Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Cross-Cohort Transportability of Multi-Family Resting-State EEG Microstate Dynamics in Parkinson's Disease

Yuxiang Wang
Taiyuan Normal University
Guimei Yin
yinguimeicn@126.com

Taiyuan Normal University
Guangxing Guo
Taiyuan Normal University
Ming Liu
Taiyuan Normal University
Zihao Wei
Taiyuan Normal University
Yukun Wang
Taiyuan Normal University
Anqi Li
Taiyuan Normal University
Yanli Zhao
Capital Medical University Affiliated Beijing Huilongguan Hospital
Shuping Tan
Capital Medical University Affiliated Beijing Huilongguan Hospital
Hua Wei
Suzhou University of Technology
Bin Wang
Taiyuan University of Technology

Research Article

Keywords: Parkinson’s disease, electroencephalography, EEG microstates, brain-state dynamics, cross-cohort transportability, distribution shift

Posted Date: October 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11136934/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Cross-Cohort Transportability of Multi-Family Resting-State EEG Microstate Dynamics in Parkinson's Disease

Yuxiang Wang[1], Guimei Yin[1*], Guangxing Guo[1], Ming Liu[1], Zihao Wei[1], Yukun Wang[1], Anqi Li[1], Yanli Zhao[2], Shuping Tan[2], Hua Wei[3], Bin Wang[4*]

[1*]College of Computer Science and Technology, Taiyuan Normal University, Jinzhong, 030619, Shanxi, China .
[2]Psychiatric Research Center, Capital Medical University Affiliated Beijing Huilongguan Hospital, Beijing, 100096, China .
[3]Department of Psychology, Suzhou University of Science and Technology, Suzhou, 215009, Jiangsu, China .
[4]College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, 030024, Shanxi, China .

*Corresponding author(s). E-mail(s): yinguimeicn@126.com; wangbin01@tyut.edu.cn.

Abstract

Resting-state electroencephalography (EEG) microstates provide a millisecond-scale representation of large-scale brain-state dynamics, but the cross-cohort stability of Parkinson's disease (PD)-related microstate signatures remains uncertain. We tested whether a multi-family representation of canonical A/B/C/D microstate sequences—integrating static descriptors, directed transitions, flow-related measures, and sequence complexity—retains disease-related information across increasingly stringent validation levels. The development analysis used repeated stratified cross-validation, followed by leakage-aware fold-wise template estimation and locked external testing on an independent OpenNeuro cohort. Random forest achieved a development-cohort ROC-AUC of 0.823, whereas performance fell to near chance during external testing (ROC-AUC = 0.513). In the same feature space, a dataset-origin classifier achieved ROC-AUC = 0.966, demonstrating substantial cross-cohort distribution shift. These findings show that resting-state microstate dynamics contain PD-related information within

a cohort, but that the present representation is not sufficiently stable across cohorts. Cross-cohort validation and explicit assessment of dataset shift are therefore essential when evaluating EEG microstate features as reproducible electrophysiological markers.

Keywords: Parkinson's disease, electroencephalography, EEG microstates, brain-state dynamics, cross-cohort transportability, distribution shift

1 Introduction

Parkinson's disease (PD) is a heterogeneous neurodegenerative disorder with electrophysiological alterations that extend beyond the motor manifestations used in routine clinical recognition. Resting-state EEG has been increasingly used to characterize PD-related neural alterations through spectral, connectivity, complexity, and machine-learning analyses. However, successful discrimination within a single dataset provides limited evidence that the identified EEG pattern will remain stable when acquisition site, channel montage, medication state, preprocessing procedure, or cohort composition changes (Yassine et al. 2023; Fiorenzato et al. 2024; Betrouni et al. 2026; Handojoseno et al. 2015; Meng et al. 2025; Di Biase et al. 2023).

EEG microstate analysis approaches resting-state activity from a temporal-topographic perspective by representing continuous EEG as a succession of short-lived, recurrent scalp configurations. Conventional descriptors quantify the prevalence and persistence of these configurations, whereas transition syntax and other sequence-level measures characterize how the system moves between recurrent states over time (Asha et al. 2024; Michel and Koenig 2018; Michel and Bréchet 2026; Haydock et al. 2025; Zhao et al. 2026; Tarailis et al. 2026). Simultaneous EEG-fMRI studies support a broader interpretation of microstates as rapidly changing large-scale network configurations, while also cautioning against assigning any individual A/B/C/D class to a single canonical fMRI network in a one-to-one manner (Britz et al. 2010; Abreu et al. 2021).

For PD, the central translational question is not simply whether microstate features separate patients from controls within one dataset, but whether the disease-related structure persists when the cohort and recording context change. This distinction is especially important for machine-learning studies, because internal cross-validation can remain optimistic when cohort-specific feature structure, preprocessing choices, or template estimation influence model development. A biomarker-oriented analysis therefore requires both within-cohort evidence and an explicit test of transportability.

The present study tested the cross-cohort stability of multi-family resting-state EEG microstate dynamics in PD. We integrated conventional static descriptors with directed state transitions, inflow/outflow and flow-balance measures, and sequence-complexity features. The analytical design was organized around three linked methodological contributions: (1) a multi-family representation that captures both state occupancy and temporal organization; (2) a validation hierarchy separating

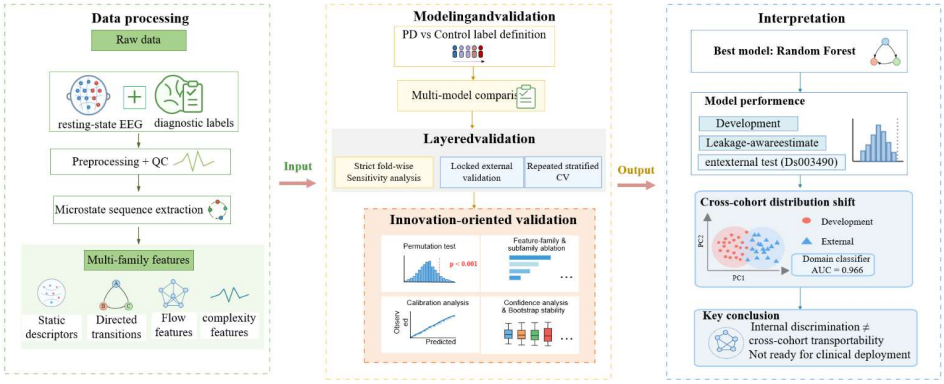


Fig. 1 Analytical framework for multi-family EEG microstate marker evaluation. The study separates exploratory within-cohort discrimination, leakage-aware internal validation, and locked external testing, followed by cross-cohort distribution-shift analysis.

exploratory repeated cross-validation from leakage-aware fold-wise template estimation and locked external testing; and (3) explicit quantification of dataset shift in the same feature space used for PD-control discrimination. This design allowed attenuation of external performance to be interpreted in relation to measurable cohort structure rather than model accuracy alone.

The overall workflow is shown in Fig. 1. Resting-state EEG recordings were quality controlled and standardized to a common duration, canonical A/B/C/D microstate sequences were derived, and multi-family temporal features were constructed. The primary analysis asked how much PD-related information was present within the development cohort and how much of that information survived progressively stricter validation. Supporting analyses were used to characterize feature contribution and interpretation without replacing the central transportability test.

1.1 Study Objective and Analytical Framework

The primary analytical objective was to determine whether PD-related information in resting-state EEG microstate dynamics was stable across validation levels and cohorts. Each preprocessed recording was converted into a symbolic sequence of canonical A/B/C/D labels and summarized by static, transition, flow, and complexity features. Binary PD-control classification was used as a quantitative probe of signal strength and transportability rather than as evidence of a clinically deployable diagnostic system.

The diagnostic group label was coded as 0 for healthy controls and 1 for PD participants. Age, sex, MoCA, and UPDRS-III were excluded from the primary predictive feature matrix and retained for cohort description and sensitivity analyses. The principal analytical sequence was predefined as development-cohort discrimination, leakage-aware internal evaluation, locked external testing, and cross-cohort distribution-shift characterization. Calibration, demographic and cognitive-status sensitivity, feature attribution, and effect-size analyses were treated as supporting analyses for interpretation rather than as independent claims of biomarker validity.

2 Materials and Methods

2.1 Dataset and Clinical Variables

The study was based on the open-access Rest Eyes Open EEG dataset available from OpenNeuro (ds004584, version 1.0.0) (Singh et al. 2023). This development dataset included 149 participants, consisting of 100 patients with Parkinson's disease and 49 healthy controls. EEG signals were recorded using a BrainVision system, and 63 scalp EEG channels were retained after channel selection and preprocessing. An independent external test was additionally performed using OpenNeuro ds003490 (version 1.1.0), which includes resting-state EEG recordings from 25 PD patients and 25 healthy controls (Cavanagh 2021). For this external analysis, one resting recording per participant was used; PD recordings were selected in the OFF-medication state when available, and controls were taken from session 1.

The available clinical information included age, sex, diagnostic group, MoCA score, and UPDRS-III score. In the present study, the primary target was diagnostic group classification, namely distinguishing PD patients from healthy controls. Therefore, the diagnostic group label was used as the prediction outcome, with controls coded as 0 and PD patients coded as 1. Age and sex were not used as predictive features in the main EEG-based model, but were retained for demographic sensitivity analysis to examine whether model performance was influenced by population characteristics.

MoCA and UPDRS-III scores were used only to describe the clinical profile of the dataset. They were not included as input features in the EEG microstate classifier, because the purpose of this study was to evaluate whether resting-state EEG microstate dynamics alone could provide useful information for PD diagnostic support.

2.2 Preprocessing With Quality Control

All recordings were processed in Python using MNE-Python and the study's scripted analysis workflow. Because transition, flow, and sequence-complexity measures can be sensitive to noise, artifacts, abnormal channels, and unequal sequence length, preprocessing was treated as an integral quality-control component of the analytical framework.

