



Personalized progression modelling and prediction in Parkinson's disease with a novel multi-modal graph approach



Jie Lian^{1,2}, Xufang Luo¹✉, Caihua Shan¹, Dongqi Han¹, Chencheng Zhang³✉, Varut Vardhanabhuti²✉, Dongsheng Li¹✉ & Lili Qiu¹

Parkinson's disease (PD) is a complex neurological disorder characterized by dopaminergic neuron degeneration, leading to diverse motor and non-motor impairments. This variability complicates accurate progression modelling and early-stage prediction. Traditional classification methods based on clinical symptoms are often limited by disease heterogeneity. This study introduces a graph-based interpretable personalized progression method, utilizing data from the Parkinson's Progression Markers Initiative (PPMI) and Stroke Parkinson's Disease Biomarker Program (PDBP). Our approach integrates multimodal inter-individual and intra-individual data, including clinical assessments, MRI, and genetic information to make multi-dimension predictions. Validated using the PDBP dataset from 12 to 36 months, our AdaMedGraph method demonstrated strong performance, achieving AUC values of 0.748 and 0.714 for the 12-month Hoehn and Yahr Scale and Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) III on the PPMI test set. Ablation analysis reveals the importance of baseline clinical assessment predictors. This novel framework improves personalized care and offers insights into unique disease trajectories in PD patients.

Parkinson's disease (PD) is a neurological disorder characterized by the gradual degeneration of dopaminergic neurons located in the substantia nigra, leading to debilitating impairments in both motor and non-motor abilities in patients¹. The progression of PD patients shows significant diversity, with some individuals experiencing a relatively mild clinical course while others deteriorate rapidly^{2,3}. The consequent inter-individual and intra-individual variation in the manifestation of PD poses a challenge for disease management and prediction, especially in the early stages. To improve our understanding of disease progression and to make accurate predictions, a personalized approach is essential for precise and effective management⁴.

Traditional methods, which cluster patients based on predominant clinical symptoms, have limitations. In some parts, this has been postulated due to disease heterogeneity and other diverse clinical features such as nonmotor manifestations, which can vary in addition to the traditional motor symptoms that patients are usually classified by. In addition, symptomatology can diverge over time, and as a result, classifications focused on early-year symptoms may differ from those observed in later

years^{2,5,6}. Therefore, failing to account for changes over time may lead to inaccuracies in classifications. In recent research, data-driven methodologies have garnered substantial attention within the domains of disease progression modelling and prediction in PD^{7,8}. To achieve precise and accurate outcome prediction, previous studies have extensively utilized machine learning (ML) techniques⁹. For instance, some research studies have utilized multi-modality data to predict disease progression in terms of assessment scores or cluster patients into different subtypes. Specifically, data including clinical assessments and bio-samples were used to model and predict disease progression using the Parkinson's Progression Markers Initiative (PPMI) and Stroke Parkinson's Disease Biomarker Program (PDBP) dataset^{10–14}. While most of the studies have focused on disease progression based on patient cohorts, we reason that adopting a personalized approach would provide more individualized care to manage PD patients, as well as derive greater insights for individuals to gain a clearer understanding of their unique disease course.

T1-weighted Magnetic Resonance Imaging (MRI) has emerged as a promising tool for tracking structural changes in the brain associated with

¹Microsoft Research, Shanghai, China. ²Department of Diagnostic Radiology, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China. ³Department of Neurosurgery, Clinical Neuroscience Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. ✉e-mail: xufluo@microsoft.com; icczhang.org; varv@hku.hk; dongsheng.li@microsoft.com



PD progression^{15–17}. Longitudinal studies have shown that T1-weighted MRI can pick up on small changes in neurodegeneration caused by Parkinson's disease over time¹⁵. This makes it a useful biomarker for tracking how the disease gets worse. Furthermore, studies have correlated measures derived from T1-weighted MRI, like cortical thickness, with clinical outcomes in PD, demonstrating good test-retest reliability and reproducibility^{16,18}. Therefore, it is imperative to include T1-weighted MRI as an essential modality in PD progression analyses.

To better understand individuals and gain a clearer understanding of their unique disease course, we assume that the exploration of relationships among patients holds potential significance in the development of personalized models for tracking disease progression in PD patients. Such an approach may unveil nuanced connections that may be ignored when focusing solely on individual attributes. Graph methods, including graph neural networks (GNNs), have been commonly used to construct relationships among patients and improve the classification accuracy of individuals in the medical field^{19–21}, exploring multi-modality data and learning patient similarities from various data sources²². Therefore, we aim to introduce a novel graph model for personalized disease progression modelling and prediction from comprehensive aspects in this study. The model can automatically identify important features to construct multiple patient similarity graphs, leveraging a combination of inter-individual and intra-individual data. Our model allows for the development of distinct prediction models tailored to each patient, thus significantly enhancing disease progression modelling and prediction accuracy. Additionally, it accommodates diverse prediction targets, encompassing motor and non-motor labels, to meet personalized requirements for various progression information types. Besides, the inclusion of multimodal inputs, particularly objective ones such as MRI, enriches comprehensive and precise individual status descriptions.

In this study, we have developed an individual-based progression model for PD patients that utilizes graph methods to achieve precise multi-label predictions of disease progression. Our model incorporates diverse data sources, including clinical assessments, medical images, and biological information, enabling us to comprehensively analyse and predict changes of multiple aspects in PD patients over 12, 24, and 36 months.

Results

Study population and the disease progression model

This study uses longitudinal data from the PPMI and PDBP as the main cohorts. We select PD and healthy control (HC) participants from the PPMI based on the availability of baseline T1-weighted Magnetic Resonance Imaging (MRI) scans, gene data, and 36-month follow-up information. For PDBP, a primary diagnosis of PD, the availability of gene data, and the Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) assessments at baseline, as well as 36-month follow-up assessments, are the inclusion criteria in our study (refer to Method and Supplementary 1A and B for PPMI and PDBP participants selection diagrams, and Fig. 1A for study design and inclusion/exclusion of participants). We utilize MRI, clinical assessment, and gene on PPMI, but owing to limited access to MRI scans within the PDBP dataset, our analysis for PDBP was restricted to utilizing gene and clinical assessment data exclusively.

In this study, we have developed two analyses for 12-, 24- and 36-month progression prediction: in the first, both PPMI and PDBP datasets have been used, with patients' baseline clinical assessment and genetic information as input features, comprising a total of 78 variables. In the second, only PPMI is included. In the first analysis, the input features include clinical assessment, MRI, and genetic modalities, totalling 234 variables.

Regarding the 12-month progression, for analysis I, the PPMI training and validation datasets comprise 330 and 82 PD participants, respectively, while the testing dataset consist of 104 PD participants. In comparison, the HC cohort comprises 168 participants, serving as a reference group. PDBP

includes 340 PD participants for external testing. In analysis II, the PPMI training and validation datasets comprise 230 and 57 participants, and the testing dataset includes 74 participants. A summary of the baseline cohort characteristics can be found in Table 1.

