Posterior-Aware Motor Phenotyping with Multimodal Imaging Validation in Parkinson's Disease*

Harsh Tirhekar[1,†], Priyanshi Yadav[2,†], and Chandrajit Bajaj[1,†,*]

[1]Department of Computer Science, Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX, USA
[2]Department of Biomedical Engineering, National Institute of Technology Raipur, Raipur, Chhattisgarh, India
[†]All authors contributed equally to this work. [*]Correspondence: bajaj@oden.utexas.edu

Abstract. Parkinson's disease exhibits heterogeneous motor presentations, yet data-driven subtyping is sensitive to modeling choices and often omits assignment uncertainty. We developed a posterior-aware Bayesian Gaussian mixture modeling framework using n=29,366 longitudinal MDS-UPDRS-III assessments from 1,847 PPMI participants. A predefined search over 2,912 configurations selected a five-state representation, and patient-block bootstrap assessed component-occupancy stability. Model-conditioned posteriors assigned 99.5% of assessments to a high-confidence category (Textbook) and 0.5% to intermediate-confidence boundary assignments (Chimera). Cross-granularity correspondence between five- and eight-component solutions was strong ($V = 0.945$). In exploratory longitudinal analyses, bradykinesia scores most frequently improved prediction of subsequent axial scores. Motor-state assignments were associated with DaTSCAN striatal binding ratios (n = 1,839) and small-magnitude FreeSurfer subcortical volume differences (n=1,706; η^2=0.005–0.010; 13/25 ROIs FDR-significant). Applying the fixed scaler and BGMM to BioFIND (n=310) without refitting yielded 99.7% high-confidence assignments. These findings support a reproducible visit-level motor-state representation with complementary imaging correlates. They do not establish five stable biological patient subtypes; repeated measures, possible severity confounding, and uncalibrated model posteriors limit interpretation.

Keywords: Parkinson's disease · Motor state modeling · Bayesian mixture models · DaTSCAN imaging · Structural MRI · Uncertainty quantification · Multi-scale clustering

1 Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder, with substantial motor heterogeneity that confounds clinical trials and limits

* An extended version of this paper is available at https://www.biorxiv.org/content/10.64898/2026.06.12.732003v1. For further details, see https://cvc-lab.github.io/projects/posterior-aware-pd-phenotyping/.

precision therapeutics [1,3,4]. Two patients with the same total severity score may exhibit fundamentally different profiles: one dominated by resting tremor with preserved gait, another by axial rigidity with no tremor [3].

The MDS-UPDRS-III [2] provides the international standard for motor assessment, comprising 33 items across five clinical domains: Tremor (10 items), Bradykinesia (10), Rigidity (4–5), Axial function (4), and Bulbar symptoms (3) [17]. Longitudinal cohorts such as PPMI [15] now contain tens of thousands of assessments, creating an opportunity for data-driven phenotyping at unprecedented scale. Despite this, four methodological gaps limit PD motor subtyping:

Gap 1: Hyperparameter sensitivity. Most clustering studies report a single configuration, e.g., k-means with k=3 [3] or GMM/ML-based subtype modeling [6], without exploring how results change across covariance structures, prior specifications, and truncation thresholds. Recent ML reviews emphasize the breadth of PD datasets and algorithms [7], while methods such as HYDRA [8] add further architectural and modeling choices.

Gap 2: Uncertainty quantification. Hard-assignment methods (k-means, HDBSCAN) assign every patient a single label with no confidence measure, yet patients near decision boundaries may exhibit mixed pathology spanning multiple phenotypes [4].

Gap 3: Multi-scale reconciliation. Different analyses yield different k (e.g., k=5 vs. k=8), typically treated as contradictory rather than complementary cross-granularity views [5].

Gap 4: Imaging validation. While neuroimaging-based subtyping methods exist [25,26], systematic validation of computationally derived motor phenotypes against both SPECT dopaminergic imaging and structural MRI morphometry using bootstrap-corrected nonparametric tests has not been performed.

Our approach. We address all four gaps with BGMMs [9], whose posterior vectors $\boldsymbol{\pi}_i$ yield uncertainty diagnostics (gap, KL [9], entropy [22]), Granger temporal analysis [20], Cramér's V correspondence [21], and multimodal imaging validation.