Continuous EEG was band-pass filtered at 1–45 Hz, supplemented by a 60-Hz notch filter for line-noise attenuation, and resampled to 250 Hz. Channel quality was assessed from across-time variance relative to each participant's channel distribution using a robust median-absolute-deviation criterion. Channels with absolute MAD-normalized variance scores greater than 3.5 were identified as atypical and reconstructed by interpolation. ICA artifact correction was performed with the FastICA implementation available through MNE-Python. Components were reviewed manually using their time courses, spectra, and scalp topographies; components with clear ocular or muscle-related patterns were removed before reconstruction of the continuous signal.

Following artifact correction, recordings were transformed to a common-average reference. To ensure that transition and sequence-derived features were calculated from equal-length signals, the primary analysis window was standardized to 120 s for all retained participants. Recordings were excluded when more than 15% of EEG

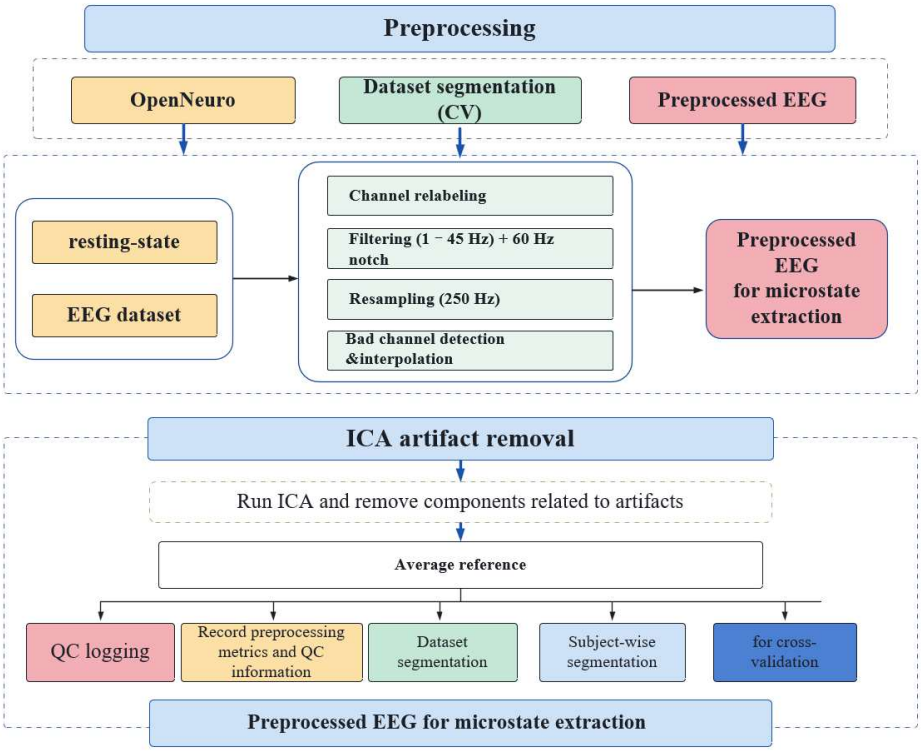


Fig. 2 Preprocessing and participant-level quality-control workflow applied before microstate extraction and feature construction.

channels were classified as bad, less than 90 s of usable data remained, or artifact correction required removal of more than 10 ICA components. These participant-level criteria were applied before microstate feature extraction.

Channel-wise temporal variance was used as the initial signal-quality statistic:

$$Var_i = \frac{1}{T-1}\sum_{t=1}^{T}(x_i(t) - \bar{x}_i)^2 \tag{1}$$

where $x_i(t)$ is the voltage recorded from channel i at sample t, T is the number of temporal samples, and $\bar{x}_i$ is the corresponding temporal mean. Variance values at either extreme of the participant-specific channel distribution were treated as potential indicators of abnormal channel behavior.

To reduce the influence of outlying channels on the reference distribution itself, channel variance was standardized relative to the median and median absolute deviation (MAD):

$$z_i^{\text{robust}} = \frac{Var_i - \text{median}(Var)}{\text{MAD}(Var)} \tag{2}$$

The resulting statistic expresses the deviation of each channel variance from the participant-level median on a MAD-normalized scale. Channels with $|z_i^{\text{robust}}| > 3.5$ were marked for interpolation. The proportion of such channels was then summarized as the participant-level bad-channel ratio:

$$R_{\text{bad}} = \frac{N_{\text{bad}}}{N_{\text{EEG}}} \times 100\% \tag{3}$$

where N_{bad} is the number of channels flagged and reconstructed by interpolation and N_{EEG} is the number of EEG channels included in the analysis. This ratio served as a participant-level quality-control measure, with recordings exceeding the prespecified 15% threshold excluded from downstream analysis.

2.3 Microstate Extraction

Microstate estimation was performed on the preprocessed continuous EEG. Global field power (GFP) was first computed at each sample to quantify the spatial dispersion of the scalp potential field:

$$GFP(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i(t) - \bar{x}(t)\right)^2} \tag{4}$$

where N is the number of EEG channels, $x_i(t)$ is the amplitude of channel i at sample t, and $\bar{x}(t)$ is the instantaneous mean potential across channels. Topographies at local GFP maxima were used for template estimation because these samples provide relatively high spatial signal strength. Local GFP maxima were detected within each 120-s, 250-Hz analysis segment, retaining all eligible peaks rather than imposing a fixed per-participant subsample; the resulting participant-specific peak count was therefore determined by the observed GFP time course after preprocessing. Microstate templates were estimated from these peak maps using polarity-invariant modified k-means clustering with $K = 4$, 100 random initializations, a maximum of 1000 iterations per initialization, and a convergence tolerance of 10^{-6} in explained-variance change. The four-cluster solution was selected a priori to match the conventional A/B/C/D framework used in adult resting-state EEG microstate studies and to maintain interpretability across the development, fold-wise, and external validation analyses (Michel and Bréchet 2026; Lehmann et al. 1987; Koenig et al. 2002, 2005; Liu et al. 2022).

Cluster maps were aligned to the canonical A/B/C/D convention by maximizing absolute spatial correlations between estimated maps and the reference A/B/C/D topographies. When several maps were aligned simultaneously, the one-to-one assignment was chosen by the Hungarian algorithm to maximize the summed absolute spatial correlation across the four classes. The same spatial-matching rule was applied to cohort-level, training-fold, and external-template backfitting steps. Analyses were implemented in Python with MNE-Python, NumPy, SciPy, scikit-learn, and the study's scripted microstate workflow.

Because polarity inversion was not considered to define a distinct microstate, template matching was based on absolute spatial correlation. Each sample was assigned to the template with the highest absolute topographic correlation:

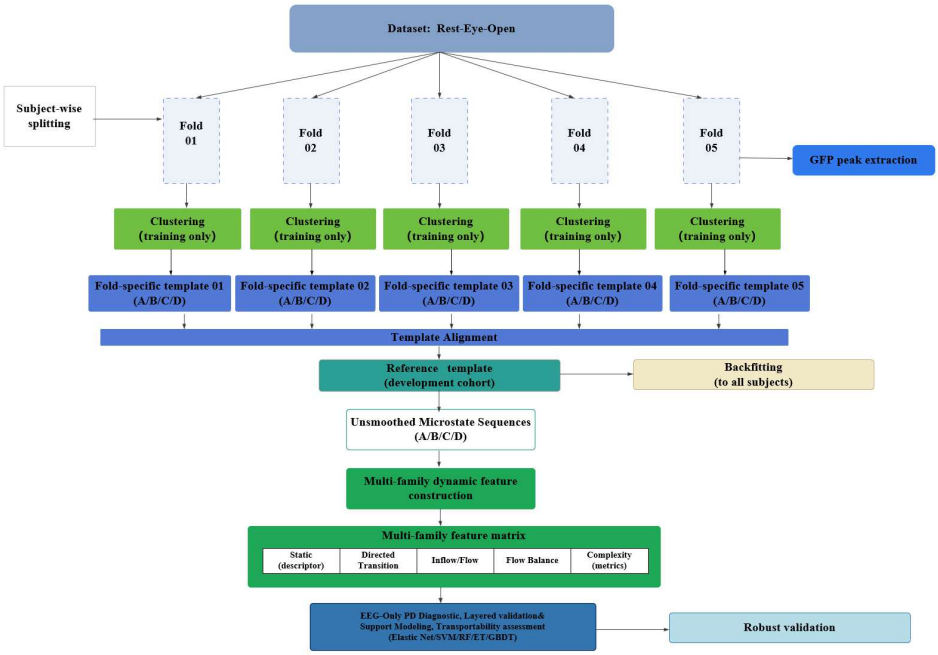


Fig. 3 Fold-aware microstate template estimation, backfitting, and multi-family feature construction. The feature representation combines static, directed-transition, flow, and complexity information before marker evaluation.

$$s(t) = \arg\max_{k} |\mathrm{corr}(X(t), M_k)| \tag{5}$$

where $s(t)$ is the assigned microstate class at sample t, $X(t)$ is the observed scalp map, and M_k is the k-th reference template.

Template estimation was explicitly incorporated into the validation hierarchy to assess the optimism that can arise when microstate maps are informed by the full development cohort. The primary repeated-CV analysis used cohort-level templates and was interpreted as an exploratory development result. In the stricter internal analysis, template construction was repeated independently within each training fold and the resulting maps were then backfitted to the corresponding held-out participants. For external validation, neither the development templates nor the fitted prediction models were re-estimated using ds003490.

The primary analyses used the unsmoothed backfitted label sequence. This choice was made because temporal smoothing can suppress short-lived state changes that contribute directly to transition probability, switching rate, flow, and entropy-related measures. As a robustness analysis, the complete feature set was recomputed after 30-ms majority smoothing and compared with the unsmoothed representation (Haydock et al. 2025; Liang et al. 2024).