To capture the multi-relationship among patients effectively and to make accurate progression prediction, we propose an ensemble model named AdaMedGraph²³. This innovative approach is specifically designed to automatically investigate the relationships between individuals. Specifically, in each step, the AdaMedGraph automatically constructs a set of feature-based similarity graphs, and the most important feature and its corresponding population graph are selected. This approach is interpretable as it automatically discovers the most crucial edge features, facilitating an understanding of the critical relationships among patients. Please refer Fig. 1D for details.

PD progression indicators and scores distribution

The 10 assessment scores in this study held significant meaning for PD patients. Thus, they have been selected as indicators of disease progression, including MDS-UPDRS I, II, and III, and the total score (sum of I, II, and III), as well as the sub-scores of MDS-UPDRS III, including MDS-UPDRS tremors, axial symptoms, and akinetic rigid (rigidity) score (Supplementary 2). Additionally, the Hoehn and Yahr Scale (HY)²⁴, Montreal Cognitive Assessment (MoCA) total score, and Epworth Sleepiness Scale total score (ESS) total score are also included. Considering the distinct clinical meaning and the range of values associated with each assessment score, it is important to establish an objective PD progression label modelling method. In pursuit of this goal, we conducted an approach that encompassed the computation of both the mean and standard deviation (std) of score variations within a healthy control (HC) group for each assessment score (excluding HY). Subsequently, these calculated values are compared with the observed changes in PD patients (see Fig. 1B).

Understanding the trend of progression indicators provides valuable insights into the short-term fluctuations of PD symptoms from various perspectives. As depicted in Fig. 2A, the MDS-UPDRS I, II, III, total, axial, and rigidity scores exhibit a worsening trend from 12 to 36 months. Similarly, the total score of the ESS displays a comparable trend. Besides, the MoCA total score demonstrates an overall decline for all participants from 12 to 36 months, with the steepest decline occurring after 24 months. The above trends indicated a progressive worsening of both motor and cognitive symptoms in the general PD cohort over an extended period, which is consistent with previous knowledge²⁵. Interestingly, the average tremor score remains stable after 12 months, experiences a slight reduction of 0.15 points at 24 months, and subsequently remains constant until month 36. This pattern indicates the absence of significant changes during the early stages of the overall tremor score.

Progression prediction comparison

To evaluate the performance of our proposed model, we compare with a 2-layer multi-layer perceptron (MLP), a logistic regression classifier (LR), a random forest classifier (RF), a support vector machine classifier (SVM), and an XGBoost model. Furthermore, since it is generally recognized that age is an important factor in the progression of Parkinson's disease, we create a graph only based on age (with a threshold of 5, see Supplementary 3) and develop an APPNP (known as APPNP-Age) as a comparative model in the second analysis in the 12-month task. We also use bootstrap techniques to evaluate the statistical significance of the AUC values. Furthermore, Brier scores were reported for all models to make calibration comparison.

In the context of 12-month progression prediction, the AdaMedGraph method demonstrates superior performance compared to our baseline models across all labels, with the exception of MoCA, for both the PPMI and PDBP datasets in analysis I. Specifically, for 12-month HY prediction, AdaMedGraph achieves AUC values of 0.748 (PPMI) and 0.743 (PDBP), surpassing XGBoost (AUC: 0.715, $P > 0.05$ and 0.544, $P < 0.001$), RF (AUC: 0.674, $P < 0.001$ and 0.620, $P < 0.001$), SVM (AUC: 0.704, $P = 0.027$ and

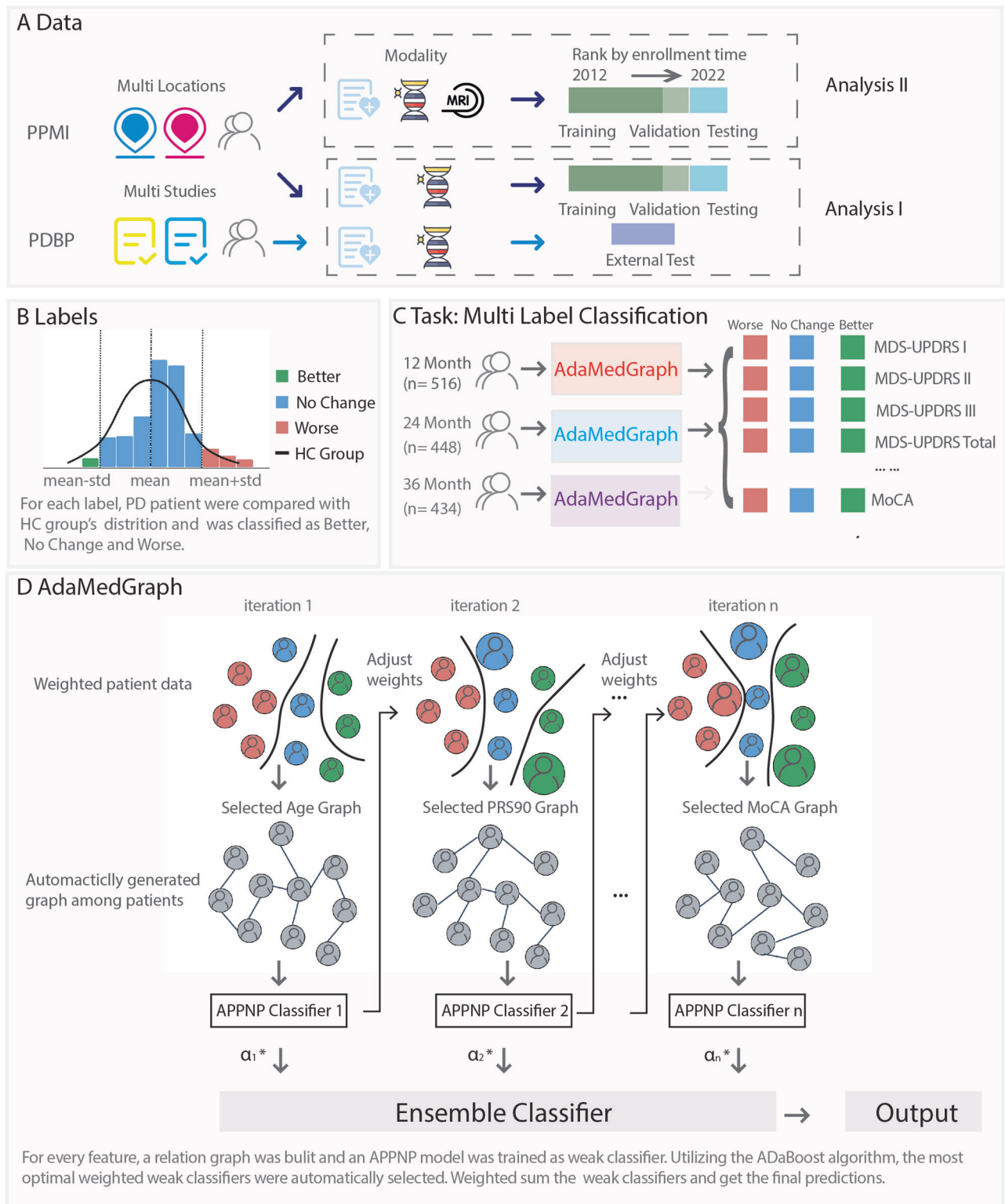


Fig. 1 | Study overview. **A** Study dataset, train-validation-test split, and external testing data; **B** labels of PD patients; **C** 12, 24 and 36 months 10 labels' prediction tasks; **D** Overview of AdaMedGraph.

