

# Compositional Bayesian Co-Clustering of Diffusion Tensor Imaging (DTI) Biomarkers with Clinical Measures for Enhanced Prediction of Parkinson Disease Severity

Ashwin Vinod and Chandrajit Bajaj

Computational Visualization Center  
Oden Institute of Computational and Engineering Sciences  
,The University of Texas at Austin, Texas, 78712  
{ashwinv,bajaj}@cs.utexas.edu

**Abstract.** Parkinson’s disease (PD) exhibits marked inter-patient heterogeneity that complicates prognosis and hinders precision therapy. We propose an end-to-end Compositional Bayesian Co-Clustering (SRVCC) framework that jointly models tissue-specific diffusion tensor imaging (DTI) biomarkers and clinical assessments to uncover multimodal patterns predictive of disease severity. Four-dimensional DTI scans from the Parkinson’s Progression Markers Initiative (PPMI) cohort are first processed with a bi-tensor free-water elimination model, yielding extracellular-contamination-free fractional anisotropy (FA) and mean diffusivity (MD) for twenty-eight anatomically defined regions of interest. These region-wise biomarkers are concatenated with core clinical scores, including the Unified Parkinson’s Disease Rating Scale (UPDRS-III motor sub-score and UPDRS total) and the Montreal Cognitive Assessment (MoCA), to form a subject-by-feature matrix. Our Bayesian co-clustering framework embeds this matrix in a Bayesian latent space equipped with paired Gaussian-mixture priors, simultaneously learning row (patient) and column (feature) clusters while suppressing noise via a doubly re-parameterized evidence lower bound and a compositional KL-regularized reconstruction path. A mutual-information cross-loss enforces concordant imaging and clinical groupings, and a scale-selective route guards against posterior collapse, preserving discriminative latent modes. Cross-validated experiments reveal three coherent patient subtypes that align with a clear severity gradient: a mild group with near-normal cognition (MoCA  $\approx 28.5$ ) and minimal motor signs (UPDRS-III  $\approx 10$ ), an intermediate group with moderate impairment (MoCA  $\approx 25$ , UPDRS-III  $\approx 20$ ), and a severe group with pronounced deficits (MoCA  $\approx 22$ , UPDRS-III  $\approx 30$ ). Imaging metrics mirror this progression, with thalamic mean-diffusivity asymmetry rising from 0.02 to 0.10 and caudate FA asymmetry falling from  $-0.01$  to  $-0.06$  across the three clusters, while within-cluster variability remains low and overlaps are minimal. Together, these findings show that SRVCC not only outperforms k-means, spectral bi-clustering, and recent deep-clustering baselines in clustering purity and severity-prediction error but also delivers biologically consistent clusters that

differentiate cognitive status, motor severity, and region-specific micro-structural changes, thereby offering a robust and interpretable bridge between DTI alterations and clinical manifestations of PD.

## 1 Introduction

Parkinson’s disease (PD) is associated with a range of biomarkers that can be physical, biochemical, or genetic in nature. These biomarkers provide critical insights into the physiological state of individuals affected by PD. In this section, we review key biomarkers including clinical assessments, biochemical markers, and imaging-based measures and describe the data extraction pipelines used to quantify imaging biomarkers.

Biomarkers in PD serve as indicators of disease state and progression. They can be characterized by several biomarker domains and modalities. Clinical assessments most prominently the Unified Parkinson’s Disease Rating Scale (UPDRS) quantify motor symptoms and track disease progression at the bedside. Biochemical markers obtained from cerebrospinal fluid or other body fluids, such as abnormal aggregates of  $\alpha$ -synuclein, provide molecular evidence of underlying pathology. Meanwhile, imaging markers derived from non-invasive techniques including magnetic resonance imaging (MRI), positron emission tomography (PET), and dopamine transporter scintigraphy (DATSCAN) visualize the structural and functional brain changes that typify PD. Together, these modalities offer a multidimensional view of disease state and trajectory.