2.4 Multi-family Microstate Feature Construction

The microstate sequence was transformed into a feature representation spanning four complementary levels of temporal organization. State-level behavior was characterized using duration, occurrence, coverage, and global explained variance. Pairwise temporal organization was represented by directed transition probabilities. Directional organization was summarized through state-specific inflow, outflow, and inflow–outflow balance, while whole-sequence irregularity and flexibility were captured by complexity measures. The resulting multi-family temporal-flow representation was designed to capture information beyond conventional static microstate summaries (Li and Zhu 2025; Pei et al. 2025; Tang et al. 2026).

For each participant, the unsmoothed microstate label sequence was used to build a directed transition probability matrix. Each off-diagonal entry described how often the EEG sequence moved from one microstate to another. Diagonal entries were not treated as transitions because the analysis focused on switching between different states. In addition to pairwise transitions, inflow, outflow, and balance were calculated to describe whether a state tended to function as a receiving state, a leaving state, or a relatively balanced node in the microstate transition network.

The off-diagonal transition probability from microstate i to microstate j was calculated as:

$$P_{ij} = \frac{\sum_{t=1}^{T-1} I[s(t) = i,\, s(t+1) = j]}{\sum_{t=1}^{T-1} I[s(t) = i]}, \qquad i \neq j \tag{6}$$

where P_{ij} is the conditional probability of leaving state i for state j. Transition counts were normalized by the number of samples originating from state i, and self-transitions ($i = j$) were excluded because the analysis focused on changes between distinct microstate classes.

2.5 EEG Microstate Marker Evaluation Models

PD-control classification was used to quantify the disease-related information carried by the EEG microstate feature space and to test its stability across validation levels. The diagnostic label was treated as a binary analytical outcome, with healthy controls coded as 0 and PD patients coded as 1. Age, sex, MoCA, and UPDRS-III were not used as predictive features in the primary model and were retained only for sample description and sensitivity analyses. Accordingly, classifier performance was interpreted as a marker-evaluation metric rather than as clinical diagnostic accuracy.

The binary diagnostic label was defined as:

$$y_i = \begin{cases} 0, & \text{healthy control}, \\ 1, & \text{PD patient}, \end{cases} \tag{7}$$

where y_i denotes the diagnostic label of participant i. A value of 0 indicates a healthy control, whereas a value of 1 indicates a patient with PD.

For each participant i, the multi-family microstate feature vector was denoted as $\mathbf{x}_i$. Multiple classifiers were compared, including Elastic Net logistic regression, calibrated support vector machine, random forest, extra trees, and gradient boosting.

Each model estimated the probability that a participant belonged to the PD group: (Breiman 2001)

$$\hat{p}_i = P(y_i = 1 \mid \mathbf{x}_i) \tag{8}$$

where $\hat{p}_i$ denotes the predicted probability of PD for participant i, $y_i = 1$ indicates the PD class, and $\mathbf{x}_i$ represents the EEG microstate feature vector of participant i.

2.6 Validation Hierarchy, Sensitivity Analyses, and Model Interpretation

Model performance was reported in a predefined hierarchy. The subject-level repeated stratified cross-validation result was treated as the exploratory within-cohort model-comparison estimate, strict fold-wise sensitivity analysis was treated as the conservative leakage-aware internal check, and locked testing on ds003490 was treated as the independent external stress test. The exploratory development analysis used five-fold stratified cross-validation repeated five times with a fixed random seed of 42. All feature standardization, feature selection when applicable, calibration, and model fitting were performed within the training portion of each split and then applied to the corresponding held-out fold. Hyperparameters were fixed from the predefined analysis configuration or selected inside the training fold only; held-out and external participants were never used to tune model settings. ROC-AUC and balanced accuracy were the primary discrimination metrics, with sensitivity and specificity reported to characterize class-specific performance. Importantly, the three validation levels were not interpreted as interchangeable estimates: the progressive attenuation from exploratory internal evaluation to strict fold-wise evaluation and then to external testing was treated as evidence about model transportability.

Balanced accuracy was calculated as the mean of sensitivity and specificity:

$$BA = \frac{\text{Sensitivity} + \text{Specificity}}{2} \tag{9}$$

where BA denotes balanced accuracy. Sensitivity reflects the proportion of PD patients correctly identified by the model, whereas specificity reflects the proportion of healthy controls correctly identified by the model.

Permutation testing was used to examine whether the observed AUC could be explained by random label structure. The empirical p-value was calculated by comparing the observed AUC with the AUC values obtained after label shuffling: (Phipson and Smyth 2010)

$$p_{\text{perm}} = \frac{1 + \sum_{r=1}^{R} I(AUC_r \geq AUC_{\text{obs}})}{R + 1} \tag{10}$$

where AUC_{obs} is the ROC-AUC obtained from the original diagnostic labels, AUC_r is the corresponding value from permutation r, and R is the total number of label permutations ($R = 499$ in the present analysis). The indicator contributes one whenever a permuted AUC is at least as large as the observed AUC. Thus, p_{perm} estimates the proportion of permutation outcomes providing performance equal to or better than that obtained with the true labels.

Feature-family ablation was used to compare the contribution of static, transition, flow, complexity, dynamic, and combined feature sets. This analysis was used to evaluate whether diagnostic information was concentrated in one feature family or distributed across several aspects of microstate dynamics.

The Brier score measured the mean squared difference between predicted probability and true diagnostic label:

$$\text{Brier} = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2 \tag{11}$$

where n is the number of participants, p_i is the predicted PD probability, and y_i is the observed binary label. Lower values indicate better agreement between predicted probabilities and observed outcomes.

Expected calibration error was further calculated to summarize the difference between predicted confidence and observed accuracy across probability bins:

$$ECE = \sum_{m=1}^{M}\frac{|B_m|}{n}\left|\text{acc}(B_m) - \text{conf}(B_m)\right| \tag{12}$$

where B_m denotes the m-th probability bin, $\text{acc}(B_m)$ is the observed accuracy within that bin, and $\text{conf}(B_m)$ is its mean predicted confidence. ECE was calculated with $M = 10$ equal-width probability bins. Lower ECE values indicate closer agreement between empirical outcomes and predicted confidence across bins.

Decision confidence was defined as:

$$c_i = \max\left(\hat{p}_i, 1 - \hat{p}_i\right) \tag{13}$$

where c_i summarizes the distance of participant i's predicted probability from the 0.5 decision boundary; larger values therefore indicate more confident model assignments.

Bootstrap-based feature stability was calculated as:

$$S_j = \frac{1}{B}\sum_{b=1}^{B} I(j \in F_b) \tag{14}$$

where F_b is the set of selected or top-ranked variables in bootstrap iteration b ($B = 500$ bootstrap iterations). The stability score S_j is therefore the proportion of bootstrap samples in which feature j reappears among the selected variables, with larger values indicating greater ranking reproducibility.

To characterize cross-cohort distribution shift after locked external testing, participants from the development cohort and ds003490 were compared in a common multi-family microstate-feature space. Principal component analysis was used to visualize the dominant axes of between-cohort variation. In parallel, a cross-validated dataset-origin classifier was trained to predict development-versus-external cohort membership from the same feature representation. Domain-classification performance was interpreted as a measure of cohort separability rather than diagnostic accuracy; a higher ROC-AUC therefore indicated stronger dataset-specific structure in the microstate-feature space.

Given the lower MoCA scores observed in the PD group, cognitive-status sensitivity was evaluated in two complementary ways. First, Spearman correlations were examined within the PD group between MoCA and leading model-prioritized microstate features. Second, the diagnostic analysis was repeated after restricting the PD sample using MoCA thresholds of ≥ 21 and ≥ 23, and the resulting repeated-cross-validation scores were compared with the all-subject sensitivity configuration. To provide directionality for model-attribution findings, leading features were also summarized using Cohen's d (PD minus HC) and standardized group means. These analyses were used as sensitivity and descriptive effect-size assessments rather than as evidence of causal biomarkers.

Together, these analyses framed binary classification as a tool for evaluating the robustness and interpretability of PD-related EEG microstate markers rather than as an endpoint of clinical diagnosis.

2.7 Statistical Analysis and Reporting

Demographic and clinical variables were summarized using mean ± standard deviation (SD), counts, or median [25th–75th percentile], as appropriate. Age and MoCA were compared between groups using two-sided independent-samples t tests, and sex distribution was compared using Fisher's exact test. UPDRS-III was summarized descriptively because it was available only for participants with PD. Preprocessing quality-control variables were summarized using the mean ± SD together with median values and prespecified quality-control thresholds. The common 120-s analysis duration was fixed by design and was treated as an analysis setting rather than an inferential outcome.

For predictive analyses, ROC-AUC and balanced accuracy were the primary discrimination metrics, with sensitivity and specificity used to describe class-specific performance. Permutation testing was used to assess whether development-cohort AUC exceeded the shuffled-label reference distribution. Model interpretation, subgroup analyses, cognitive-status sensitivity analyses, external testing, and dataset-origin classification were reported in the same order as specified in the analytical framework above. Exact p values are reported where available, rather than replacing non-significant values with qualitative labels. Group contrasts, classifier performance, feature attribution, and cross-cohort comparisons were interpreted as associational or descriptive evidence and were not used to make causal claims.

3 Results

3.1 Participants and Clinical Scores

The development cohort comprised 149 participants, including 49 healthy controls and 100 participants with PD. Age did not show clear evidence of a between-group difference ($p = 0.081$), and sex distribution was also comparable by Fisher's exact test ($p = 0.104$). MoCA scores were lower in the PD group than in controls (24.31 ± 4.02 vs. 26.67 ± 1.86, $p < 0.001$). UPDRS-III was available only for the PD group, with

a median of 11.0 [7.0–17.5]. Age, sex, MoCA, and UPDRS-III were not entered as predictors in the primary EEG-based microstate classifier.

