 (2023).

27 Li, X. *et al.* Trends and hotspots in non-motor symptoms of Parkinson's disease: a 10-year bibliometric analysis. *Front Aging Neurosci* **16**, 1335550, doi:10.3389/fnagi.2024.1335550 (2024).

28 Dadu, A. *et al.* Identification and prediction of Parkinson's disease subtypes and progression using machine learning in two cohorts. *npj Parkinson's Disease* **8**, doi:10.1038/s41531-022-00439-z (2022).

29 Abumalloh, R. A. *et al.* Parkinson's disease diagnosis using deep learning: A bibliometric analysis and literature review. *Ageing Res Rev* **96**, 102285, doi:10.1016/j.arr.2024.102285 (2024).

30 Freire-Álvarez, E. *et al.* Artificial intelligence for identification of candidates for device-aided therapy in Parkinson's disease: DELIST-PD study. *Computers in Biology and Medicine* **185**, doi:10.1016/j.compbiomed.2024.109504 (2025).

31 Hani, M., Betrouni, N., Mahmoudi, S. & Benjelloun, M. 246-259 (SCITEPRESS - Science and Technology Publications).

32 Salmanpour, M. R., Shamsaei, M., Hajianfar, G., Soltanian-Zadeh, H. & Rahmim, A. Longitudinal clustering analysis and prediction of Parkinson's disease progression using radiomics and hybrid machine learning. *Quantitative Imaging in Medicine and Surgery* **12**, 906-919, doi:10.21037/qims-21-425 (2022).

Table 1: Summary of APD Classification Based on CDEPA Criteria and PPMI Variables.

CDEPA Domain	CDEPA item	Corresponding PPMI variable(s)	PPMI variable value	CDEPA APD certainty
General characteristics	Evolution time	Time SYMPTOM_DATE to VISIT_DATE	≥ 10 years < 10 years	Probable Not APD
Disability	Daily living activities	Total Modified Schwab & England score	≤ 60% > 60 & <80% ≥ 80%	Definitive Probable Not APD
Treatment-related motor symptoms	Motor fluctuations	Time in OFF state (4.3 MDS-UPDRS IV)	≥ 2 1 0	Definitive Probable Not APD
		Time with Dyskinesia (4.1 MDS-UPDRS IV)	> 25% ≤ 25% & > 0% 0%	Definitive Probable Not APD
		Functional impact of dyskinesias (item 4.2 from MDS-UPDRS Part IV)	4 < 4 & > 0 0 or 101 (u.t.r.)	Definitive Probable Not APD
Disease-related motor symptoms	Dysphagia	Dysphagia (Other clinical features)	1 (Yes) 1 (Yes) 0 (No) or 2	Definitive Probable Not APD
		Chewing and swallowing (2.3 MDS-UPDRS II)	4 3 0	Definitive Probable Not APD
	Falls	Experiencing falls not related to freezing (Freezing & Falls)	≥ 2 < 2	Definitive Not APD
	Freezing of gait	Currently experiencing freezing of gait (Freezing & Falls)	≥ 1 0	Probable Not APD
		Freezing (2.13. MDS-UPDRS II)	≥ 2 0 or 1	Probable Not APD
		Freezing of gait (3.11. MDS-UPDRS III)	≥ 1 0	Probable Not APD
	Dysarthria	Speech (3.1. MDS-UPDRS III)	≥ 3 < 3 or 101	Probable Not APD
	Posture/ Equilibrium	Walking and balance (2.12. MDS-UPDRS II)	≥ 3 <3	Possible Not APD
		Postural stability (3.12 MDS-UPDRS III)	≥ 3 <3 or 101 (u.t.r.)	Possible Not APD
Disease-related non-motor symptoms	Dysautonomia & orthostatic hypotension	SCOPA_AUT - Overall	≥ 6 < 6	Possible Not APD
		Orthostatic_hypotension	1 (Yes) 0 (No)	Possible Not APD
		Postural hypotension (Other clinical features)	1 (Yes) 0 (No)	Possible Not APD
		Lightheadedness on standing (1.12. MDS-UPDRS I)	≥ 1 0	Possible Not APD
	Somnolence	Daytime sleepiness (1.8. MDS-UPDRS I)	≥ 2 > 10	Possible Not APD
		Epworth Sleepiness Scale – Overall Score	< 2 ≤ 10	Possible Not APD
Neuropsychiatric & cognitive symptoms	Dementia/ Cognitive impairment	Cognitive impairment (1.1. MDS-UPDRS I)	≥ 3 < 3 & >0 0	Definitive Possible Not APD
		MoCA Total Score	≤ 15 > 15 & ≤ 26 > 26	Definitive Possible Not APD
	Hallucinations & delirium	Hallucinations and psychosis (1.2. MDS-UPDRS I)	≥ 3 < 3 & > 0 0 or 101 (u.t.r.)	Probable Possible Not APD
	Apathy	Apathy (1.5. MDS-UPDRS I)	≥ 3 < 3 or 101	Possible Not APD