0.628, $P < 0.001$), LR (AUC: 0.671, $P < 0.001$ and 0.554, $P < 0.001$), and MLP (AUC: 0.626, $P < 0.001$ and 0.433, $P < 0.001$). Additionally, in analysis II, in predicting MDS-UPDRS I, II, and III and total score progression on the PPMI dataset, AdaMedGraph achieves AUC values of 0.712, 0.647, 0.654, and 0.676, respectively, outperforming the baseline models. Notably, our model's superior performance compared to APPNP-Age shows that incorporating multi-relationship considerations can help with accurate

prediction compared to only considering manually selected important relations. More labels and 24, 36-month prediction results and p-value of statistic test can be found in Table 2 and Supplementary 4–13. Additionally, the calibration analyses indicated that our proposed approach achieved the highest Brier score in all tasks except for the 12 and 24 month MoCA on PDBP in analysis I, and the 36 month Tremor in analysis II. Please refer to Supplementary 14–19 for further details.

Table 1 | Demographic and baseline clinical characteristics

	PPMI (Analysis I)	PDBP (Analysis I)	P-values ^a	PPMI (Analysis II)
Sample size (n)	516	340	/	361
Age (years)	61.7, 9.8	64.7, 9.0	$P < 0.001$	62.0, 9.6
Sex (%F)	40.90%	40.90%	$P = 0.997$	37.7
PRS 90	1.37, 1.87	0.62, 1.08	$P < 0.001$	1.04, 1.63
LEDD	/	/	/	124.2, 321.0
MDS-UPDRS I	6.5, 5.0	8.1, 5.4	$P < 0.001$	6.2, 4.6
MDS-UPDRS I Label (Worse, No-change, Better)	14.3%, 65.5%, 20.2%	23.8%, 55.3%, 20.9%	/	12.7%, 66.2%, 21.1%
MDS-UPDRS II	6.7, 4.7	8.8, 6.6	$P < 0.001$	6.3, 4.6
MDS-UPDRS II Label (Worse, No-change, Better)	22.1%, 32.7%, 45.2%	27.3%, 34.1%, 38.5%	/	20.7%, 31.9%, 47.4%
MDS-UPDRS III	20.4, 9.7	22.4, 12.6	$P < 0.001$	20.5, 9.3
MDS-UPDRS III Label (Worse, No-change, Better)	37.8%, 18.0%, 44.2%	41.8%, 21.2%, 37.0%	/	36.0%, 17.2%, 46.8%
MDS-UPDRS Total	33.6, 14.7	39.4, 20.1	$P < 0.001$	33.0, 13.8
MDS-UPDRS Total Label (Worse, No-change, Better)	28.1%, 28.3%, 43.6%	33.8%, 30.3%, 35.9%	/	25.2%, 30.2%, 44.6%
MDS-UPDRS Axial	2.9, 2.3	3.0, 2.9	$P < 0.001$	2.8, 1.9
MDS-UPDRS Axial Label (Worse, No-change, Better)	30.0%, 25.4%, 44.6%	37.0%, 24.7%, 38.2%	/	28.5%, 26.6%, 44.9
MDS-UPDRS Rigidity	13.1, 7.1	14.5, 8.7	$P < 0.001$	13.1, 7.0
MDS-UPDRS Rigidity Label (Worse, No-change, Better)	36.4%, 18.8%, 44.8%	40.0%, 21.2%, 38.8%	/	34.1%, 18.8%, 47.1%
MDS-UPDRS Tremor	3.8, 3.3	4.3, 3.5	$P < 0.001$	4.0, 3.4
MDS-UPDRS Tremor Label (Worse, No-change, Better)	38.2%, 37.8%, 24.0%	42.3%, 30.3%, 27.4%	/	38.8%, 35.2%, 26.0%
MoCA	26.7, 2.8	26.6, 2.9	$P < 0.001$	27.0, 2.3
MoCA Label (Worse, No-change, Better)	18.0%, 69.8%, 12.2%	23.2%, 70.9%, 5.9%	/	17.4%, 71.5%, 11.1%
ESS	6.2, 3.9	7.4, 4.6	$P < 0.001$	6.1, 3.9
ESS Label (Worse, No-change, Better)	9.7%, 70.9%, 19.4%	10.9%, 75.6%, 13.5%	/	7.5%, 74.0%, 18.5%

^aComparison the baseline characteristics for Analysis I cohort between the PPMI and PDBP study, $P < 0.05$ means significant.

The importance of different modalities contributing to PD progression prediction

To further validate the relative importance of input features, we conduct ablation analyses. We systematically remove clinical assessment, MRI, and gene-based features, respectively, to observe their impact on the accuracy of 12-month progression prediction. The results reveal that removing clinical assessment predictors has the most substantial negative effect on accuracy for nearly all labels, with the following AUC values: 0.774 vs. 0.448 for HY, 0.712 vs. 0.641 for MDS-UPDRS I, 0.647 vs. 0.513 for MDS-UPDRS II, 0.654 vs. 0.510 for MDS-UPDRS III, 0.676 vs. 0.575 for MDS-UPDRS total, 0.632 vs. 0.544 for axial, 0.657 vs. 0.579 for rigidity, 0.710 vs. 0.560 for tremor, and 0.677 vs. 0.502 for ESS (Table 3). On the other hand, removing genetic-based features, i.e., 90-SNP polygenic risk score (PRS90) and gene mutations, have the most pronounced negative impact on MoCA prediction accuracy, with AUC values of 0.655 vs. 0.604. Finally, the removal of MRI-based predictive features results in the least reduction in accuracy for all labels except MDS-UPDRS Total, tremor, and HY.