Table 1 Demographic and clinical characteristics of the development cohort

Metric	Whole cohort	Controls	PD	p value
Age (years)	69.32 ± 7.97	70.92 ± 7.62	68.53 ± 8.06	0.081
Sex (M/F)	94/55	26/23	68/32	0.104
MoCA score	25.09 ± 3.63	26.67 ± 1.86	24.31 ± 4.02	< 0.001
UPDRS-III score	–	–	11.0 [7.0–17.5]	–

Note. Age and MoCA are reported as mean ± SD and were compared using two-sided independent-samples t tests. Sex was compared using Fisher's exact test. UPDRS-III is reported as median [25th–75th percentile] and was available only for participants with PD; therefore, no between-group test was performed.

Table 1 summarizes the cohort characteristics and the prespecified group comparisons used for sample description. The demographic and clinical variables were not included in the primary predictive feature matrix.

3.2 Preprocessing Quality

Preprocessing quality was summarized at the level of the full main-analysis cohort. The bad-channel ratio was $2.49 \pm 1.65\%$ (median 3.17%, maximum 6.35%), remaining below the prespecified exclusion threshold of 15%. The number of ICA components removed was 1.93 ± 0.85 per participant (median 2), below the exclusion threshold of more than 10 components. Available recording duration before cropping was 160.45 ± 46.12 s, with a median of 145.72 s [125.76–177.43]; recordings were required to contain at least 90 s of usable data. All retained recordings were subsequently standardized to a common 120-s analysis window before microstate extraction.

Table 2 Preprocessing quality-control metrics for the main analysis

Metric	Main-analysis value
Bad-channel ratio (%)	2.49 ± 1.65; median 3.17
ICA components removed	1.93 ± 0.85; median 2
Available duration before cropping (s)	160.45 ± 46.12; median 145.72 [125.76–177.43]
Common analysis duration (s)	120

Note. Values are reported as mean ± SD with the median where applicable. Recordings were excluded when the bad-channel ratio exceeded 15%, more than 10 ICA components required removal, or less than 90 s of usable EEG remained. All retained recordings were standardized to a common 120-s analysis window before microstate extraction.

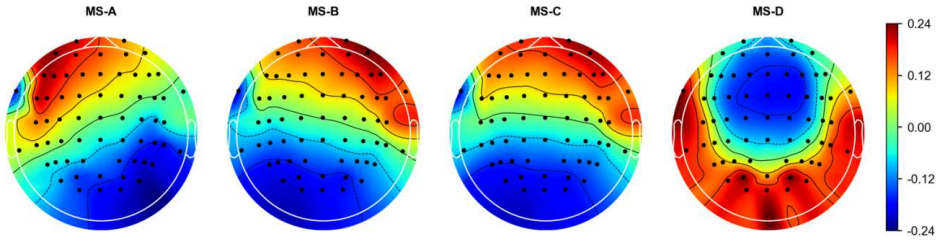


Fig. 4 Canonical A/B/C/D EEG microstate templates used to generate subject-level label sequences after template alignment and backfitting.

Table 2 summarizes the principal preprocessing quality-control quantities and the thresholds used to retain recordings for the main analysis. These criteria were applied before microstate extraction so that downstream transition, flow, and complexity measures were derived from quality-controlled, equal-length EEG segments.

3.3 Microstate Sequences and Multi-family Features

Four canonical microstate maps were extracted and aligned to the classical A/B/C/D convention. These templates were used as reference maps for backfitting continuous EEG signals into symbolic microstate label sequences. In the final prediction framework, the main route used unsmoothed sequences so that brief switching events could be retained for transition, flow, and complexity features. A 30-ms smoothing sensitivity analysis was added as a supplementary robustness check.

Based on these sequences, the feature matrix was organized into static, transition, flow, and complexity families. Static features described state dominance; directed transitions described switching between microstate classes; flow features summarized inflow, outflow, and balance; and complexity features captured sequence irregularity. These feature families, together with finer subfamilies, were used for diagnostic prediction, ablation, and interpretability analyses.

3.4 Within-Cohort Discrimination Performance

In the exploratory development analysis, random forest showed the highest discrimination among the evaluated classifiers, with repeated-CV ROC-AUC = 0.823 and balanced accuracy = 0.734. This result was treated as the development benchmark for the subsequent validation hierarchy rather than as an estimate of deployable diagnostic performance.

Permutation testing provided a complementary check that the development-cohort discrimination exceeded a shuffled-label reference distribution. Because the permutation analysis and other supporting modules used their own predefined resampling configurations, their displayed AUC values were interpreted within those analyses rather than as interchangeable estimates of the primary repeated-CV ROC-AUC.

The development cohort therefore contained detectable PD-related information in the multi-family microstate representation. Age, sex, MoCA, and UPDRS-III were

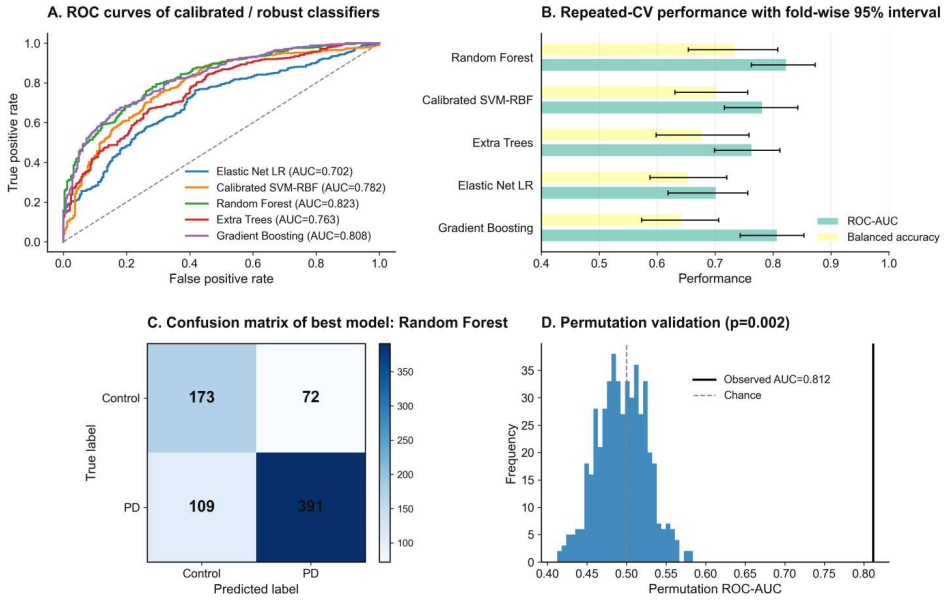


Fig. 5 Exploratory within-cohort discrimination using the multi-family microstate feature set. A, ROC curves across evaluated classifiers. B, repeated-CV ROC-AUC and balanced accuracy. C, repeated-CV confusion summary for the random forest model; counts represent accumulated out-of-fold predictions across the five repeated cross-validation runs rather than unique participants. D, permutation-based comparison with shuffled-label performance. These panels characterize development-cohort signal and are not evidence of external diagnostic validity.

excluded from the predictive feature matrix, so the observed discrimination was derived from EEG microstate features rather than direct demographic or clinical-score input (Gschwandtner et al. 2023; Elkholy et al. 2023).

The key question was whether this within-cohort signal survived stricter evaluation. The leakage-aware fold-wise analysis and locked external test were therefore treated as progressively stronger tests of transportability, while feature ablation, calibration, confidence, stability, and attribution were used only to characterize the development representation.

3.5 Feature-Family Contribution and Supporting Validation Analyses

Feature-family ablation was used to test whether the combined representation captured information that was not confined to one conventional feature family. Under the dedicated ablation configuration, the combined feature set achieved the highest ROC-AUC among the tested families, whereas static, transition, flow, and complexity components showed lower performance when evaluated separately.

The ablation AUCs were used for within-analysis comparison of feature families and should not be numerically equated with the primary development ROC-AUC of 0.823, because the analyses served different validation purposes. The result supports a distributed representation of PD-related information across complementary aspects

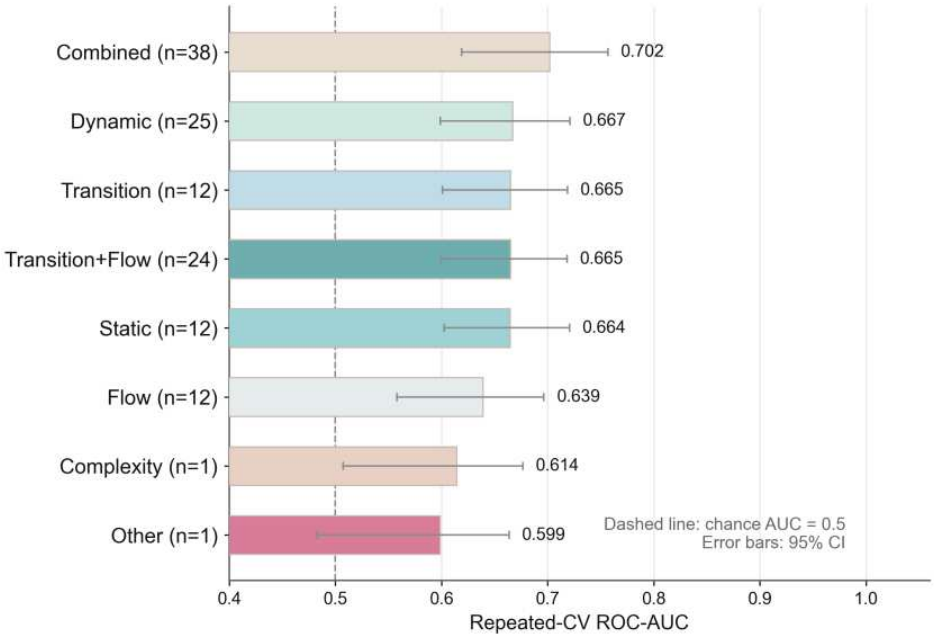


Fig. 6 Feature-family ablation under the dedicated family-comparison configuration. The combined representation retained the highest ROC-AUC among the tested feature families, supporting a distributed contribution across static and dynamic descriptors. Error bars indicate 95% confidence intervals; very small feature families should be interpreted descriptively.

of microstate organization rather than a claim that any individual feature family is a standalone biomarker (Costa et al. 2023).