Summary of APD classification rules derived from the CDEPA criteria and their mapping to PPMI variables. Each CDEPA item was linked to one or more corresponding variables from the PPMI dataset, enabling the assignment of APD certainty levels (Not, Possible, Probable, or Definitive) based on structured thresholds. Mapping rules were defined through expert consensus and applied to motor, functional, cognitive, and autonomic measures collected in PPMI. This operationalization allowed a consistent and reproducible translation of the clinical criteria into data-driven labels. The abbreviation “u.t.c.” refers to cases that were unable to be classified due to missing or ambiguous data. Abbreviations: CDEPA, Clinical Diagnostic Criteria for Advanced Parkinson’s Disease; PPMI, Parkinson’s Progression Markers Initiative; APD, Advanced Parkinson’s Disease; MDS-UPDRS, Movement Disorder Society–Unified Parkinson’s Disease Rating Scale; SCOPA-AUT, Scales for Outcomes in Parkinson’s Disease–Autonomic Dysfunction; MoCA, Montreal Cognitive Assessment.

Table 2: Demographic and clinical data of participants in the PPMI study (baseline)

	NAs %	**All patients**	**Definitive APD** (Def)	**Probable APD** (Pro)	**Possible APD** (Pos)	**Not APD** (Not)	**Statistic**	**Group differences**
N	—	1302	5 (0.4%)	88 (6.8%)	229 (17.6%)	980 (75.3%)	—	
Genetically confirmed PD, n (%)†		323 (24.8)	4 (80.0)	50 (56.8)	62 (27.1)	207 (21.1)		
Monitoring time, years	7.4	3.9 (4.2)	9 (5.8)	6.6 (3.8)	2.2 (2.5)	1.1 (3.6)	572.9***	Def-Not*** Def-Pos*** Def-Pro*** Not-Pro**
Age, years	0.3	62.8 (9.7)	62.2 (9.6)	61.8 (10.1)	65.4 (9.2)	62.5 (9.7)	7.0***	Def-Not*
Males, n (%)	0	800 (61.4)	138 (61.6)	194 (60.4)	152(66.7)	316 (59.4)	—	---
Disease duration, years (mean (min, max))	7.4	2.2 (0, 5)	2.9 (0, 2)	2.9 (0, 54)	1.7 (0, 9)	1.7 (0, 9)	17.4***	Def-Not*** Def-Pos*** Def-Pro*** Not-Pro*** Pos-Pro***
LEDD	3.3	159.5 (304.5)	174.9 (322.1)	176.3 (309.5)	135.1 (270.0)	136.4 (300.1)	1.2	----
H&Y, median (IQR)	0.4	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	44.8***	Def-Not*** Def-Pos**
MDS-UPDRS-I Questionnaire	0.5	5.0 (3.6)	5.6 (3.5)	5.2 (4.1)	6 (3.9)	4.2 (3)	16.9***	Def-Not*** Def-Pos*** Def-Pro*** Not-Pro***
MDS-UPDRS-I EDL	1.2	1.6 (2.0)	1.8 (2.1)	1.7 (2.2)	2.2 (2.3)	1.1 (1.4)	21.5***	Def-Not*** Def-Pos*** Def-Pro*** Not-Pro**
MDS-UPDRS-II	0.6	6.6 (4.9)	8.1 (5.7)	7 (5.3)	7.2 (4.9)	5.5 (4.0)	18.8***	Def-Not*** Def-Pos*** Def-Pro*** Not-Pro***
MDS-UPDRS-III	0.38	21.7 (10.0)	21.4 (9.8)	21.2 (11.2)	21.6 (9.6)	22.1 (9.6)	0.6	Def-Not ** Not-Pro*
MDS-UPDRS-IV	80.3	1.8 (2.8)	2.4 (2.9)	1.5 (2.6)	1.3 (2.6)	1.6 (3.0)	2.1	Def-Prob** Def-Pos*
MoCA	6.1	26.8 (2.7)	26.5 (3.3)	26.7 (2.9)	26 (2.7)	7.3 (2.1)	14***	Def-Not ***