AdaMedGraph suggests important edge (relation) and node (prediction) features

Our AdaMedGraph model demonstrates the capacity to automatically select the most significant relational features (see Method section), allowing us to discern feature importance based on patient similarity. Furthermore, by leveraging the Graph Explainer, we can identify the top contributing node features for prediction in each graph. In the 12-month HY prediction, two edge features, namely baseline left foot toe tapping (TTAPL) and HY, are chosen. For the TTAPL graph, the primary predictive features comprise the left posterior insula, the Hopkins Verbal Learning Test (HVLT) discrimination recognition score, the left amygdala, the right basal forebrain, and postural stability score. On the other hand, for the HY

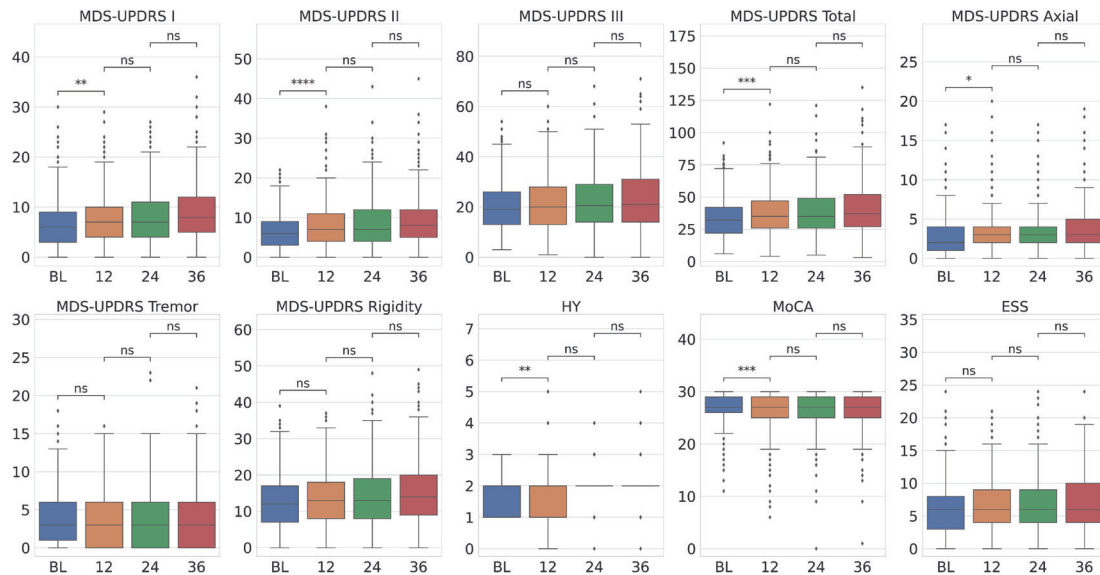
graph, the leading prediction features include left upper extremity rigidity, the right precentral gyrus, the right superior occipital gyrus, fatigue score, and rest tremor amplitude in the right lower extremities.

Gene features correlate with both motor and non-motor labels

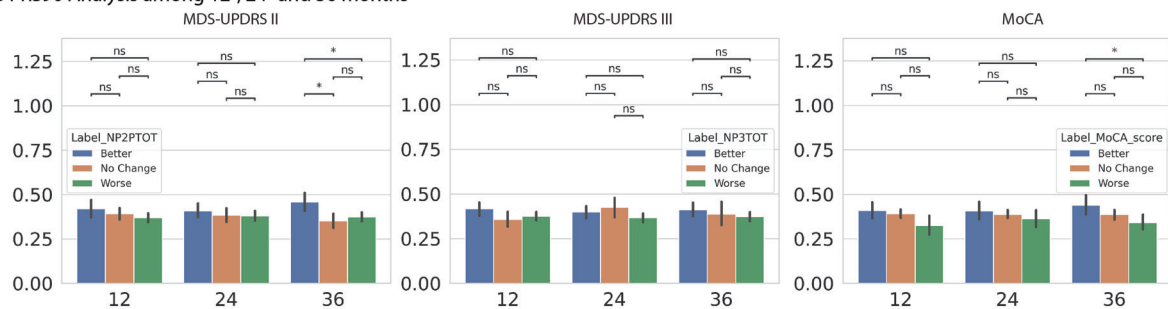
The utilization of PD genetic risk factors has been prevalent in disease prediction, yet the relationship between these disease biomarkers and the progression of the disease remains elusive. To gain further insights, we conduct a comparative analysis of PRS90 among our proposed groups: Better, No Change, and Worse, based on their MDS-UPDRS II, III, and MoCA scores. We have implemented ANOVA tests to compare the differences within the three groups. Interestingly, the Better group consistently exhibits the highest mean PRS90 values for 12-, 24-, and 36-month changes in all three assessment labels. Conversely, no significant differences are observed in the mean scores between the No Change and Worse groups. Notably, when examining the 36-month changes of MoCA and MDS-UPDRS II, the mean PRS value of the Better group is found to be significantly higher than that of the Worse group ($p < 0.05$). Further details can be found in Fig. 2B.

Also, we conduct a comparative analysis between the APOE4 positive and negative groups using the t-test to investigate the influence of APOE4 status on sleep and cognitive function. Regarding cognitive function, we observe a trend of slight improvement in the APOE4 positive group after 12 months, followed by a deterioration after 24 and 36 months compared to the APOE4 negative group. This aligns with the previous conclusion that APOE4 status may be a predictor of cognitive decline in PD²⁶. In terms of sleep patterns, the APOE4 positive group exhibits worsened ESS and day sleep status after 12 months, while the night sleep status displays a slightly better trend. For a comprehensive understanding, please refer to Fig. 2C for further details.

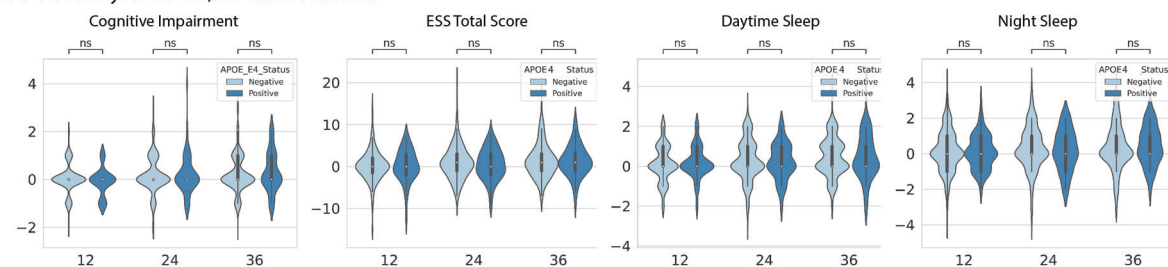
A Label changes among 12-, 24- and 36 months



B PRS90 Analysis among 12-, 24- and 36 months



C APOE4 analyses for 12-, 24- and 36 months



D LEDD changes analysis

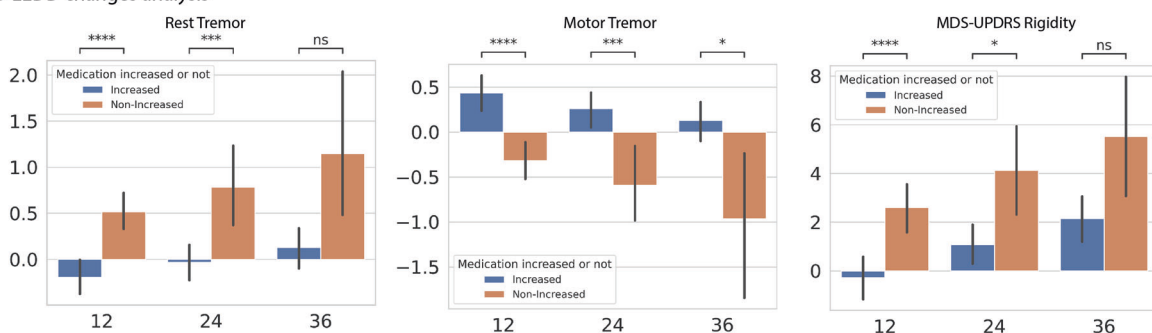


Fig. 2 | Plots of 10 clinical assessment scores changing trends, PRS90 association analyses, APOE4 association analyses and LEDD changes analyses. A Box plots of 10 clinical assessment scores at the baseline, 12-, 24- and 36-month with quartiles. **B** PRS90 association plots of mean and 95% confidence interval with MDS-UPDRS II, III and MoCA after 12-, 24- and 36-month, with (C) APOE4 association plots