Probability calibration, decision-confidence analysis, and bootstrap feature stability were treated as supporting analyses and are provided in Online Resource 1 (Figs. S2–S4), keeping the main text focused on the validation hierarchy, external testing, and cross-cohort shift.

3.6 Joint Feature Importance and Stability Analysis

To further interpret the fitted PD-control classifier, permutation feature importance was examined together with bootstrap-based feature stability. Permutation importance quantified the decrease in ROC-AUC after perturbing each feature, whereas bootstrap stability reflected how frequently a feature was repeatedly selected as a top-ranked variable across resampling. Therefore, the two analyses provided complementary information: permutation importance described model contribution, while bootstrap stability described the robustness of feature selection.

As shown in Fig. 7, several microstate features showed both relatively high permutation importance and high bootstrap selection frequency. These features spanned directed transitions, static descriptors, GEV-related measures, flow-related variables, and switching-rate information. Tree SHAP attribution additionally prioritized A-to-D transition, C-to-B transition, D outflow, D-to-C transition, B-to-C transition, C

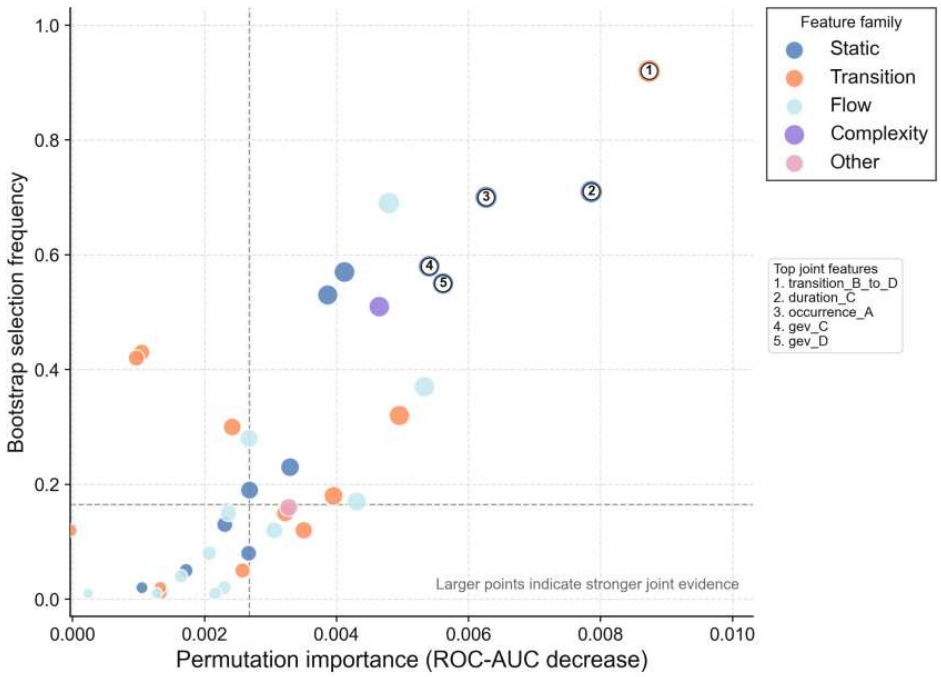


Fig. 7 Joint model-level feature importance and resampling stability. The x-axis shows the decrease in ROC-AUC after feature perturbation, and the y-axis shows bootstrap top-feature selection frequency. Features toward the upper-right contributed more strongly and were selected more consistently, but these quantities do not establish biological causality.

duration, A outflow, and D duration among the leading contributors. This distributed pattern is consistent with the proposed multi-family temporal-flow representation.

This joint pattern is consistent with the multi-family feature design used in the present study. The fitted model did not rely only on a fixed set of conventional microstate descriptors. Instead, PD-related diagnostic information appeared to involve a combination of static state properties, directed transition organization, flow directionality, GEV-related template expression, and temporal complexity. Features supported by more than one interpretability analysis were therefore considered more reliable model-level contributors than features ranked highly by only one method.

At the same time, these results should be interpreted cautiously. Permutation importance and bootstrap stability describe how the trained model used the available EEG microstate features, but they do not establish direct biological causality. Thus, the highlighted features should be regarded as interpretable contributors to the current EEG-based microstate classifier rather than as confirmed neurophysiological biomarkers.

3.7 Demographic Sensitivity and Quality-Control Checks

Demographic sensitivity analysis was performed to examine whether prediction performance was concentrated in a single age or sex subgroup. As shown in Fig. 8, subgroup

ROC curves were calculated for female participants, male participants, participants below the median age, and participants above the median age. The subgroup AUC values were broadly comparable, ranging from 0.815 to 0.841.

The female subgroup showed an AUC of 0.841, while the male subgroup showed an AUC of 0.817. For age-stratified analysis, the age-above-median subgroup showed an AUC of 0.829, and the age-below-median subgroup showed an AUC of 0.815. These relatively close AUC values suggest that the model's discrimination ability was not obviously restricted to one demographic subgroup.

This analysis is important because age and sex were not used as predictive features in the main EEG-based model. The subgroup results therefore provide an additional check that the prediction signal was not simply driven by direct demographic input. However, these subgroup analyses were still based on repeated cross-validation predictions within the development dataset. The independent ds003490 test provided a stronger generalization check and showed that demographic robustness inside one cohort did not guarantee cross-cohort transferability.

In addition to demographic checks, preprocessing summaries showed high channel matching, similar bad-channel ratios, and broadly comparable artifact-correction profiles between controls and PD participants. These results did not reveal an obvious preprocessing-quality imbalance that paralleled the diagnostic prediction, although residual acquisition or preprocessing effects cannot be excluded.

3.8 External Validation on an Independent OpenNeuro Cohort

The locked development representation was applied to OpenNeuro ds003490, an independent resting-state EEG cohort with 25 PD patients and 25 healthy controls. Neither the development templates nor fitted prediction models were re-estimated using external diagnostic labels. For channel compatibility, 62 shared scalp channels were retained in development-template order, and the external recordings were processed using the same filtering, referencing, resampling, and 120-s cropping strategy.

External discrimination was near chance. Random forest achieved ROC-AUC = 0.513 and balanced accuracy = 0.520, while logistic regression achieved ROC-AUC = 0.411 and balanced accuracy = 0.480. The random-forest probability distributions showed substantial overlap between external PD and control participants. Thus, the central finding was not the development AUC in isolation, but the attenuation of discrimination as validation became more stringent.

The external result was treated as a transportability finding rather than as a model-selection failure. We therefore examined whether the loss of PD-control discrimination coincided with systematic differences between the development and external feature distributions.

3.9 Cross-Cohort Distribution-Shift Analysis

We next tested whether the attenuation in external PD-control performance was accompanied by measurable cohort structure in the same feature space. Principal component analysis showed systematic displacement between the development and

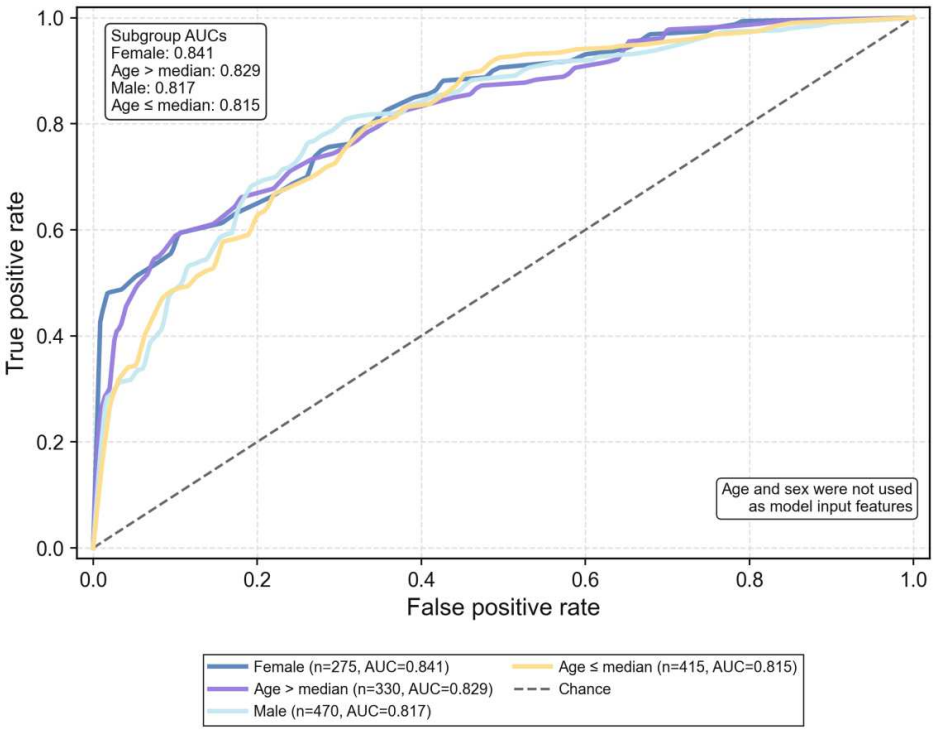


Fig. 8 Supporting demographic sensitivity analysis. ROC curves compare development-cohort performance across sex and age strata; similar subgroup AUCs indicate that the internal discrimination was not obviously concentrated in one tested demographic subgroup.

ds003490 cohorts; the first two components explained 42.7% and 18.9% of the variance, respectively.