Table 2: Demographic and clinical data of participants in the PPMI study at baseline. *p<.05, **p<.01, ***p<.001. Note: Variables are grouped by domain, and baseline values are presented as mean (standard

deviation) for continuous variables and percentages for categorical variables. †Includes patients with genetically confirmed Parkinson's disease: LRRK2 (n = 173), GBA (n = 100), SNCA (n = 29), PARKIN (n = 12), PINK1 (n = 1), and combined LRRK2+GBA (n = 8).Abbreviations: APD = Advanced Parkinson's Disease; MDS-UPDRS = Movement Disorder Society-Unified Parkinson's Disease Rating Scale; EDL = Experiences of Daily Living; MoCA = Montreal Cognitive Assessment; H&Y = Hoehn and Yahr Stage. The "Statistic" column reports the t-value for all comparisons, except for H&Y, where the U-value from Wilcoxon test was used. The value of the statistic is F except for H&Y which is χ^2

Table 3. Performance of classification models predicting future Definite APD from baseline data at different follow-up years.

Year	Model Type	Best Model	AUC	Balanced Accuracy	F1 Score
Y5	Multiclass	LMT	0.7	0.6	0.41
Y6	Multiclass	PART	0.66	0.61	0.41
Y7	Binary	XGBoost	0.87	0.76	0.67
Y7	Multiclass	XGBoost	0.67	0.61	0.41
Y8	Binary	XGBoost	0.8	0.73	0.66
Y8	Multiclass	PART	0.67	0.65	0.44
Y9	**Binary**	**XGBoost**	**0.88**	**0.82**	**0.82**
Y9	Multiclass	SVM (RBF)	0.81	0.62	0.36
Y10	Binary	XGBoost	0.79	0.71	0.78
Y11	Binary	LMT	0.81	0.66	0.79

For each target year (Y5–Y11), supervised classification models were trained to predict whether a patient would meet criteria for Definite APD at that timepoint, using only baseline clinical variables. The table shows the best-performing binary and multiclass classifiers for each year, along with their respective Area Under the ROC Curve (AUC), balanced accuracy, and F1 score. Binary models, which distinguish Definite APD from all other stages, consistently outperformed multiclass models that aimed to classify all four APD categories (Not, Possible, Probable, Definite). Performance variability reflects differences in class imbalance and predictive separability across follow-up years. See Methods and Results sections of the main manuscript for further details. Abbreviations: AUC, Area Under the Curve; F1 score, harmonic mean of precision and recall; XGBoost, Extreme Gradient Boosting; LMT, Logistic Model Tree; PART, Partial Decision Tree; SVM (RBF), Support Vector Machine with Radial Basis Function kernel.

Table 4: Validation metrics for best APD classification models at Year 9

Validation metrics	Binary Model (XGBoost)	Multiclass Model (RBF-SVM)
AUC	0.881	0.808
Balanced Accuracy	0.824	0.62
Recall	0.791	0.378
F1 score	0.819	0.357
Matthews coefficient	0.649	0.459
Cohen's Kappa	0.647	0.427

Performance metrics of the best-performing classification models for Advanced Parkinson's Disease (APD) prediction at Year 9. The Binary Model (XGBoost) classifies participants into Definite APD vs Not Definite APD using an Extreme Gradient Boosting algorithm (XGBoost). The Multiclass Model (RBF-SVM) assigns participants to one of four APD certainty categories (Not, Possible, Probable, Definite) using a Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel, which enables non-linear separation of complex class boundaries. Metrics shown include: AUC (Area Under the Receiver Operating Characteristic Curve), Balanced Accuracy (mean of sensitivity and specificity), Recall (true positive rate), F1 score (harmonic mean of precision and recall), Matthews coefficient (Matthews Correlation Coefficient, MCC), and Cohen's Kappa (agreement coefficient adjusted for chance).

Table 5: Confusion matrices for best APD classification models at Year 9

Binary Model (XGBoost)

	Not Definite APD (0)	Definite APD (1)
True: Not Definite APD (0)	**36** (TN)	6 (FN)
True: Definite APD (1)	9 (FP)	**34** (TP)

Multiclass Model (RBF-SVM)

	Not APD	Pos APD	Pro APD	Def APD
True: Not APD	0	0	4	0
True: Pos APD	0	0	3	0
True: Pro APD	0	0	**31**	**4**
True: Def APD	0	0	**16**	**27**

The Binary Model (XGBoost) discriminated Definite APD from all other categories (Not, Possible, and Probable APD), showing a favorable balance between true positives and true negatives, with low false negatives. The Multiclass Model (RBF-SVM) attempted to classify participants into four CDEPA-derived categories (Not, Possible, Probable, and Definite APD), but showed substantial misclassification, especially for early stages (Not and Possible APD), which were never correctly identified. These results highlight the superior performance and reliability of binary classifiers compared with multiclass approaches for predicting future Definite APD. Abbreviations: APD = Advanced Parkinson's Disease; CDEPA = Clinical Diagnostic Criteria for Advanced Parkinson's Disease; XGBoost = Extreme Gradient Boosting; SVM = Support Vector Machine; RBF = Radial Basis Function; TN = true negative; TP = true positive; FN = false negative; FP = false positive.