Main Contributions: We developed (1) a BGMM framework with 2,912-configuration sweep and bootstrap yielding a consistent k_{eff}=5 motor phenotypic partition; (2) a three-tier posterior triage (TEXTBOOK/ CHIMERA/ AMBIGUOUS); (3) Granger analysis identifying bradykinesia as a leading temporal predictor of subsequent motor domain decline; (4) cross-granularity partition correspondence (V=0.945) between k=5 and k=8 assignments; (5) DaTSCAN SPECT validation (n=1,839, $p<10^{-8}$); (6) FreeSurfer 7 MRI validation (n=1,706, 13/25 FDR-significant); and (7) BioFIND transfer without refitting (99.7%). Our pipeline is shown in fig. 1.

2 Related Work

Classical PD motor subtyping. The clinical gold standard divides patients into tremor-dominant (TD) and postural instability/gait difficulty (PIGD) subtypes using MDS-UPDRS-III ratio cutoffs [17], but roughly 30% of patients

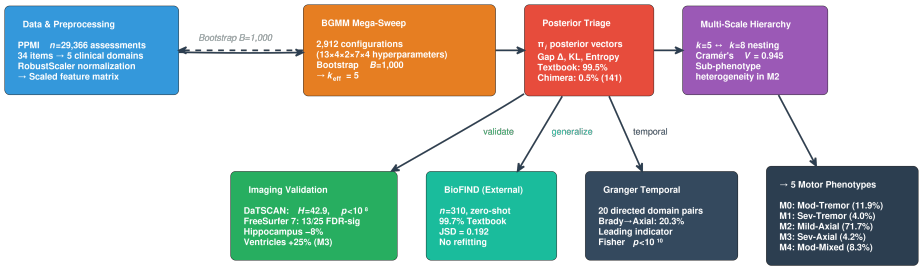

Fig. 1. Posterior-aware phenotyping pipeline: PPMI motor-domain preprocessing, BGMM configuration search and bootstrap validation, posterior triage, cross-granularity analysis, multimodal imaging validation, BioFIND transfer, and exploratory Granger analysis.

transition between subtypes over 5 years [3]. Adams et al. [10] proposed updated criteria but still rely on discrete cutoffs.

Data-driven and imaging-based approaches. Dadu et al. [6] applied machine learning to identify and predict PD subtypes and progression in two cohorts. Shokrpour et al. [7] reviewed ML datasets, algorithms, and challenges across acoustic, biomarker, imaging, movement, and multimodal PD studies. Vijayakumari et al. [8] investigated PD subtypes and REM sleep behavior disorder severity using brainstem tractography. Young et al. [25] introduced SuStaIn for disease subtyping from imaging, and Vogel et al. [26] identified brain-network subtypes via tau-PET. However, none bridge data-driven motor clustering with systematic multimodal imaging validation.

Bayesian mixture models. BGMMs provide posterior probability vectors and automatic k_{eff} via Dirichlet priors [9], but no prior work swept hyperparameters at the scale here (2,912 configs). The five-domain MDS-UPDRS-III structure [17,18,11] is validated via McDonald's ω [19], while temporal domain relationships are assessed via Granger predictability [20,12]. Cross-granularity partition correspondence is quantified via Cramér's V [21,5].

3 Methods

3.1 Data and Preprocessing

We analyze n=29,366 MDS-UPDRS-III assessments from 1,847 PD patients (median 14 visits/patient) in PPMI [15]. Visits with >20% missing items are excluded; remaining values are imputed per visit via median. The 34 motor assessment variables (33 MDS-UPDRS-III item-level scores, with bilateral limb items scored separately per side) are aggregated into five validated clinical domains [17]. All domain scores are RobustScaler-normalized; the fitted scaler is retained for out-of-cohort inference.

3.2 BGMM Systematic Configuration Search with Bootstrap Validation

We construct a configuration space $\mathcal{C}$ by varying five hyperparameter axes: truncation $K \in \{3, 5, 8, 10, 12, 15, 20, 25, 30, 40, 50, 70, 100\}$ (13 values); covariance $\in$ {full, tied, diag, spherical}; prior type (Dirichlet Process, DP, vs. Dirichlet distribution, DD); weight concentration α_0 (7 values, 10^{-4}–10^{2}); and mean precision (4 values), yielding $|\mathcal{C}| = 13{\times}4{\times}2{\times}7{\times}4 = 2{,}912$ configurations. Each is evaluated via Silhouette, Davies-Bouldin, Calinski-Harabasz, and k_{eff} (components with weight >1%). The top model is bootstrap-validated via B=1,000 patient-level resamplings.