A cross-validated dataset-origin classifier distinguished development from external participants with ROC-AUC = 0.966, and the corresponding probabilities showed marked cohort separation. High dataset-origin performance is not a desirable diagnostic result; instead, it demonstrates that cohort membership was strongly encoded in the microstate representation. This finding provides a direct link between limited external generalization and non-invariance of the measured feature space.

The present design cannot isolate a single source of this shift because acquisition protocol, channel configuration, medication state, cohort composition, recording characteristics, and preprocessing compatibility differed jointly between datasets. The analysis therefore supports distribution shift as a plausible contributor to performance attenuation without assigning a causal source.

3.10 Cognitive-Status Sensitivity Analysis

Because MoCA scores were lower in the PD group, cognitive status was examined as a potential source of heterogeneity rather than as a predictor. Within the PD

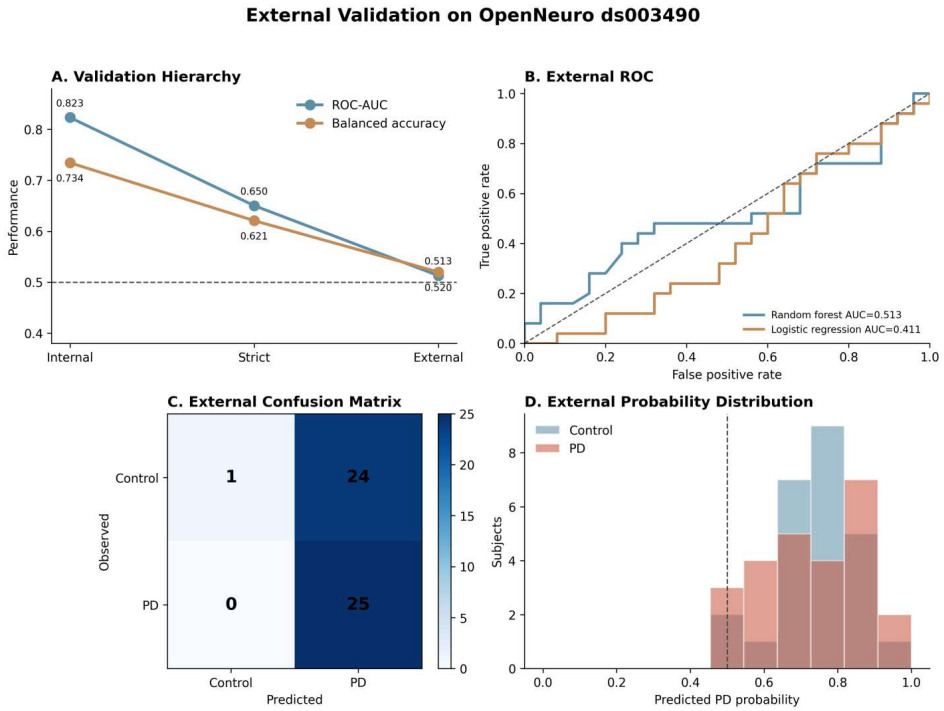


Fig. 9 Locked external validation on OpenNeuro ds003490. A, validation hierarchy from exploratory repeated CV to leakage-aware fold-wise evaluation and independent testing. B–D, external ROC curves, random-forest confusion matrix, and predicted-probability distributions. The near-chance external performance provides the principal evidence that the development signal was not robustly transportable.

group, associations between MoCA and leading model-prioritized features were weak (approximately −0.17 to 0.15), and the displayed uncertainty intervals crossed zero.

Classification was also repeated under a dedicated cognitive-restriction configuration. The all-subject sensitivity analysis yielded ROC-AUC = 0.478 and balanced accuracy = 0.456; restricting PD participants to MoCA ≥21 or ≥23 did not restore discrimination. These values arise from the sensitivity configuration shown in Fig. 11 and should not be directly compared as alternative estimates of the primary development ROC-AUC of 0.823. The analysis therefore indicates weak direct MoCA-feature associations while leaving residual uncertainty about the contribution of cognitive heterogeneity.

3.11 Direction and Effect Size of Discriminative Microstate Features

To complement model-attribution results with group-level directionality, we compared leading microstate features between PD and control participants using standardized effect sizes (Cohen's d, PD minus HC) and standardized group means. The effects were directionally heterogeneous rather than concentrated in a single feature family.

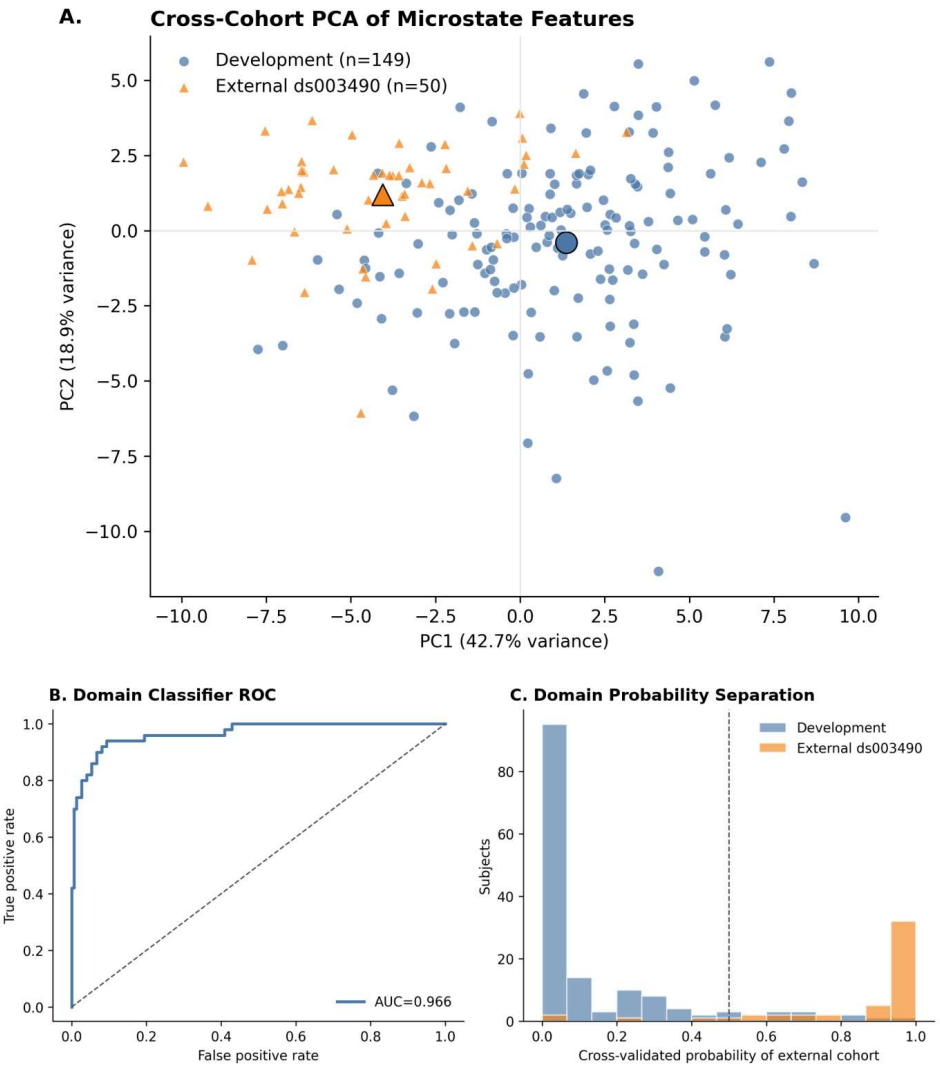


Fig. 10 Cross-cohort distribution shift in the multi-family microstate feature space. A, principal component analysis of development and external participants. B, cross-validated dataset-origin ROC curve. C, dataset-origin probability distributions. The dataset-origin ROC-AUC of 0.966 indicates strong cohort separability and provides a feature-space explanation for limited external transportability.

Duration_C and duration_D were higher in PD (d approximately 0.37 and 0.32, respectively), whereas GEV_C showed essentially no group difference (d approximately 0.00) and GEV_D showed only a small negative difference (d approximately -0.06).

Several transition-, flow-, and switching-related measures were lower in PD. The largest negative effects were observed for the B-to-C transition (d approximately -0.46), switching_rate (d approximately -0.42), the D-to-C transition (d approximately

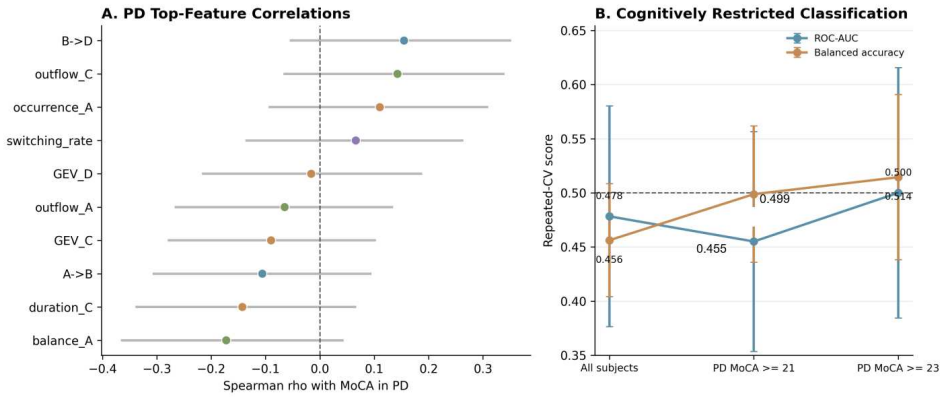


Fig. 11 Cognitive-status sensitivity analysis. A, within-PD Spearman associations between MoCA and leading model-prioritized features. B, performance under the dedicated cognitive-restriction configuration for all participants and after PD MoCA thresholds of $\geq$21 and $\geq$23. These sensitivity estimates are interpreted within this analysis and are not direct re-estimates of the primary development ROC-AUC.