Suppl. Table 1: Demographic and clinical data of participants in the PPMI study (year 7)

	NAs %	All patients	Definitive APD (Def)	Probable APD (Pro)	Possible APD (Pos)	Not APD (Not)	Statistic	Group differences
N	—	396	108 (27%)	143 (36.1%)	79 (19.9%)	69 (17.4%)	—	
Genetically confirmed PD, n (%)†	0	136 (34.3)	71 (52.2)	41 (30.1)	16 (11.7)	8 (5.9)		
Monitoring time, years	0	9.2 (2.5	6.6 (2.6	9.7 (1.8	10.3 (1.5	10.6 (1.2	85.1***	Def-Not*** Def-Pos*** Def-Pro*** Not-Pro**
Age, years	0	61 (9.4)	62.4 (10.6)	60.8 (9.2)	61.9 (7.7)	58.5 (9.0)	2.8*	Def-Not*
Males, n (%)	0	247 (62.4)	61 (25.7)	86 (21.7)	54 (13.6)	46 (11.6)	—	
Disease duration, years (mean (min, max))	2.1	2.4 (0, 20)	4.6 (0, 20)	2.6 (0, 12)	0.9 (0, 2)	0.8 (0, 2)	49.5***	Def-Not*** Def-Pos*** Def-Pro*** Not-Pro*** Pos-Pro***
LEDD	0.2	150.8 (301.5)	308.4 (384.1)	132.4 (283.9)	70.7 (226.0)	44.3 (139.2)	15.9***	Def-Not*** Def-Prob*** Def-Pos***
H&Y, median (IQR)	0	2 (1)	2 (0)	2 (1)	2 (1)	1 (1)	21.8**	Def-Not*** Def-Pos**
MDS-UPDRS-I Questionnaire	0.1	4.7 (3.5)	6.7 (3.7)	4.8 (3.5)	3.7 (2.6)	2.7 (2.1)	24.2***	Def-Not*** Def-Pos*** Def-Pro*** Not-Pro***
MDS-UPDRS-I EDL	0.2	1.5 (1.9)	2.6 (2.5)	1.4 (1.6)	1.2 (1.5)	0.5 (1.0)	20.33***	Def-Not*** Def-Pos*** Def-Pro*** Not-Pro**
MDS-UPDRS-II	0.1	6.7 (5.2)	10 (6.5)	6.5 (4.4)	5.2 (3.4)	3.7 (2.6)	30.1***	Def-Not*** Def-Pos*** Def-Pro*** Not-Pro***
MDS-UPDRS-III	0	20 (9.4)	22 (11.3)	20.4 (8.5)	19.6 (8.4)	16.7 (8.1)	4.7**	Def-Not ** Not-Pro*
MDS-UPDRS-IV	21.1	1.9 (2.6)	2.7 (2.9)	0.8 (1.8)	0.2 (0.6)	1 (2.0)	6.5***	Def-Pro** Def-Pos*
MoCA	1.7	26.8 (2.9)	25.9 (4.0)	26.9 (2.5)	26.9 (2.3)	27.8 (2.3)	5.9***	Def-Not***