3.3 Posterior Assignment-Confidence Categories

For the optimal BGMM, per-sample posterior vectors $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{ik})$ are computed. Each assessment is assigned to an assignment-confidence category:

$$\text{Category}(i) = \begin{cases} \textsc{Textbook} & \text{if } \max_j \pi_{ij} > 0.8 \\ \textsc{Chimera} & \text{if } 0.5 < \max_j \pi_{ij} \leq 0.8 \\ \textsc{Ambiguous} & \text{if } \max_j \pi_{ij} \leq 0.5 \end{cases} \tag{1}$$

Three complementary diagnostics quantify uncertainty: (i) *posterior gap* $\Delta_i = \pi_{i(1)} - \pi_{i(2)}$ [9]; (ii) *KL divergence from uniform* $D_{\text{KL}}(\boldsymbol{\pi}_i \| \mathbf{u})$; (iii) *posterior entropy* $H(\boldsymbol{\pi}_i) = -\sum_j \pi_{ij} \ln \pi_{ij}$ [22]. These are model-conditioned probabilities, uncalibrated against external ground truth. The thresholds (>0.8 and 0.5–0.8) were selected heuristically for descriptive analysis and were not optimized against a clinical ground truth. Chimeras are analyzed to identify which motor-state pair they bridge.

3.4 Exploratory Longitudinal Predictability via Granger Analysis

For patients with ≥8 visits (n_L=1,677), we test Granger predictability [20] between all 20 directed domain pairs (lag L=3, AIC-selected) using statsmodels [14]. Patient-level p-values are aggregated via Fisher's method: $\chi^2_{\text{Fisher}} = -2\sum_{i=1}^{n_L} \ln p_i \sim \chi^2_{2n_L}$. Results are reported as the percentage of patients with significant ($p < 0.05$) predictability per direction.

3.5 Cross-Granularity Partition Correspondence

We fit BGMM at k=5 and k=8 and compute a contingency table. Correspondence between five- and eight-component assignments is quantified via Cramér's $V = \sqrt{\chi^2/(n \cdot (\min(k_1, k_2) - 1))}$ [21]; $V > 0.5$ indicates strong cross-granularity correspondence. Cramér's V is symmetric and does not establish directional hierarchy.

3.6 Multimodal Imaging Validation Protocol

Phenotype assignments are linked to two complementary imaging modalities to assess complementary imaging associations. (i) *DaTSCAN SPECT:* Striatal binding ratios (SBR) for putamen and caudate nuclei are extracted from [^{123}I]FP-CIT SPECT scans for n=1,839 PPMI patients with both motor and imaging data. SBR differences across phenotypes are tested via Kruskal-Wallis with bootstrap confidence intervals (B=1,000). (ii) *Structural MRI:* FreeSurfer 7 subcortical segmentations (ASEG atlas, 25 ROIs) from T1-weighted MRI (n=1,706) are normalized by estimated total intracranial volume (eTIV) and compared across phenotypes using permutation-corrected Kruskal-Wallis tests (B=10,000) with Benjamini-Hochberg FDR correction. Effect sizes are quantified via η^2. Cortical thickness across 64 Desikan-Killiany parcels is also examined.

4 Results and Discussion

All computations were executed on a TACC Vista node (144 ARM Neoverse-V2 cores, 243 GB RAM, 120-worker parallelism).

4.1 Bootstrap-Stable k_{eff}=5 Macro-Phenotypes

The top-ranked configuration (full covariance, DD prior, α_0=0.1, K=5) yields k_{eff}=5.0±0.0 under bootstrap (mean posterior gap=0.995), confirming all five components remain occupied; this does not establish five as the unique biological granularity.

Table 1. Posterior divergence diagnostics.

Flag	Count	%	Criteria
Textbook	29,225	99.5	$\max \pi > 0.8$
Chimera	141	0.5	$0.5 < \max \pi \leq 0.8$
Ambiguous	0	0.0	$\max \pi \leq 0.5$
Mean gap Δ	0.995		$\overline{\max \pi}$=0.999

Table 2. Domain reliability [19].