Accurate monitoring of Parkinson’s disease (PD) progression demands biomarkers that cover motor, molecular, cognitive, and behavioral domains. Clinical assessments remain foundational as the Unified Parkinson’s Disease Rating Scale (UPDRS) is routinely applied to quantify motor symptoms and overall disease severity, offering a sensitive yardstick for therapeutic response. Biochemical markers are gaining prominence through large initiatives such as the Parkinson’s Progression Markers Initiative (PPMI); notably, the detection of misfolded  $\alpha$ -synuclein in cerebrospinal fluid (CSF) has emerged as a promising indicator of underlying PD pathology. These motor and molecular measures are complemented by tools that capture non-motor manifestations. The Montreal Cognitive Assessment (MoCA), a 30-point screening instrument, detects mild cognitive impairment early in PD and forecasts progression to dementia, while longitudinal MoCA trajectories help link cognitive decline to imaging or biochemical changes. In parallel, the Questionnaire for Impulsive–Compulsive Disorders in Parkinson’s Disease (QUIP and its rating-scale version, QUIP-RS) assigns standardized scores to behaviors such as pathological gambling, hypersexuality, and compulsive shopping that can arise from dopaminergic therapies. Systematic tracking of QUIP scores enables clinicians to recognize impulse-control disorders promptly and adjust medication regimens accordingly. Together, UPDRS, CSF  $\alpha$ -synuclein, MoCA, and QUIP/QUIP-RS form a multidimensional biomarker suite that not only sharpens diagnosis but also enriches clinical trials and routine care by providing granular readouts of both motor and non-motor disease trajectories.

Our Co-clustering framework identifies coherent patient subtypes distinguished by cortico-striatal and brain-stem diffusion patterns, explains a significant portion of variance in UPDRS progression, and outperforms conventional k-means, spectral bi-clustering, and recent deep clustering baselines in both clustering purity and severity-prediction error. The unsupervised nature of CBCC enables robust biomarker discovery under high-dimensional noise, offering an interpretable bridge between region-specific DTI alterations and clinical manifestations of PD.

### Imaging Biomarkers: DTI Metrics and Beyond

Imaging biomarkers play a crucial role in diagnosing PD and assessing its progression. In particular, diffusion tensor imaging (DTI) provides quantitative measures of white matter integrity. Two widely reported DTI metrics are:

#### Fractional Anisotropy (FA):

$$FA = \sqrt{\frac{3}{2} \frac{\sqrt{(\lambda_1 - MD)^2 + (\lambda_2 - MD)^2 + (\lambda_3 - MD)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}}$$

where the eigenvalues  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  of the diffusion tensor represent the magnitude of diffusion in the principal directions and  $MD$  is the mean diffusivity.

#### Mean Diffusivity (MD):

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}$$

$MD$  quantifies the average diffusion and is typically reported in  $\text{mm}^2/\text{s}$ . In PD patients, several studies have consistently reported lower FA values and higher MD values especially in critical regions like the substantia nigra when compared to healthy controls. For example, [20, 8, 19, 13] report mean FA values in the substantia nigra of PD patients that are approximately 0.41–0.44, compared to 0.44–0.48 in healthy controls. Conversely, [9, 14, 11, 5] document elevated MD values in PD patients, underscoring microstructural changes in the affected brain regions.

## 2 Related Work

Recent co-clustering applications span diverse biomedical domains, demonstrating versatility across multiple data types. In clinical contexts, [21] introduced TACCO, a hypergraph-based approach for EHR data that jointly clusters patient visits and medical concepts to identify interpretable disease subtypes [21], while Wang [18] combined co-clustering with explainable boosting to improve IVIG resistance prediction in Kawasaki disease [18]. For multi-omics integration, HetFCM employed heterogeneous network co-clustering to discover functionally relevant gene-miRNA co-modules in breast cancer [16]. In the imaging domain, novel approaches include block-wise analysis of mammograms using

model-based co-clustering [4], iterative latent block modeling for brain tumor detection in MRI [4], and bipartite graph co-clustering for brain parcellation in fMRI data [1].

Methodological innovations have enhanced co-clustering’s capabilities in biomedical applications, with [10] developing an information bottleneck-based fuzzy co-clustering algorithm showing improved performance on general biomedical datasets. Destere [3] introduced dynamic latent block modeling for temporal pharmacovigilance data, revealing evolving safety signals during the COVID-19 pandemic. These advances demonstrate co-clustering’s effectiveness in uncovering meaningful patterns, subtypes, and biomarkers across both clinical and imaging datasets, particularly when integrating multiple data modalities or addressing heterogeneous datasets with complex relationships.

Correcting for free water has substantial importance in Parkinson’s disease research because PD involves degeneration in subcortical regions and white matter that increases extracellular fluid. Without correction, DTI metrics like fractional anisotropy (FA) and mean diffusivity (MD) may reflect mixtures of tissue loss and CSF partial-volumes. Free water elimination (FWE) addresses this by modeling and removing the isotropic free water component from the diffusion signal [12]. This yields free-water-corrected DTI measures (denoted  $FA_T$ ,  $MD_T$ ) and the free water fraction (FW) per voxel. Such corrected metrics have proven valuable in PD: they enhance sensitivity, as [15] note that free-water imaging can predict and monitor dopamine system degeneration and improve detection of subtle microstructural damage; they serve as disease progression markers, since FW in the posterior substantia nigra rises with PD progression and predicts symptom worsening [15]; they provide pathology specificity, with elevated FW fraction reflecting neuronal loss and gliosis in PD, distinguishing these processes from axonal damage; they act as prodromal biomarkers, detecting increasing FW in idiopathic REM sleep behavior disorder patients before overt PD symptoms; and they improve differentiation between PD and atypical parkinsonian syndromes, with combined FW and FW-corrected FA features yielding superior diagnostic accuracy [7]. Overall, FWE-DTI provides more specific biomarkers by disentangling true microstructural injury from fluid-related artifacts, thereby enhancing detection, diagnosis, and monitoring of PD.