with cognitive impairment, ESS total score, daytime and night sleep after 12-, 24- and 36-month. **D** LEDD changes plot of mean and 95% confidence interval with rest tremor, motor tremor and rigidity analysis plot. Statistics: ns: $p > 0.05$, *: $0.001 < p \leq 0.05$, **: $0.0001 < p \leq 0.001$, ***: $0.00001 < p \leq 0.0001$.

Table 2 | Prediction accuracy of 12-month PPMI and PDBP using clinical assessment and gene data.

Label	MLP		LR		SVM		RF		XGBoost		ADaMedGraph	
	PPMI	PDBP	PPMI	PDBP	PPMI	PDBP	PPMI	PDBP	PPMI	PDBP	PPMI	PDBP
HY	0.626	0.433	0.671	0.554	0.704	0.628	0.674	0.620	0.715	0.544	0.748	0.743
MDS-UPDRS I	0.583	0.589	0.653	0.608	0.675	0.634	0.666	0.587	0.672	0.539	0.723	0.649
MDS-UPDRS II	0.583	0.581	0.623	0.509	0.590	0.555	0.621	0.533	0.606	0.577	0.663	0.597
MDS-UPDRS III	0.645	0.527	0.693	0.606	0.711	0.617	0.720	0.590	0.674	0.568	0.714	0.618
MDS-UPDRS Total	0.643	0.543	0.664	0.568	0.658	0.532	0.656	0.569	0.655	0.539	0.688	0.601
MDS-UPDRS Axial	0.400	0.494	0.693	0.599	0.683	0.608	0.694	0.608	0.642	0.560	0.696	0.624
MDS-UPDRS Rigidity	0.732	0.518	0.728	0.619	0.713	0.590	0.719	0.596	0.682	0.582	0.763	0.625
MDS-UPDRS Tremor	0.486	0.536	0.699	0.589	0.643	0.619	0.660	0.621	0.700	0.550	0.706	0.647
MoCA	0.624	0.477	0.718	0.520	0.668	0.517	0.680	0.514	0.665	0.467	0.681	0.517
ESS	0.537	0.532	0.593	0.596	0.585	0.602	0.542	0.567	0.567	0.576	0.622	0.634

Bold: The statistical significance comparing the baseline models with our proposed ADaMedGraph model can be found in Supplementary Table 11.

Table 3 | Prediction accuracy of 12-month PPMI using clinical assessment, gene and MRI, and ablation analysis of modality.

Label	MLP	XGBoost	LR	SVM	RF	APPNP-Age	ADaMedGraph	No Gene	No MRI	No Tabular
HY	0.597	0.668	0.758	0.685	0.622	0.643	0.774	0.759	0.748	0.448
MDS-UPDRS I	0.548	0.662	0.625	0.647	0.666	0.710	0.712	0.657	0.705	0.641
MDS-UPDRS II	0.583	0.559	0.629	0.623	0.538	0.640	0.647	0.624	0.631	0.513
MDS-UPDRS III	0.646	0.557	0.595	0.603	0.567	0.645	0.654	0.582	0.615	0.510
MDS-UPDRS Total	0.524	0.562	0.653	0.666	0.619	0.637	0.676	0.645	0.634	0.575
MDS-UPDRS Axial	0.573	0.600	0.620	0.624	0.538	0.624	0.632	0.604	0.616	0.544
MDS-UPDRS Rigidity	0.641	0.576	0.645	0.647	0.575	0.644	0.657	0.635	0.629	0.579
MDS-UPDRS Tremor	0.638	0.704	0.653	0.672	0.687	0.701	0.710	0.655	0.650	0.560
MoCA	0.660	0.667	0.693	0.676	0.595	0.578	0.655	0.604	0.645	0.613
ESS	0.592	0.610	0.468	0.626	0.611	0.584	0.677	0.588	0.627	0.502

Bold: The statistical significance comparing the baseline models with our proposed ADaMedGraph model can be found in Supplementary Table 8.

Medication association with progression group vs non-progression group

To investigate the relationship between levodopa equivalent daily dose (LEDD) use and symptom progression, we categorize all participants into two groups: the medication increased group (individuals with higher LEDD intake compared to baseline) and the medication non-increased group. T-tests were then utilised to compare the statistical differences between these groups. As shown in Fig. 2D, the medication non-increased group exhibits significantly worsened progression in terms of 12-month and 24-month rest tremor, and rigidity scores. Interestingly, when it comes to motor tremor, all individuals in the medication increased group display significantly worsened performance compared to those in the medication non-increased group.

Discussion

In this study, we present a personalized graph model for PD progression modelling and prediction. Moreover, we have proposed a novel progression label system, allowing our model to make precise disease state prediction. To our knowledge, our approach is the first personalized multi-aspect disease progression prediction model, encompassing several contributions. Firstly, we introduce new graph-based methodologies that offer a new perspective on incorporating diverse data sources for individual-based prediction while also providing interpretable feature importance. This innovative approach makes disease progression models uniquely for each patient since it takes into account each patient's similarity to neighbouring patients and use information from neighbours for prediction. Additionally, our in-depth analyses explore the relationship between genes, changes in medication dosage, and multiple facets of PD progression labels, providing a detailed

understanding of short-term changes in PD symptoms. These insights provide a deeper comprehension of various data types and their contribution to PD progression.

Our proposed model exhibits the capability to select important features, considering both edges (representing patients' similarity) and nodes (indicating prediction contribution). In the context of 12-month HY prediction, we observe that TTAPL and HY are selected as important edges, both of which are clinical assessments. Additionally, the top 5 predictive (node) features, involving a combination of MRI and clinical assessment features, suggests that short-term HY progression is closely associated with patients' baseline clinical assessment scores and MRI information. These findings align with the results of our ablation study, providing additional validation for the model's efficacy in determining feature importance. In the 24-month HY prediction task, the baseline feature HY has been the final selection. HY is a widely used staging scale that encapsulates various symptoms like tremors, rigidity, and reduced arm swing. This outcome suggests that short-term changes in HY are most strongly associated with their own values. Similarly, in the 24-month MDS-UPDRS I task, which focuses on non-motor aspects, our chosen baseline features included daytime sleepiness, volume of the right postcentral gyrus, global spontaneity of movement score and MDS-UPDRS I total score. This ensemble of features imparts valuable insights into the interplay between non-motor changes and variables such as sleep patterns and brain subregions, offering a promising foundation for future investigations in this domain.