Domain	Items	α	ω	$\bar{\lambda}$
Tremor	10	.937	**.950**	.811
Rigidity	4	.999	**.999**	.998
Bradykinesia	10	.888	**.927**	.747
Axial	4	.525	**.887**	.814
Bulbar	3	.375	**.755**	.712

High-confidence assessments (Textbook; n=29,225, 99.5%) and intermediate-confidence assessments (Chimera; n=141, 0.5%) are the primary triage categories; no assessments were low-confidence (Ambiguous). Intermediate-confidence assessments primarily involve two posterior boundary pairs: M2–M3 (n=87, 62%) and M1–M3 (n=54, 38%). M3 (Sev-Axial) participates in both of the most frequent intermediate-posterior boundary pairs (table 1). The high-confidence rate reflects the strong domain-score separation inherent in MDS-UPDRS-III items, which are operationally designed to capture distinct symptom dimensions, rather than an absence of clinical heterogeneity; this may motivate prospective

evaluation of motor-state-stratified trial analyses. Posteriors remain uncalibrated and were evaluated on the model-fitting dataset.

Domain reliability. McDonald's ω indicated acceptable-to-high internal consistency (table 2); α was low for the Axial (α=0.525) and Bulbar (α=0.375) domains, where ω substantially exceeds α (Axial: Δ=0.36; Bulbar: Δ=0.38), supporting the use of ω alongside α [19].

Phenotype profiles. fig. 2(a) shows RobustScaler-transformed domain score profiles: M3 (Sev-Axial) has the strongest separation (RobustScaler score =2.47 for Axial), and M2 (Mild-Axial, 71.7%) shows uniformly lower motor burden at assessed visits.

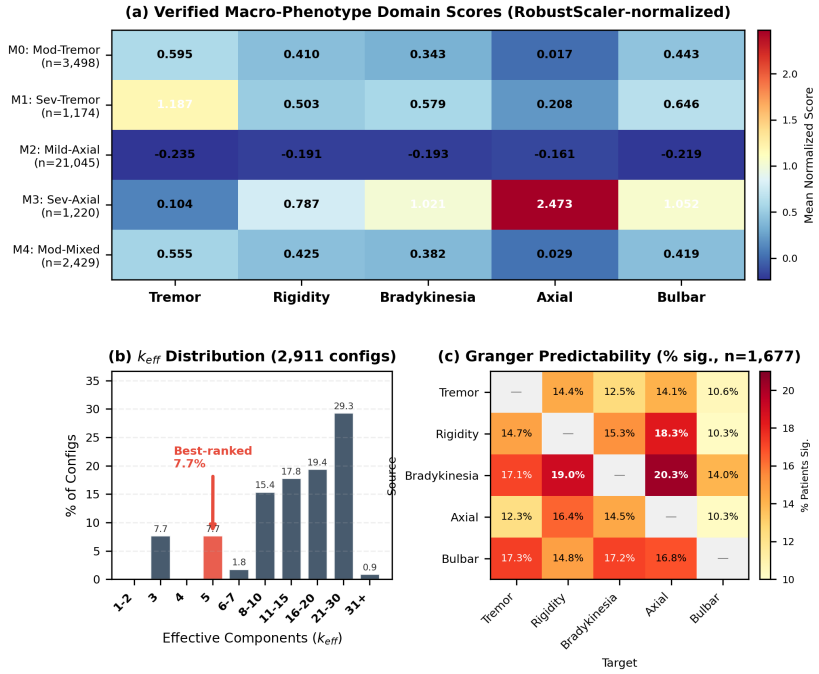

Fig. 2. Explainability panels: motor-state domain profiles, k_{eff} distribution across 2,911 converged configurations, and Granger predictability matrix.

4.2 Exploratory Longitudinal Predictability

In exploratory pairwise VAR analyses (fig. 2c, table 3), bradykinesia scores most frequently improved prediction of subsequent axial (20.3%) and rigidity scores (19.0%), while tremor showed weaker influence (10.6–14.4%) [12]. These percentages describe nominal within-patient predictability and do not establish causal ordering or a disease-mechanism hierarchy; cohort-level Fisher aggregation ($p<10^{-10}$) reflects aggregation across n_L=1,677 patients.

Table 3. Top 5 Granger-predictive directions (n=1,677).

Direction	% sig.
Bradykinesia → Axial	20.3
Bradykinesia → Rigidity	19.0
Rigidity → Axial	18.3
Bulbar → Tremor	17.3
Bulbar → Bradykinesia	17.2

Table 4. Cross-granularity partition correspondence (V=0.945; n=assessments).