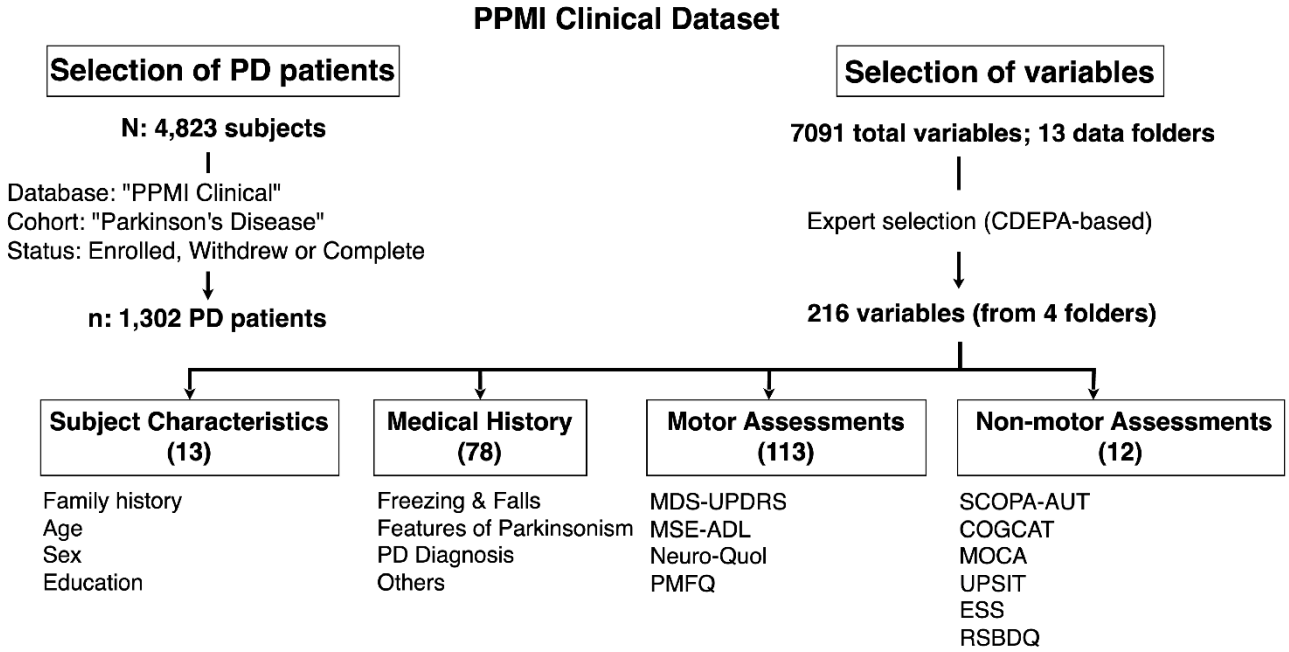
Supplementary Table 1. Demographic and clinical data of participants in the PPMI study at year 7. *p<.05, **p<.01, ***p<.001. Note: Variables are grouped by domain, and baseline values are presented as mean (standard deviation) for continuous variables and percentages for categorical variables. Abbreviations: APD = Advanced Parkinson's Disease; MDS-UPDRS = Movement Disorder Society-Unified Parkinson's Disease Rating Scale; EDL = Experiences of Daily Living; MoCA = Montreal Cognitive Assessment; H&Y = Hoehn and Yahr Stage. The "Statistic" column reports the t-value for all comparisons, except for H&Y, where the U-value from Wilcoxon test was used. †Among the 136 participants with genetically confirmed PD, the most frequent mutations were LRRK2 (n = 74), GBA (n = 40), SNCA (n = 13), PARKIN (n = 5), PINK1 (n = 1), and combined LRRK2 + GBA (n = 3).

Suppl. Table 2. Participant Retention and Causes of Withdrawal Across the 13-Year Follow-Up Period

	Y0	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13
Participants per visit	1302	775	634	580	543	498	435	396	362	338	321	293	244	224
Withdrawals since previous visit	0	527	141	54	37	45	63	39	34	24	17	28	49	20
% Withdrawals from previous visit	0	40.5	18.2	8.5	6.4	9.0	14.4	9.0	8.6	6.6	5.0	8.7	16.7	8.2
Withdrawals with documented cause	0	34	47	42	19	45	26	38	34	21	10	13	18	20
% of withdrawals with documented cause	0	6.5	33.3	77.8	51.4	100.0	41.3	97.4	100.0	87.5	58.8	46.4	36.7	100
Due to adverse event	0	3	1	0	0	0	0	0	0	0	0	0	1	0
Due to death	0	3	5	7	4	10	6	10	6	9	4	5	8	13
Due to family or social issues	0	3	4	4	2	5	3	4	2	1	1	0	0	0
Lost to follow-up	0	1	8	9	2	8	2	5	5	3	1	3	1	2
Due to protocol non-compliance	0	4	1	1	0	1	0	1	1	0	0	0	1	0
Due to travel issues	0	1	8	4	5	5	5	6	4	1	1	3	2	0
Due to burden of study procedures	0	7	9	8	1	1	0	1	3	1	1	1	1	1
Due to PD-related disability	0	4	3	3	2	6	3	4	3	3	2	1	2	4
Due to site closure	0	0	0	0	0	2	2	0	0	0	0	0	0	0
Due to general disinterest	0	7	5	6	1	6	5	7	8	3	0	0	2	0
Other	0	1	3	0	2	1	0	0	2	0	0	0	0	0