In addition, our model can provide explanations at the individual level. For instance, for the 12-month HY prediction TTAPL graph, we analyse patient node 77, which is a "No Change node" with 124 neighbours, and it is

notable that 90 of those neighbours are also No Change nodes. A high degree of clustering of same groups indicates that our automatically generated graph is reliable. (For additional correlation analysis of TTAPL changes and HY, please refer to in Supplementary 20.)

The progression of PD among patients demonstrated noteworthy variations. As indicated in Fig. 2A, while the majority of symptoms worsen over time, certain aspects, such as tremor and MoCA scores, exhibit fluctuations or remain stable within shorter periods, which aligns with findings from recent studies^{1,27,28}. The trend in tremor scores can be attributed to the heterogeneity within PD, where certain patients may experience worsening of rest tremor while simultaneously showing improvements in motor tremor.

The association of PD disease genetic biomarkers with the detailed progression of specific symptom labels has remained elusive. To have more insights, we present compelling evidence that a higher PRS90 for PD is linked to favourable short-term outcomes, characterized by either symptom improvement or stability, in both motor and non-motor aspects. These findings are consistent with observations from previous research¹³. This association may be mediated through the established relationship between genetic risk and early-onset disease, as well as the link between early-onset disease and a slower disease progression²⁹. Our analyses of sleep and cognitive impairment also align with previous research²⁶, which suggest that APOE4 status may be a predictor of cognitive decline in PD. Such insights contribute to our understanding of the interplay between genetics and PD progression.

Our analysis of LEDD medication reveals a noteworthy positive impact on rigidity. In contrast, for motor tremor, an increase in LEDD shows a non-positive effect, and rest tremor exhibits an inverse correlation at 12- and 24-months, with no significant difference observed at 36 months. These variations in tremor patterns can be attributed to the fact that tremor severity is not solely determined by dopamine deficiency but is also influenced by other factors that affect its response to dopaminergic treatments³⁰. (For additional correlation analysis of LEDD changes, please refer to Supplementary 21). Furthermore, the findings showed PD patients with non-increased LEDD had experienced significantly worse progression in terms of rest tremor and rigidity after 12 and 24 months, which is consistent with the broader understanding of PD progression and medication management. PD is a neurodegenerative condition characterised by the gradual death of dopaminergic neurons in the substantia nigra of the brain. This cumulative loss causes a drop in dopamine levels, which is directly associated with the worsening of motor symptoms³¹. Patients whose LEDD did not increase over time probably received insufficient dopamine supplementation, resulting in decreased control of rest tremor and stiffness. This is consistent with the understanding that ongoing assessment and adjustment of PD medications are crucial for managing the progressive worsening of symptoms³².

This study presents several limitations that should be acknowledged. Firstly, our focus is only limited to 12-, 24-, and 36-month progression intervals due to the limited number of longer follow-up cases. Extending the analysis to encompass a more extended timeline might reveal additional trends and patterns in disease progression, contributing to a more comprehensive understanding of PD's long-term dynamics and aiding in the development of better management strategies for patients. Secondly, the heterogeneity of treatment status within the available cohort results in a mixture of ON and OFF state MDS-UPDRS measurements, making it challenging to confidently control for the influence of medication effects, which could potentially impact MDS-UPDRS scores and progression status assignment. Lastly, the method uses for train-validation-test separation led to unequal distributions in the training, validation, and testing datasets, posing challenges to prediction accuracy. These limitations should be considered while interpreting the results and implications of this study.

In considering avenues for future research, several important directions emerge. Firstly, our current model assumes a static representation of each patient as a single node, without accounting for

longitudinal changes over time. Our future work will focus on enriching our model by incorporating temporal information. This will enable a more dynamic and nuanced understanding of disease progression. Secondly, the impact of levodopa therapy on MDS-UPDRS III tremor scores is a significant consideration in PD management³³. Recognising this, we intend to further investigate this relationship by stratifying the PD cohort based on levodopa treatment status, leveraging a larger dataset. This approach will provide insights into how treatment influences symptomatology and disease trajectory. Lastly, the challenge of missing data is common in PD studies and poses a barrier to robust analysis. In our future endeavours, we aim to tackle this issue by integrating missingness predictors into our analytical framework. By doing so, we can improve the accuracy and reliability of our findings, ensuring a more comprehensive understanding of PD progression.

In conclusion, our study introduces a novel approach for personalized progression modelling and prediction in PD, effectively addressing the challenge of disease heterogeneity among individuals. By incorporating multi-modal data, considering both intra-individual and inter-individual variability, and accounting for medication effects, our method offers a promising avenue for improving individualized care for PD patients. It enables a comprehensive framework for more accurate outcome predictions and the identification of critical features that impact individual short-term disease progression, thereby shedding light on patients' unique disease trajectories. Our model not only supports the concept of heterogeneous disease progression but also holds valuable implications for patient stratification and management. As we look towards future validation in real-world cohorts, there is significant potential for our model to serve as a valuable prognostication tool in clinical practices, thereby contributing to gaining insights into PD progression and aiding in enhancing patient care.

Methods

Data

PPMI and PDBP data have been used in this study. The PPMI study is a comprehensive observational study conducted on a precisely defined cohort of participants (for a detailed description, please refer to the website)³⁴. PPMI enrolled individuals diagnosed with early-stage PD within the past two years in 24 distinct sites. At the baseline visits (first visit to the assessment centre), the PD participants had not yet received treatment, and they were not expected to require symptomatic therapy for the next six months. PDBP is a collection of studies, with each having its own set of inclusion and exclusion criteria³⁵.

Our study data has been retrieved from the AMP-PD platform (<https://amp-pd.org/>) using the terms of the AMP-PD Data Use Agreement³⁶ and the PPMI website (<https://www.ppmi-info.org/>). The data used in this study have been collected before the commencement of our study and were obtained in an anonymized format. In the original PPMI and PDBP studies, participants had previously consented in writing to the sharing of their data. These investigations have been conducted in accordance with protocols authorized by the Indiana University Institutional Review Board (IRB) for PPMI and the respective PDBP centres.

Input modality and processing

To gain a deeper understanding of different modalities in PD progression and make accurate disease progression predictions, our analyses focus on the following multiple modalities extracted from the PPMI dataset: gene data, MRI, and clinical assessment data (please refer to Supplementary 22). In addition, demographic data of age and sex are also used as input.

With the aim of comprehensively capture the diverse range of symptom manifestations, our study utilizes a wide range of clinical assessments as input, including (1) MDS-UPDRS parts I, II, and III, along with their respective subpart scores, (2) Modified Schwab and England Percent Activities of Daily Living (SE-ADL), (3)HY, (4) MoCA total score (adjusted as described in Supplementary 23), (5) ESS score and status, and (6) measures of autonomic function, cognitive function, and impulse control (refer to Supplementary 22 for detailed information), and (7) Dopamine Transporter Scan (DaTScan)

striatum binding ratios (calculated previously). Furthermore, we accounted for medication effects by converting information on dopaminergic medication into the LEDD. Due to the lack of (6) and (7) in the PDBP dataset, we only included (1) to (5) for PDBP-related predictions.