Label	n	%	Eight-state assign. distrib.
M0 Mod-Trem	3,498	11.9	S1(89.7),S4(7.3),other(3.0)
M1 Sev-Trem	1,174	4.0	S7(**99.6**),other(0.4)
M2 Mild-Ax	21,045	71.7	S0(75),S6(13),S3(10)
M3 Sev-Ax	1,220	4.2	S2(80.2),S5(18.8)
M4 Mod-Mix	2,429	8.3	S4(93.1),S6(3.5)

4.3 Cross-Granularity Partition Correspondence

Cramér's V=0.945 indicates strong cross-granularity correspondence between the five- and eight-component partitions (table 4). Sev-Tremor maps to a single eight-state assignment S7 (99.6% purity), while Mild-Axial splits into three eight-state assignments (S0, S3, S6), revealing heterogeneity within the majority population (fig. 3b). 62% of intermediate-confidence assessments involve the M2–M3 posterior boundary pair (fig. 3a), and 38% involve M1–M3.

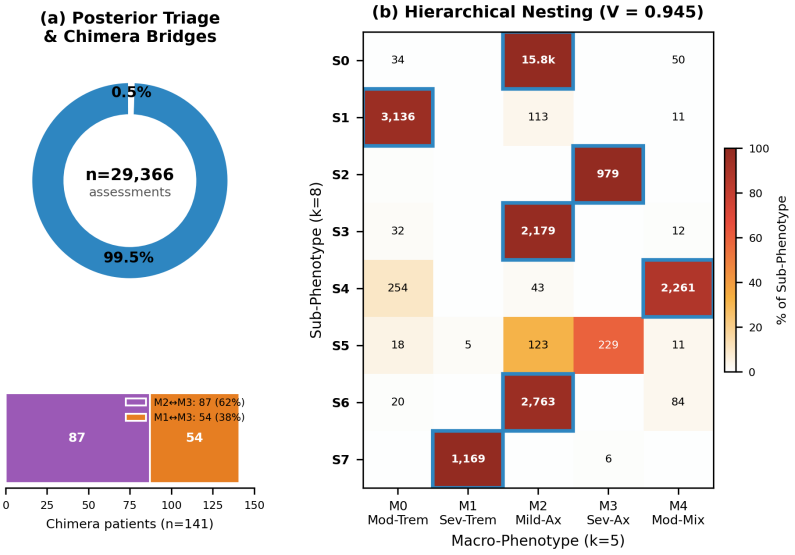

Fig. 3. Posterior triage and cross-granularity partition correspondence. (a) 99.5% Textbook vs. 0.5% Chimera; posterior boundary pairs: M2↔M3 (62%), M1↔M3 (38%). (b) 8×5 contingency matrix (V=0.945); blue borders: ≥85% purity.

4.4 Multimodal Imaging Validation

DaTSCAN SPECT reveals significant SBR differences (putamen: H=42.9, $p = 1.07\times10^{-8}$; caudate: H=25.3, p=4.4×10^{-5}; fig. 4a). M2 (Mild-Axial) shows

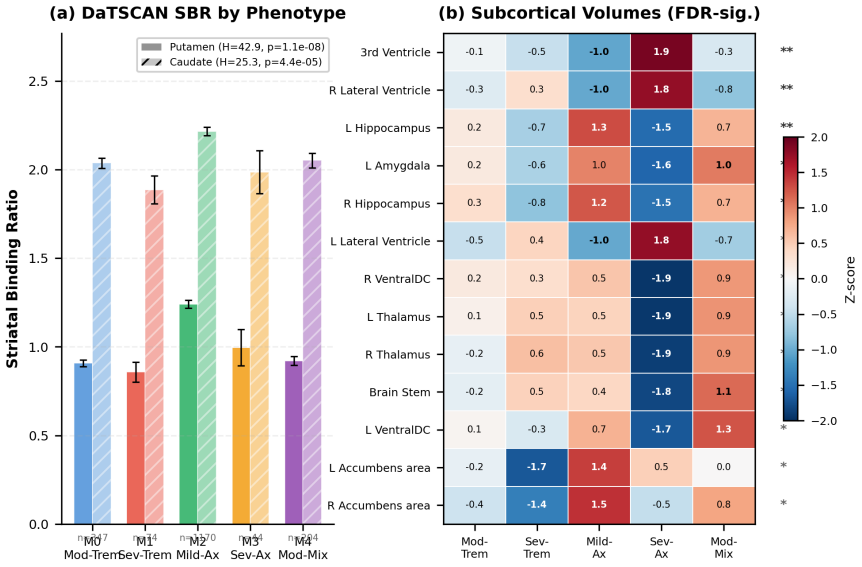

Fig. 4. Multimodal imaging validation. Motor states differ in DaTSCAN SBR and FDR-significant subcortical MRI volumes; MRI effect sizes are small.