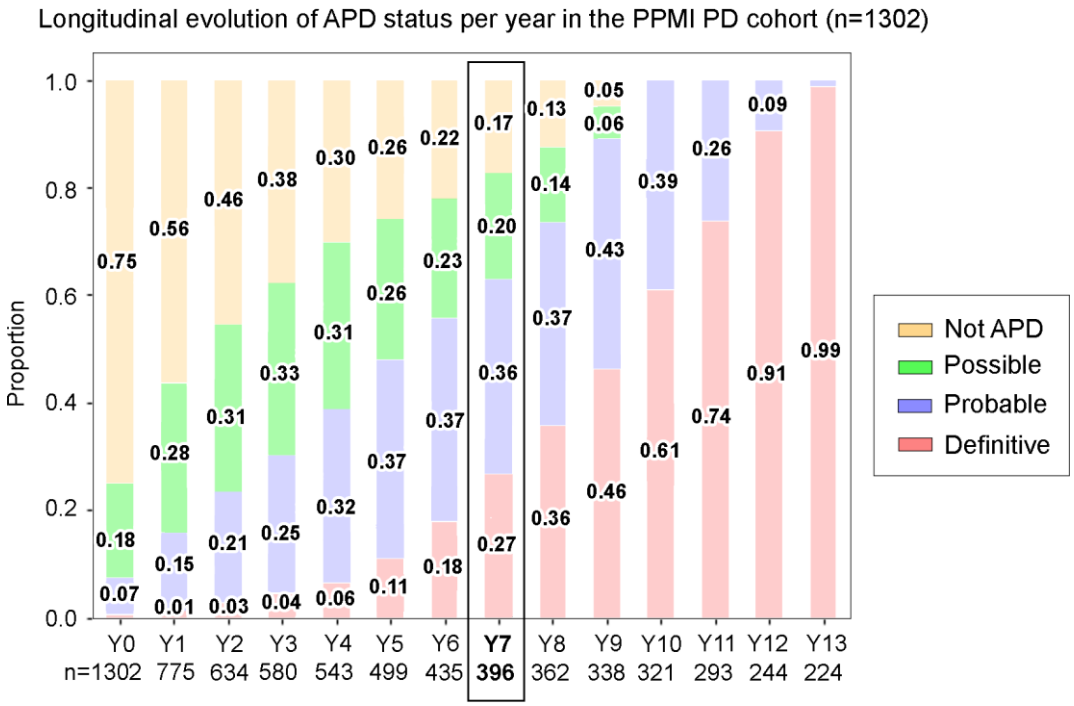
This table summarizes the number of participants completing each scheduled visit (Y0 to Y13), the absolute and relative number of withdrawals between visits, and the causes of withdrawal when documented. The percentage of withdrawals with a documented reason varied across visits, ranging from 6.5% to 100 % relative to total dropouts. The most frequent causes of discontinuation were death, general disinterest, and burden of study procedures. Notably, the proportion of missing reasons was highest in early visits and progressively decreased over time, especially from Year 7

Figure 1. Subject and Variable Selection Workflow from PPMI Data for APD Classification with CDEPA