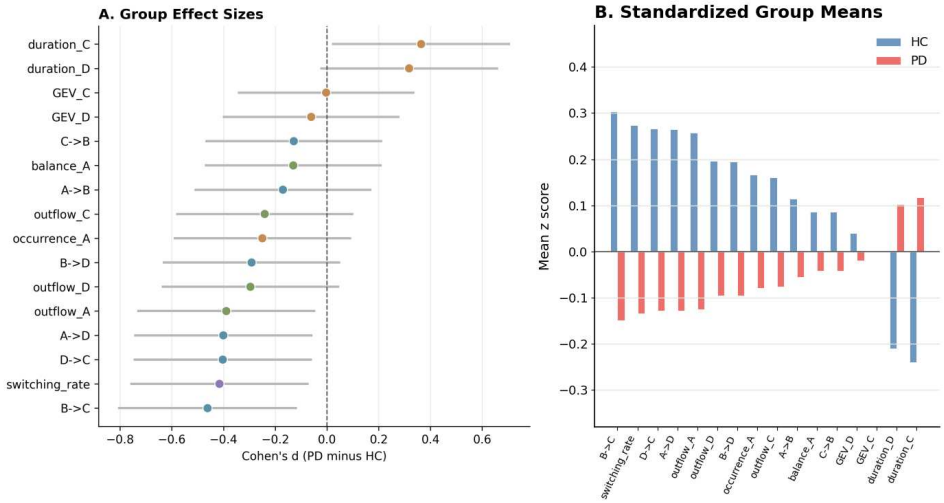


Fig. 12 Descriptive direction and magnitude of leading PD-control microstate differences. A, Cohen's d values computed as PD minus HC. B, standardized group means for the same features. These summaries provide directionality for model-attribution results and are not presented as independently validated biomarkers.

-0.40), the A-to-D transition (d approximately -0.40), and outflow_A (d approximately -0.39). Smaller negative effects were observed for outflow_D, the B-to-D transition, occurrence_A, outflow_C, the A-to-B transition, balance_A, and the C-to-B transition. Standardized group means showed the same overall directional pattern. Because uncertainty intervals were broad for most features, these results are interpreted as descriptive group-level context for model attribution rather than as independently validated neurophysiological biomarkers.

4 Discussion

4.1 Principal Findings

The central result is the ordered change in evidence across the validation hierarchy. The multi-family microstate representation carried PD-related information within the development cohort (random-forest ROC-AUC = 0.823), but the signal weakened under leakage-aware fold-wise evaluation (ROC-AUC = 0.650) and fell to near chance during locked external testing (ROC-AUC = 0.513). The development AUC alone therefore overstates the robustness of the marker representation.

Three findings support the study's methodological logic. First, feature-family ablation indicated that the combined representation captured information distributed across static and dynamic aspects of microstate organization. Second, progressively stricter validation separated within-cohort discrimination from cross-cohort transportability. Third, the dataset-origin classifier achieved ROC-AUC = 0.966, showing that the same feature space was strongly structured by cohort membership. Taken together, the external attenuation and strong cohort separability argue for explicit distribution-shift analysis when evaluating EEG microstate markers.

4.2 Multi-Family Microstate Dynamics as Electrophysiological Markers

The present findings suggest that restricting microstate analysis to state-occupancy measures may overlook diagnostically relevant temporal organization. Duration, occurrence, coverage, and global explained variance describe how individual states are expressed, whereas transition, flow, switching, and complexity measures quantify relationships among successive states (Michel and Bréchet 2026; Haydock et al. 2025). In the development dataset, the combined representation outperformed individual feature families, supporting a distributed rather than single-descriptor account of PD-related microstate differences.

The group-level effect-size analysis provides a complementary physiological description independent of model ranking. Duration_C and duration_D were higher in PD, whereas several transition-, flow-, and switching-related measures were lower, with the largest negative effects involving the B-to-C transition, switching rate, D-to-C transition, A-to-D transition, and outflow_A. Taken together, this pattern is compatible with longer dwell times in some recurrent global configurations and reduced flexibility of transitions among configurations. It should not, however, be interpreted as direct evidence that a specific canonical network is impaired.

EEG-fMRI studies support the broader interpretation that microstates index rapidly changing large-scale network configurations (Britz et al. 2010; Abreu et al. 2021). At the same time, the functional assignment of individual A/B/C/D classes is not one-to-one across studies: for example, simultaneous EEG-fMRI work has linked canonical microstates to several resting-state networks without establishing a universal mapping of a particular class to the default-mode or executive-control network (Britz et al. 2010). The present C/D duration and switching findings are therefore interpreted at the level of recurrent whole-brain configuration dynamics, not as proof of

dysfunction in a uniquely specified fMRI network. This conservative interpretation is also consistent with current methodological reviews emphasizing standardization and caution in functional labeling of microstate classes (Michel and Koenig 2018; Michel and Bréchet 2026).

4.3 Validation Attenuation and Cross-Cohort Distribution Shift

The separation between within-cohort discrimination and cross-cohort transportability has methodological implications beyond the present classifier. Permutation testing established that the development result exceeded a shuffled-label null distribution, but it did not establish external validity. Likewise, calibration and confidence analyses characterize the fitted development model rather than the invariance of its feature representation. A biomarker-oriented claim therefore requires evidence that both the feature distribution and the disease-related contrast remain sufficiently stable across cohorts.

The domain-classification result shows that they did not. A dataset-origin ROC-AUC of 0.966 indicates that development and external participants were readily separable from the microstate features themselves. This makes it difficult to attribute the external AUC of 0.513 solely to an absence of PD-related neurophysiology. Instead, the result is consistent with non-invariance of the measured feature space across datasets. Potential contributors include acquisition protocol, channel montage and channel matching, medication state, cohort composition, recording duration and characteristics, preprocessing compatibility, and microstate-template estimation. Because these factors vary together between datasets, the present design cannot identify a single causal source of shift.

This distinction matters for interpretation. The study does not show that microstate dynamics are intrinsically unsuitable as PD markers; it shows that the current representation and pipeline did not preserve PD-control discrimination across these two cohorts. Resolving that question will require harmonized multi-site data, explicit site-aware analyses, prospective external validation, and experiments that can separate acquisition effects from biological heterogeneity. Domain adaptation or harmonization could improve transfer in future work, but applying such procedures here would answer a different question from the present one, which was whether the unadapted microstate representation itself was transportable.

4.4 Clinical and Cognitive Sensitivity

Age and sex were excluded from predictive inputs, and subgroup AUCs were broadly similar within the development dataset. These findings reduce concern that internal discrimination was trivially driven by direct demographic input, but they do not substitute for external validation. MoCA requires similar caution. PD participants had lower MoCA scores, and excluding MoCA from the feature matrix does not eliminate the possibility that cognitive status is indirectly reflected in EEG dynamics.

Within-PD correlations between MoCA and leading features were weak, while cognitively restricted classification remained near chance in the sensitivity configuration.

These findings do not support a strong direct monotonic MoCA-feature association, but they also do not establish independence from cognitive heterogeneity. The current results should therefore be treated as sensitivity evidence rather than proof that cognition is irrelevant to the observed PD-control contrast (Oikonomou et al. 2026; Aarsland et al. 2021).

4.5 Model Interpretation and Biological Caution

Permutation importance, bootstrap stability, and Tree SHAP provided complementary model-level views of the feature space. Leading contributors spanned directed transitions, flow measures, durations, GEV-related variables, and switching rate, reinforcing the distributed nature of the representation (Ponce-Bobadilla et al. 2024; Lundberg and Lee 2017). However, these tools answer how the fitted classifier used the available features; they do not identify causal disease mechanisms.

The same caution applies to descriptive effect sizes. Longer C/D duration and lower switching-related measures may reflect altered large-scale brain-state organization in PD, but the present cross-sectional analysis cannot determine whether those differences arise from PD pathology, medication state, cognitive heterogeneity, acquisition context, or interactions among these factors. Independent cohorts, longitudinal designs, medication-state comparisons, and mechanistic multimodal studies are required before any individual feature can be considered a stable neurophysiological marker (Michel and Bréchet 2026; Cai et al. 2026).

4.6 Limitations

The findings should be interpreted in light of several constraints. Most importantly, the development and external datasets differed substantially in both sample size and cohort context, and the locked classifier did not retain robust discrimination in ds003490. Although the dataset-origin analysis confirmed pronounced distribution shift (ROC-AUC = 0.966), the present design cannot determine how much of that shift arose from acquisition differences, medication state, cognitive composition, disease severity, or other cohort characteristics.

Methodological choices also remain to be tested independently. These include fixing the solution to four canonical microstates, retaining unsmoothed label sequences for the primary temporal analysis, and aligning templates across validation folds and cohorts. Unsmoothed sequences preserve brief transitions but may also be more sensitive to residual noise. Finally, calibration and model-attribution measures such as permutation importance, bootstrap stability, and SHAP summarize properties of the fitted model and should not be interpreted as clinical probabilities or causal neurophysiological evidence.

Future work should prioritize standardized multi-site acquisition, transparent reporting of microstate clustering and backfitting choices, harmonized preprocessing, medication- and disease-stage stratification, and prospective cross-cohort validation. Such designs will be necessary to determine which components of microstate dynamics represent stable PD-related electrophysiology and which primarily reflect dataset-specific measurement context.

5 Conclusion

This study shows why within-cohort EEG microstate discrimination and cross-cohort transportability should be evaluated separately. A multi-family representation combining static, transition, flow, and complexity features captured PD-related information in the development cohort, but performance attenuated under leakage-aware evaluation and approached chance during locked external testing. Strong dataset-origin discrimination in the same feature space indicated substantial cohort-specific structure.