highest SBR (1.24±0.76); M1 (Sev-Tremor, 0.86±0.48) and M0 (0.91±0.35) show greatest depletion.

Structural MRI (fig. 4b): 13/25 subcortical ROIs reach FDR-significance (η^2=0.005–0.010). M3 has lower bilateral hippocampal volumes (−8%, p_{FDR} = 0.004) [23], greater ventricular volume (+25%, η^2=0.010), and reduced thalamic/amygdala volumes [24]. Together, these multimodal results provide complementary imaging correlates for the discovered motor states; effect sizes are small (η^2=0.005–0.010).

4.5 Out-of-Cohort Model Transfer: BioFIND

We apply the PPMI-trained scaler and fixed BGMM to BioFIND [16] (n=310) **without refitting**. Inference yields 99.7% high-confidence assignments (JSD=0.192), demonstrating technical transferability under moderate distribution shift without establishing external phenotype validity.

Limitations. Four limitations warrant explicit acknowledgment. *(1) Motor-only input representation:* Motor phenotypic state discovery is based exclusively on MDS-UPDRS-III motor items; multimodal imaging serves as downstream biological validation, not as clustering input. This design is deliberate, using imaging as input would introduce circularity into the imaging validation step, but it constrains the richness of the initial partition. The discovered motor states should be interpreted as data-driven summaries of MDS-UPDRS-III structure, biologically supported by, rather than defined from, imaging data. *(2) Visit-level*

repeated-measures design: The BGMM treats n=29,366 assessments as observations from 1,847 patients with a median of 14 visits. While patient-level bootstrap partially mitigates intra-patient dependence, the model may capture longitudinal disease-state transitions alongside stable inter-patient variation. Future work will apply patient-aggregated or linear mixed-effects formulations to disentangle state dynamics from stable phenotypic structure. *(3) Effect sizes and BioFIND scope:* MRI effect sizes are small ($\eta^2 \leq 0.010$), expected for de novo baseline participants; BioFIND is cross-sectional. Longitudinal imaging replication remains an important future step. *(4) Possible severity and non-motor confounding:* Because total motor burden was not explicitly residualized before clustering, the states may partially reflect severity strata; cognitive and non-motor heterogeneity were outside the motor-only input [13]. Imaging associations therefore do not by themselves establish severity-independent biological subtypes.

5 Conclusion

We developed a posterior-aware BGMM framework yielding a five-state assessment-level motor representation with strong cross-granularity correspondence (V=0.945), detectable DaTSCAN associations, small-magnitude subcortical MRI correlates (η^2=0.005–0.010), and technical transfer to BioFIND without refitting. These groups should be interpreted as visit-level motor states rather than established biological patient subtypes; future work should pursue patient-level longitudinal models, severity-adjusted validation, and cross-validated posterior calibration.

Ethics Statement. This study is a secondary analysis of deidentified PPMI and BioFIND/AMP-PD data accessed under the respective data-use agreements; no new participant recruitment or intervention was performed.

Disclosure of Interests. The authors have no competing interests to declare.

Data and Code. PPMI data at `ppmi-info.org`; BioFIND via AMP-PD at `amp-pd.org`. Code: `https://github.com/CVC-Lab/posterior-aware-pd-phenotyping`.

Acknowledgments. We are very much appreciative the support from Dr. Dan Stanzione and UT-TACC Vista cluster resources (allocation PHY24002), and to the PPMI, BioFind patient databases. Many Thanks are also due to the Peter O' Donnell Foundation, and gifts from Jim Holland, and Michael-Connie Rasor.