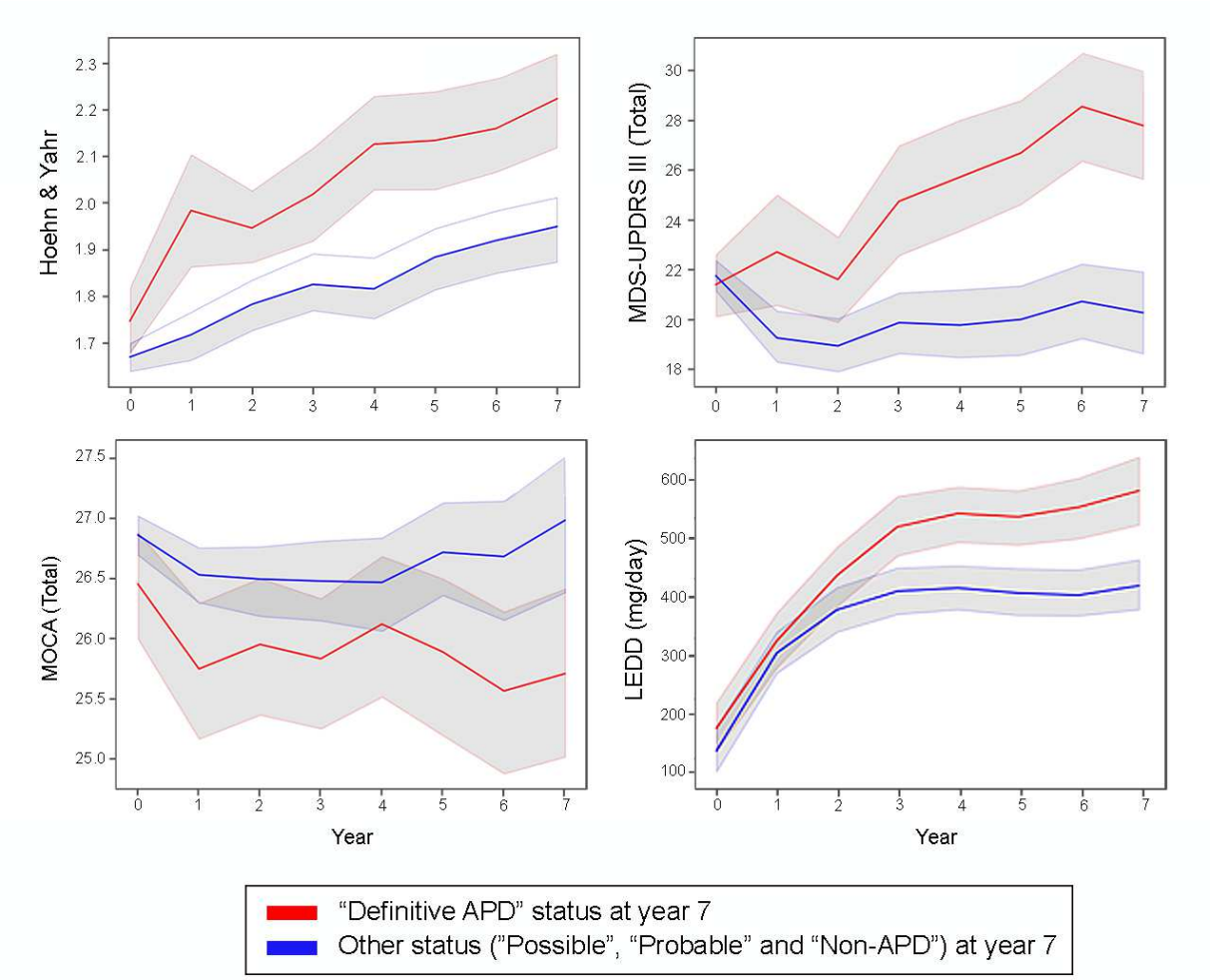
Pipeline for preparing the Parkinson's Progression Markers Initiative (PPMI) dataset for Advanced Parkinson's Disease (APD) classification based on CDEPA criteria. The process includes participant selection, mapping of CDEPA items to structured PPMI variables, and grouping into motor and non-motor assessments. Numbers in parentheses indicate the number of variables in each group. Abbreviations: MDS-UPDRS, Movement Disorder Society – Unified Parkinson's Disease Rating Scale; MSE-ADL, Modified Schwab & England Activities of Daily Living Scale; Neuro-QoL, Quality of Life in Neurological Disorders; PMFQ, Parkinson's Medication Form Questionnaire; SCOPA-AUT, Scales for Outcomes in Parkinson's Disease – Autonomic; COGCAT, Cognitive Categorization Test; MoCA, Montreal Cognitive Assessment; UPSIT, University of Pennsylvania Smell Identification Test; ESS, Epworth Sleepiness Scale; RSBDQ, REM Sleep Behavior Disorder Questionnaire.