In terms of gene data, we included a PRS90 for PD³⁷ (calculated by AMP-PD), as well as the binary representation of monogenic mutation status for GBA, LRRK2, SNCA, and APOE4. To assess the predictive capability of MRI, we have employed the MALP-EM³⁸ software to perform whole-brain segmentation on our T1 weighted MRI scans (see Supplementary 24 for details) and obtained the volumes of 138 brain segments (see Supplementary 25) as potential MRI-based disease progression biomarkers.

Image processing

We utilized the standard processing process of MALP-EM for whole brain segmentation. All T1-weighted MRI scans were corrected for bias field inhomogeneity using N4 bias correction methods, then brain extraction before being registered and labelled. After that, the brain was segmented into 138 regions. The volumes of the segmented regions were determined directly from the segmentation results. We performed visual checks to ensure that the segmentations were accurate and consistent. The segment volume values were normalized using z-score normalization method with a mean of 0 and a standard deviation of 1.

Study design

For analyses I, we partition 80% of the PPMI data as the training set, while reserving 20% of the training set as validation and an additional 20% as the internal testing set. The PDBP dataset is employed as the external testing dataset. For Analyses II, we exclusively utilize the PPMI data, with an 80% training and validation (8:2) and 20% testing split. To closely mimic real-world scenarios, we perform the train-validation-test separation for the PPMI dataset based on their HY label distribution, ensuring that the proportions of the three groups (i.e., ‘Worse,’ ‘No Change,’ and ‘Better’) are consistent among the train-validation-test sets (please see the Progression Label Modelling section for details). Subsequently, we rank these three groups based on their visiting time separately. The earlier-visiting participants of three groups are allocated to the training and a validation set is randomly sampled from the training set, while the later-visit participants are assigned to the testing set.

Progression label modelling

For each label of each task (e.g. 12-month MoCA), expect HY, we construct a reference interval based on that label scores’ progression in the HC group, using the mean and standard deviation of HC changes. Subsequently, we compare each PD patient’s assessment score change with the reference interval. The patient’s label is categorized as “No Change” if the patient falls within the reference interval, or as “Better” or “Worse” if the patient demonstrates less or greater changes compared to the reference interval (mean $\pm$ std), respectively (Please refer Supplementary 26 for detail of reference interval thresholds).

Since each score follows its own trend (e.g., MoCA scores tend to decline with age), it is essential to consider these trends when defining labels. The reference interval is designed to account for such assessment-based trends, thus avoiding bias. By incorporating this approach, we achieve a personalized and unbiased assessment of PD patients’ progression.

Given that the HY is a widely recognized staging system for PD, we designate a PD patient’s progression status as “Worse” if their HY score increases, “No Change” if the staging score remains unchanged, and “Better” if the HY values decreases. In this case, we establish ten separate 3-class classification tasks, specifically targeting the time points of 12, 24, and 36 months individually.

Model

GNNs have gained considerable attention in the medical domain owing to their versatility and efficacy. Nevertheless, the construction of an appropriate graph structure remains crucial for ensuring the success of subsequent analyses. Prior research has often relied on human knowledge to manually define the graph’s structure^{39,40}, which is challenging especially for complex diseases like PD.

The AdaMedGraph method utilizes the AdaBoost algorithm while employing the GNN (Approximate Personalized Propagation of Neural Predictions (APPNP) in our study) as the weak classifier. The entire procedure of our algorithm compasses multiple rounds of iterations. In each iteration n , every input feature automatically generates one relational graph. These relational graphs represent each patient as a node, utilizing their baseline features as node attributes. The establishment of connections (edges) between these nodes is contingent upon the similarity of the relational features. Subsequently, each of these relational graphs is integrated into the training of an APPNP classifier. In the following steps, our algorithm autonomously chooses the most precise classifier as the weak classifier of iteration n and calculates its importance α_n based on its error rate. After all the iterations, we get our final classifier by calculating the weighted sum of the output of each weak classifier. Besides the weights of weak classifier, α_n also presents the importance of the associated relational graph feature.

We explain the graph construction process in more detail: For each feature, we construct a graph where each patient is represented as a node based on that feature as follows. To establish the correlation edges, we compare the feature values (the feature to build graph) of each node (patient) with all other nodes. If the difference in feature values between two patients falls within a predefined threshold, they are connected by an edge; otherwise, no connection is established between them. Consequently, we generate multiple feature correlation graphs by iterating over all features, and each patient serves as a node in a single graph. We trained our APPNP graph classifiers using those graphs. Please refer to Fig. 1 and Supplementary 27 for further details.

Statistics analysis

We conduct several sets of analyses to explore the potential correlation between medication and gene data on individual participants’ disease progression. Firstly, we compare the differences in scores between the increased medication and non-increased medication groups. Next, we evaluate the impact of APOE4 mutation status on patients’ changes in day/night sleep and cognitive impairment scores. Furthermore, we investigate the disparities in PRS90 values among the Worse, No Change, and Better groups concerning changes in MoCA, HY, and MDS-UPDRS total scores. For all these analyses, we employ ANOVA and two-sided t -test to assess statistical significance.

Data availability

The data utilized in this article was from the publicly accessible Parkinson’s Progression Markers Initiative database (www.ppmi-info.org/data) and the APM-PD platform (<https://amp-pd.org/>). The code in this paper is available via <https://github.com/SereneLian/AdaMedGraph>.

Received: 5 February 2024; Accepted: 7 November 2024;

Published online: 01 December 2024

References

1. Armstrong, M. J. & Okun, M. S. Diagnosis and Treatment of Parkinson Disease: A Review. *JAMA* **323**, 548–560 (2020).
2. Kalia, L. V., Kalia, S. K. & Lang, A. E. Disease-modifying strategies for Parkinson’s disease. *Mov. Disord.* **30**, 1442–1450 (2015).