References

1. Bloem, B.R., Okun, M.S., Klein, C.: Parkinson's disease. The Lancet **397**(10291), 2284–2303 (2021)
2. Goetz, C.G., et al.: Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS). Mov. Disord. **23**(15), 2129–2170 (2008)
3. Fereshtehnejad, S.M., et al.: Evolution of prodromal Parkinson's disease and dementia with Lewy bodies: a prospective study. Brain **142**, 2051–2067 (2019)

4. Simuni, T., et al.: A biological definition of neuronal α-synuclein disease: towards an integrated staging system for research. Lancet Neurol. **23**(2), 178–190 (2024)
5. Weingarten, C.P., et al.: Neuroimaging of Parkinson's disease: expanding views. Neurosci. Biobehav. Rev. **59**, 16–52 (2015)
6. Dadu, A., et al.: Identification and prediction of Parkinson's disease subtypes and progression using machine learning in two cohorts. npj Parkinson's Dis. **8**, 172 (2022)
7. Shokrpour, S., MoghadamFarid, A., Bazzaz Abkenar, S., et al.: Machine learning for Parkinson's disease: a comprehensive review of datasets, algorithms, and challenges. npj Parkinson's Dis. **11**, 187 (2025). https://doi.org/10.1038/s41531-025-01025-9
8. Vijayakumari, A.A., Sakaie, K.E., Fernandez, H.H., Walter, B.L.: Parkinson's disease subtypes and their association with probable rapid eye movement sleep behavior disorder severity: a brainstem tractography and machine learning investigation. Brain Imaging Behav. **19**(1), 189–194 (2025). https://doi.org/10.1007/s11682-024-00956-w
9. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2006)
10. Adams, C., et al.: Updated Parkinson's disease motor subtypes classification and correlation to cerebrospinal homovanillic acid and 5-hydroxyindoleacetic acid levels. Clin. Parkinsonism Relat. Disord. **8**, 100187 (2023)
11. Regnault, A., et al.: Optimizing the MDS-UPDRS Part III for early-stage Parkinson's: early supportive evidence for a limb-related bradykinesia/rigidity sub-score. npj Parkinson's Dis. **11**, 297 (2025)
12. Bartl, M., et al.: Longitudinal change and progression indicators using the Movement Disorder Society-Unified Parkinson's Disease Rating Scale in two independent cohorts with early Parkinson's disease. J. Parkinsons Dis. **12**(1), 437–452 (2022). https://doi.org/10.3233/JPD-212860
13. Aarsland, D.: Cognitive impairment in Parkinson's disease and dementia with Lewy bodies. Parkinsonism Relat. Disord. **22 Suppl 1**, S144–S148 (2016). https://doi.org/10.1016/j.parkreldis.2015.09.034
14. Seabold, S., Perktold, J.: statsmodels: Econometric and statistical modeling with Python. In: SciPy Conf., pp. 92–96 (2010)
15. Marek, K., et al.: The Parkinson Progression Marker Initiative (PPMI). Prog. Neurobiol. **95**(4), 629–635 (2011)
16. Kang, U.J., et al.: The BioFIND study: characteristics of a clinically typical Parkinson's disease biomarker cohort. Mov. Disord. **31**(7), 924–932 (2016)
17. Stebbins, G.T., et al.: How to identify TD and PIGD groups with the MDS-UPDRS. Mov. Disord. **28**(5), 668–670 (2013)
18. Park, J., et al.: Validation study of the official Korean version of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale. J. Clin. Neurol. **16**(4), 633–645 (2020). https://doi.org/10.3988/jcn.2020.16.4.633
19. McNeish, D.: Thanks coefficient alpha, we'll take it from here. Psychol. Methods **23**(3), 412–433 (2018)
20. Granger, C.W.J.: Investigating causal relations by econometric models and cross-spectral methods. Econometrica **37**(3), 424–438 (1969)
21. Cramér, H.: Mathematical Methods of Statistics. Princeton Univ. Press (1946)
22. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423 and 623–656 (1948)
23. Li, J., et al.: Cortical and subcortical morphological alterations in motor subtypes of Parkinson's disease. npj Parkinson's Dis. **8**, 167 (2022)

24. Uribe, C., et al.: Patterns of cortical thinning in nondemented Parkinson's disease patients. Mov. Disord. **31**(5), 699–708 (2016)
25. Young, A.L., et al.: Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. Nature Commun. **9**, 4273 (2018)
26. Vogel, J.W., et al.: Four distinct trajectories of tau deposition identified in Alzheimer's disease. Nature Med. **27**(5), 871–881 (2021)