Figure 2. Longitudinal Distribution of APD Certainty Levels in the PPMI Dataset



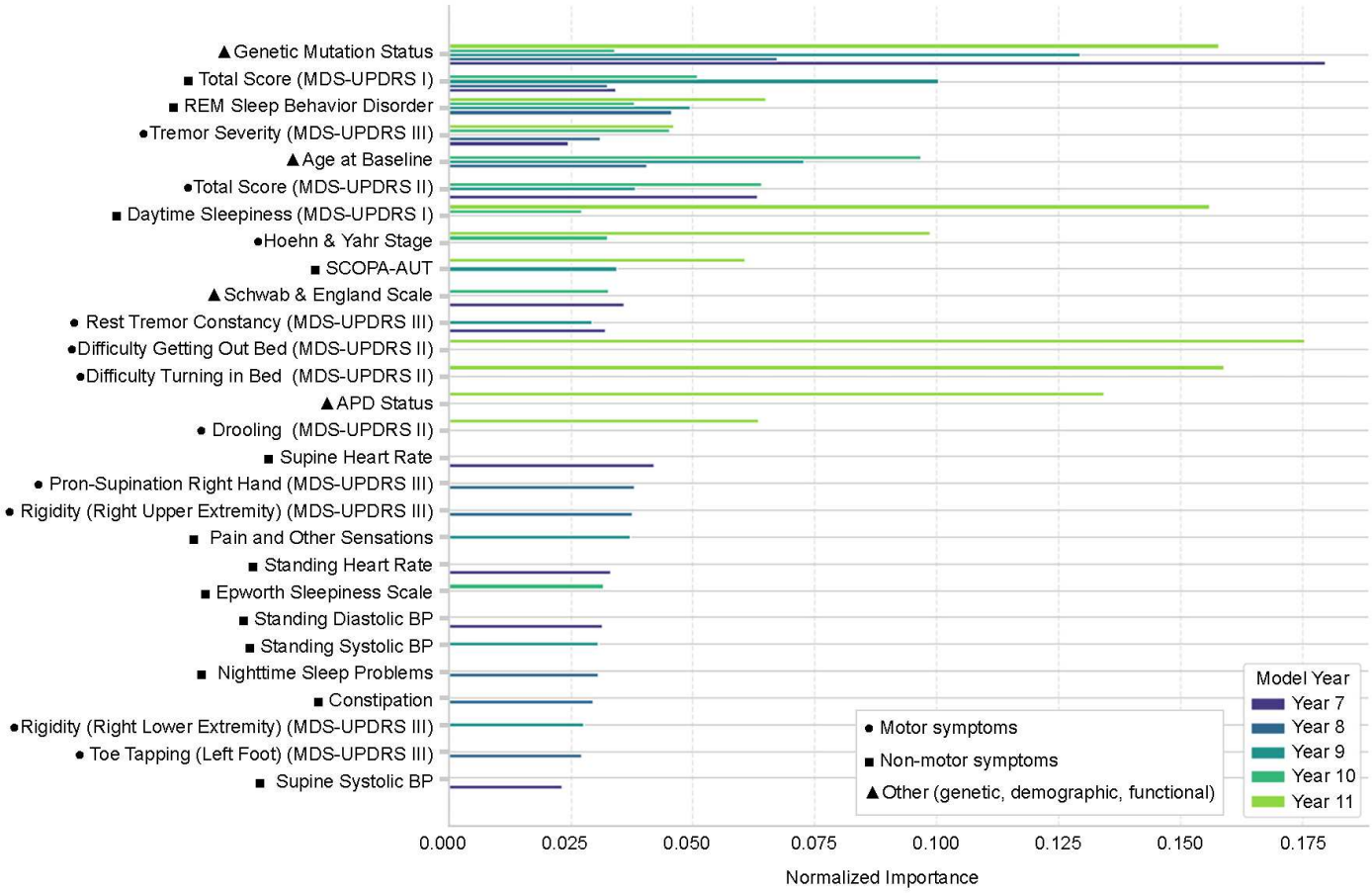
Longitudinal evolution of Advanced Parkinson's Disease (APD) status per year in the PPMI Parkinson's disease cohort (n = 1,302). Bars represent the yearly proportion of participants classified as Not APD, Possible APD, Probable APD, or Definite APD according to CDEPA-based operational definitions. Numbers within bars indicate proportions for each category at each timepoint, and sample sizes (n) per year are shown below the x-axis. Year 7 is highlighted as the main reference timepoint for long-term predictive analyses. The progressive shift toward higher APD certainty levels over time reflects the natural history of disease progression, with widening confidence intervals in later years due to participant attrition. Abbreviations: APD, Advanced Parkinson's Disease; PPMI, Parkinson's Progression Markers Initiative; Y, year from baseline.

Figure 3. Trajectories of key clinical disability markers by Year 7 APD status in the PPMI cohort



Longitudinal trajectories of four clinical markers—Hoehn & Yahr stage (global functional status), MDS-UPDRS Part III total score (motor severity), MoCA total score (cognitive performance), and Levodopa Equivalent Daily Dose (LEDD, mg/day)—comparing participants who met criteria for Definite APD at Year 7 (red) versus those classified as Possible, Probable, or Not APD at Year 7 (blue) in the PPMI cohort. Shaded areas represent 95% confidence intervals. Abbreviations: APD = Advanced Parkinson's Disease; MDS-UPDRS = Movement Disorder Society–Unified Parkinson's Disease Rating Scale; MoCA = Montreal Cognitive Assessment; LEDD = Levodopa Equivalent Daily Dose; PPMI = Parkinson's Progression Markers Initiative.

Figure 4. Top Predictive Features for Long-Term Classification of Definite APD from Baseline PPMI Data



Normalized feature importance for the top 28 baseline predictors in machine learning models classifying Definite Advanced Parkinson's Disease (APD) at Years 7, 8, 9, 10, and 11 in the Parkinson's Progression Markers Initiative (PPMI) cohort. Features are grouped by domain: motor symptoms (●), non-motor symptoms (■), and other variables including genetic, demographic, and functional measures (▲). Importance values were calculated using the Information Gain metric from XGBoost models trained separately for each follow-up year. Higher values indicate greater contribution of the feature to classification accuracy

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SuppFIle.docx