3. Postuma, R. B. et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov. Disord.* **30**, 1591–1601 (2015).
4. Frohlich, H. et al. Leveraging the Potential of Digital Technology for Better Individualized Treatment of Parkinson's Disease. *Front Neurol.* **13**, 788427 (2022).
5. Fereshtehnejad, S. M. et al. New Clinical Subtypes of Parkinson Disease and Their Longitudinal Progression: A Prospective Cohort Comparison With Other Phenotypes. *JAMA Neurol.* **72**, 863–873 (2015).
6. Mestre, T. A. et al. Parkinson's Disease Subtypes: Critical Appraisal and Recommendations. *J. Parkinsons Dis.* **11**, 395–404 (2021).
7. Mestre, T. A. et al. Reproducibility of data-driven Parkinson's disease subtypes for clinical research. *Parkinsonism Relat. Disord.* **56**, 102–106 (2018).
8. Pourzinal, D. et al. Systematic review of data-driven cognitive subtypes in Parkinson disease. *Eur. J. Neurol.* **29**, 3395–3417 (2022).
9. Landolfi, A. et al. Machine Learning Approaches in Parkinson's Disease. *Curr. Med Chem.* **28**, 6548–6568 (2021).
10. Dadu, A. et al. Identification and prediction of Parkinson's disease subtypes and progression using machine learning in two cohorts. *NPJ Parkinsons Dis.* **8**, 172 (2022).
11. Lawton, M. et al. Developing and validating Parkinson's disease subtypes and their motor and cognitive progression. *J. Neurol. Neurosurg. Psychiatry* **89**, 1279–1287 (2018).
12. Li, J. et al. Cortical and subcortical morphological alterations in motor subtypes of Parkinson's disease. *NPJ Parkinsons Dis.* **8**, 167 (2022).
13. Sadaei, H. J. et al. Genetically-informed prediction of short-term Parkinson's disease progression. *NPJ Parkinsons Dis.* **8**, 143 (2022).
14. Severson, K. A. et al. Discovery of Parkinson's disease states and disease progression modelling: a longitudinal data study using machine learning. *Lancet Digit Health* **3**, e555–e564 (2021).
15. Yang, J., Burciu, R. G. & Vaillancourt, D. E. Longitudinal Progression Markers of Parkinson's Disease: Current View on Structural Imaging. *Curr. Neurol. Neurosci. Rep.* **18**, 83 (2018).
16. Gracien, R. M. et al. Longitudinal quantitative MRI assessment of cortical damage in multiple sclerosis: A pilot study. *J. Magn. Reson Imaging* **46**, 1485–1490 (2017).
17. Shu, Z. Y. et al. Predicting the progression of Parkinson's disease using conventional MRI and machine learning: An application of radiomic biomarkers in whole-brain white matter. *Magn. Reson. Med.* **85**, 1611–1624 (2021).
18. Zhang, Q. et al. White matter biomarker for predicting de novo Parkinson's disease using tract-based spatial statistics: a machine learning-based model. *Quant. Imaging Med. Surg.* **14**, 3086–3106 (2024).
19. Aittokallio, T. and B. Schwikowski, Graph-based methods for analysing networks in cell biology. *Brief. Bioinform* **7**, 243–255 (2006).
20. Gregorich, M. et al. Individual-specific networks for prediction modelling - A scoping review of methods. *BMC Med. Res. Methodol.* **22**, 62 (2022).
21. Lian, J. et al. Early stage NSCLS patients' prognostic prediction with multi-information using transformer and graph neural network model. *Elife* **11**, e80547 (2022).
22. Wu, L., et al. Graph neural networks: foundation, frontiers and applications. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (ACM, 2022)*.
23. Lian, J. et al. AdaMedGraph: Adaboosting Graph Neural Networks for Personalized Medicine. Preprint at <https://arxiv.org/abs/2311.14304> (2023).
24. Goetz, C. G. et al. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov. Disord.* **19**, 1020–1028 (2004).
25. Holden, S. K. et al. Progression of MDS-UPDRS Scores Over Five Years in De Novo Parkinson Disease from the Parkinson's Progression Markers Initiative Cohort. *Mov. Disord. Clin. Pr.* **5**, 47–53 (2018).
26. Umeh, C. C. et al. APOE4 Allele, Sex, and Dementia Risk in Parkinson's Disease: Lessons From a Longitudinal Cohort. *J. Geriatr. Psychiatry Neurol.* **35**, 810–815 (2022).
27. Patel, B. et al. A comprehensive review of the diagnosis and treatment of Parkinson's disease dysphagia and aspiration. *Expert Rev. Gastroenterol. Hepatol.* **14**, 411–424 (2020).
28. Yu, Z. et al. The significance of uric acid in the diagnosis and treatment of Parkinson disease: An updated systemic review. *Medicine* **96**, e8502 (2017).
29. Wickremaratchi, M. M., Ben-Shlomo, Y. & Morris, H. R. The effect of onset age on the clinical features of Parkinson's disease. *Eur. J. Neurol.* **16**, 450–456 (2009).
30. Pasquini, J. et al. Progression of tremor in early stages of Parkinson's disease: a clinical and neuroimaging study. *Brain* **141**, 811–821 (2018).
31. Sako, W. et al. Comparative efficacy and safety of adjunctive drugs to levodopa for fluctuating Parkinson's disease - network meta-analysis. *NPJ Parkinsons Dis.* **9**, 143 (2023).
32. Jenner, P. Treatment of the later stages of Parkinson's disease - pharmacological approaches now and in the future. *Transl. Neurodegener.* **4**, 3 (2015).
33. Luo, S. et al. Dissecting the Domains of Parkinson's Disease: Insights from Longitudinal Item Response Theory Modeling. *Mov. Disord.* **37**, 1904–1914 (2022).
34. Marek, K. et al. The Parkinson's progression markers initiative (PPMI) - establishing a PD biomarker cohort. *Ann. Clin. Transl. Neurol.* **5**, 1460–1477 (2018).
35. Gwinn, K. et al. Parkinson's disease biomarkers: perspective from the NINDS Parkinson's Disease Biomarkers Program. *Biomark. Med.* **11**, 451–473 (2017).
36. Iwaki, H. et al. Accelerating Medicines Partnership: Parkinson's Disease. Genetic Resource. *Mov. Disord.* **36**, 1795–1804 (2021).
37. Nalls, M. A. et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
38. Ledig, C. et al. Robust whole-brain segmentation: application to traumatic brain injury. *Med. Image Anal.* **21**, 40–58 (2015).
39. Kim, S. Y. Personalized Explanations for Early Diagnosis of Alzheimer's Disease Using Explainable Graph Neural Networks with Population Graphs. *Bioengineering* **10**, 701 (2023).
40. Rakhimberdina, Z., Liu, X. & Murata, A. T. Population Graph-Based Multi-Model Ensemble Method for Diagnosing Autism Spectrum Disorder. *Sensors*, **20**, 6001 (2020).

Author contributions

J.L.: Conceptualization, Methodology, Data curation, Analysis, Validation, Visualization, Software, Writing- Reviewing and Editing. X.L.: Conceptualization, Methodology, Data curation, Analysis, Validation, Visualization, Software, Writing- Reviewing and Editing, Supervision. C.S.: Methodology, Analysis, Validation, Visualization, Software, Writing- Reviewing and Editing. Dongqi Han: Methodology, Analysis, Validation, Visualization, Software, Writing- Reviewing and Editing. C.Z.: Conceptualization, Validation, Writing- Reviewing and Editing and Supervision. V.V.: Conceptualization, Validation, Writing- Reviewing and Editing and Supervision. D.L.: Conceptualization, Validation, Writing- Reviewing and Editing and Supervision. L.Q.: Conceptualization, Validation,

Reviewing, Writing-Editing, and Supervision. All authors have directly access to the data reported in the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41531-024-00832-w>.

Correspondence and requests for materials should be addressed to Xufang Luo, Chencheng Zhang, Varut Vardhanabhuti or Dongsheng Li.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024