The main contribution is therefore methodological rather than a claim of a deployable diagnostic biomarker: multi-family microstate features should be judged through progressively stringent validation and explicit distribution-shift assessment. Harmonized multi-site data and prospective external validation will be required before such features can be considered reproducible translational markers.

Supplementary information

Supporting analyses are provided in Online Resource 1, including power-line-noise removal, probability calibration, decision-confidence analysis, bootstrap feature stability, and cross-cohort top-feature distribution shift.

Acknowledgements

The authors gratefully acknowledge the support of the Beijing High-Level Public Health Technical Talent Development Project, the Provincial Key Laboratory Project of Intelligent Optimization Computing and Blockchain Technology, the Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (STIP; No. 2023L242), and the 2026 Taiyuan Normal University Graduate Student Practice and Innovation Project (Nos. SYYJSYC-2661 and SYYJSYC-2664).

Declarations

Funding

This work was supported by the National Natural Science Foundation of China (No. 62176177), the Beijing Natural Science Foundation (No. 7202072), and the Natural Science Foundation of Shanxi Province (No. 202303021221172).

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Ethics approval and consent to participate

The present study is a secondary analysis of publicly available, de-identified EEG datasets and involved no new participant recruitment or access to identifiable information. The original data collections were conducted by the source investigators under their respective institutional approvals.

Consent for publication

Not applicable.

Data availability

All EEG data analyzed in this study are publicly available from OpenNeuro. The development cohort was OpenNeuro ds004584 (version 1.0.0; https://doi.org/10.18112/openneuro.ds004584.v1.0.0). The independent external-validation cohort was OpenNeuro ds003490 (version 1.1.0; https://doi.org/10.18112/openneuro.ds003490.v1.1.0). All analyses used de-identified public data, and no new participant data were collected for the present study.

Materials availability

Not applicable.

Code availability

The custom analysis code used in this study is available from the corresponding authors upon reasonable request.

Author contributions

YXW drafted the manuscript and contributed to conceptualization, methodology, software development, formal analysis, and visualization. GMY contributed to conceptualization, methodology, supervision, and project administration. GXG contributed to methodology and validation. ML contributed to data curation, formal analysis, and validation. ZHW contributed to software development, data curation, and visualization. YKW and AL contributed to data curation and validation. YLZ and SPT contributed to resources and validation. HW contributed to conceptualization, methodology, and supervision. BW contributed to supervision and project administration. GMY, GXG, YLZ, SPT, HW, and BW reviewed and edited the manuscript. All authors read and approved the final manuscript.

References

Aarsland D, Batzu L, Halliday GM, et al (2021) Parkinson disease-associated cognitive impairment. Nat Rev Dis Primers. https://doi.org/10.1038/s41572-021-00280-3

Abreu R, Jorge J, Leal A, Koenig T, Figueiredo P (2021) EEG microstates predict concurrent fMRI dynamic functional connectivity states. Brain Topogr 34:41–55. https://doi.org/10.1007/s10548-020-00805-1

Asha SA, Sudalaimani C, Devanand P, et al (2024) Analysis of EEG microstates as biomarkers in neuropsychological processes—Review. Comput Biol Med 173:108266. https://doi.org/10.1016/j.compbiomed.2024.108266

Betrouni N, Derambure P, Defebvre L, et al (2026) Electroencephalography microstates detect early cognitive impairment in Parkinson's disease. Parkinsonism Relat Disord 146:108280. https://doi.org/10.1016/j.parkreldis.2026.108280

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Britz J, Van De Ville D, Michel CM (2010) BOLD correlates of EEG topography reveal rapid resting-state network dynamics. NeuroImage 52:1162–1170. https://doi.org/10.1016/j.neuroimage.2010.02.052

Cai J, Zhao Y, Song J, et al (2026) EEG microstates are associated with motor function in Parkinson's disease: a cross-sectional observational study. Brain Behav 16:e71151. https://doi.org/10.1002/brb3.71151

Cavanagh JF (2021) EEG: 3-Stim Auditory Oddball and Rest in Parkinson's. OpenNeuro [Data set]. https://doi.org/10.18112/openneuro.ds003490.v1.1.0

Costa TDDC, Machado CBDS, Lemos Segundo RP, et al (2023) Are the EEG microstates correlated with motor and non-motor parameters in patients with Parkinson's disease? Neurophysiol Clin 53:102839. https://doi.org/10.1016/j.neucli.2022.102839

Di Biase L, Ricci L, Caminiti ML, et al (2023) Quantitative high density EEG brain connectivity evaluation in Parkinson's disease: the phase locking value (PLV). J Clin Med 12:1450. https://doi.org/10.3390/jcm12041450

Elkholy MM, Aboubakr HH, Abd Elmonem NA, et al (2023) EEG spectral and connectivity parameters as cognitive biomarkers in Parkinson disease. Egypt J Neurol Psychiatry Neurosurg 59:54. https://doi.org/10.1186/s41983-023-00656-0

Fiorenzato E, Moaveninejad S, Weis L, et al (2024) Brain dynamics complexity as a signature of cognitive decline in Parkinson's disease. Mov Disord 39:305–317. https://doi.org/10.1002/mds.29678

Gschwandtner U, Bogaarts G, Roth V, et al (2023) Prediction of cognitive decline in Parkinson's disease patients with electroencephalography connectivity characterized by time-between-phase-crossing. Sci Rep 13:5093. https://doi.org/10.1038/s41598-023-32345-6

Handojoseno AM, Shine JM, Nguyen TN, et al (2015) Analysis and prediction of the freezing of gait using EEG brain dynamics. IEEE Trans Neural Syst Rehabil Eng 23:887–896. https://doi.org/10.1109/TNSRE.2014.2381254

Haydock D, Kadir S, Leech R, et al (2025) EEG microstate syntax analysis: a review of methodological challenges and advances. NeuroImage 309:121090. https://doi.org/10.1016/j.neuroimage.2025.121090

Koenig T, Prichep L, Lehmann D, et al (2002) Millisecond by millisecond, year by year: normative EEG microstates and developmental stages. NeuroImage 16:41–48. https://doi.org/10.1006/nimg.2002.1070

Koenig T, Studer D, Hubl D, et al (2005) Brain connectivity at different time-scales measured with EEG. Philos Trans R Soc Lond B Biol Sci 360:1015–1023. https://doi.org/10.1098/rstb.2005.1649

Lehmann D, Ozaki H, Pal I (1987) EEG alpha map series: brain micro-states by space-oriented adaptive segmentation. Electroencephalogr Clin Neurophysiol 67:271–288. https://doi.org/10.1016/0013-4694(87)90025-3

Li Z, Zhu X (2025) Application of EEG microstates in Parkinson's disease. Parkinsonism Relat Disord 136:107869. https://doi.org/10.1016/j.parkreldis.2025.107869

Liang Z, Tang B, Chang Y, et al (2024) State-related electroencephalography microstate complexity during propofol- and esketamine-induced unconsciousness. Anesthesiology 140:935–949. https://doi.org/10.1097/ALN.0000000000004896

Liu Z, Si L, Xu W, et al (2022) Characteristics of EEG microstate sequences during propofol-induced alterations of brain consciousness states. IEEE Trans Neural Syst Rehabil Eng 30:1631–1641. https://doi.org/10.1109/TNSRE.2022.3182705

Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems 30, pp 4765–4774.

Meng L, Wang D, Ma J, et al (2025) Unraveling Parkinson's disease motor subtypes: a deep learning approach based on spatiotemporal dynamics of EEG microstates. Neurobiol Dis 210:106915. https://doi.org/10.1016/j.nbd.2025.106915

Michel CM, Bréchet L (2026) EEG microstates: from methodological foundations to clinical translation. Trends Neurosci 49:433–449. https://doi.org/10.1016/j.tins.2026.04.003

Michel CM, Koenig T (2018) EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: a review. NeuroImage 180:577–593. https://doi.org/10.1016/j.neuroimage.2017.11.062

Oikonomou P, Akhoundi FH, Olfati N, et al (2026) Characteristics and mechanisms of cognitive impairment in Parkinson disease. Nat Rev Neurol 22:90–109. https://doi.org/10.1038/s41582-025-01163-x

Pei G, Hu M, He Y, et al (2025) Disruption of low-frequency narrowband EEG microstate networks in Parkinson's disease with mild cognitive impairment. J Neuroeng Rehabil 22:218. https://doi.org/10.1186/s12984-025-01747-0

Phipson B, Smyth GK (2010) Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. Stat Appl Genet Mol Biol 9:Article39. https://doi.org/10.2202/1544-6115.1585

Ponce-Bobadilla AV, Schmitt V, Maier CS, et al (2024) Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. Clin Transl Sci 17:e70056. https://doi.org/10.1111/cts.70056

Singh A, Cole R, Espinoza A, Cavanagh J, Narayanan N (2023) Rest eyes open. OpenNeuro [Data set]. https://doi.org/10.18112/openneuro.ds004584.v1.0.0

Tang Z, Du L, Tan X, et al (2026) Research on depression recognition based on EEG microstate functional connectivity. J Integr Neurosci 25:48264. https://doi.org/10.31083/JIN48264

Tarailis P, Artoni F, Koenig T, et al (2026) Short-term and long-term test-retest reliability of memory, complexity, and randomness of EEG microstates sequence. Cogn Neurodyn 20:19. https://doi.org/10.1007/s11571-025-10391-2

Yassine S, Gschwandtner U, Auffret M, et al (2023) Identification of Parkinson's disease subtypes from resting-state electroencephalography. Mov Disord 38:1451–1460. https://doi.org/10.1002/mds.29451

Zhao Y, Xu J, Xu X, et al (2026) EEG microstate alterations in Parkinson's disease dementia. Sci Rep 16:11278. https://doi.org/10.1038/s41598-026-40029-0

Supplementary Files

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

- OnlineResource1.pdf