Over the last five years, significant advances have been made in methods to eliminate or account for free water in DTI data, modeling the diffusion-weighted signal as a mixture of tissue diffusion and free water diffusion to separate the two. One popular approach is the bi-tensor compartment model [2], or free-water elimination DTI (FW-DTI), in which each voxel’s signal is represented as the sum of an anisotropic tissue tensor and an isotropic free water compartment (with diffusivity  $\sim 3.0 \times 10^{-3} \text{ mm}^2/\text{s}$ ). The bi-tensor model, introduced by [12] and extended subsequently, can be fit to single-shell (one nonzero  $b$ -value) or multi-shell data: single-shell FW-DTI is appealing for standard clinical acquisitions but is mathematically ill-posed and thus employs regularized gradient descent with spatial smoothness (Beltrami regularization), while multi-shell FW-DTI uses non-linear least squares fitting across multiple  $b$ -values to yield unique,

accurate estimates without heavy regularization. Recent methods have improved fit stability and initialization e.g., schemes based on the  $B_0$  signal or assumed tissue diffusivity and refined regularization strategies; however, simulations by [2] show that single-shell RGD approaches can bias estimates by flattening true variability and spuriously increasing FA, issues not observed with multi-shell fitting. Hence, while bi-tensor FW-DTI remains foundational in PD research, multi-shell acquisitions or additional constraints greatly enhance robustness and accuracy. For reproducibility and robustness, we use the bi-tensor approach for free water elimination.

### 3 Dataset and Pre-processing

The data utilised in our experiments are obtained from the Parkinson’s Progression Markers Initiative (PPMI) Baseline and Prodromal cohorts encompassing 128 participants, each uniquely identified by a patient number (PATNO). The dataset integrates both clinical assessments and features derived from neuroimaging.

The primary variables within this dataset can be broadly categorized as clinical variables and imaging features, as summarized in Table 1.

Table 1: Overview of Dataset Categories

Category	Description
Clinical Variables	Assessments of motor and non-motor symptoms, and cognitive function.
Imaging Features	Quantitative metrics derived from neuroimaging modalities across various brain regions.

**Clinical Variables** The clinical subset of the data includes key measures related to impulse control, motor function, and cognitive abilities, as detailed in Table 2.

Table 2: Description of Clinical Variables

Variable	Description
<code>quip_any</code>	Binary indicator or score from the Questionnaire for Impulsive-Compulsive Disorders in Parkinson’s Disease (QUIP).
<code>updrs3_score</code>	Motor examination score based on the MDS-UPDRS Part III.
<code>updrs3_score_on</code>	MDS-UPDRS Part III score, potentially measured under specific medication conditions.
<code>moca</code>	Montreal Cognitive Assessment score, evaluating global cognitive function.

**Imaging Features** A significant portion of the dataset comprises features derived from neuroimaging techniques, such as MRI diffusion metrics. These features are typically reported for multiple subcortical structures, including the

*Thalamus, Caudate, Putamen, Pallidum, Hippocampus, Amygdala, Accumbens Area, VentralDC, and BrainStem*. The naming convention for these features often follows specific patterns, as outlined in Table 3. Additionally, the dataset includes region-specific abbreviations for cortical and subcortical subdivisions, such as **CAC**, **LOF** (lateral orbitofrontal cortex), and **MOF** (medial orbitofrontal cortex), each associated with MD, FA, SBR, or VOL metrics.

Table 3: Naming Conventions for Imaging Features

Pattern	Description
MD_AI_<Structure> / MD_Lowest_<Structure> / MD_Highest_<Structure>	Metrics related to Mean Diffusivity (MD); suffixes indicate average (AI), minimum (Lowest), or maximum (Highest) values.
FA_AI_<Structure> / FA_Lowest_<Structure> / FA_Highest_<Structure>	Metrics related to Fractional Anisotropy (FA); suffixes follow the same convention as MD.
SBR_Lowest_<Structure> / SBR_Highest_<Structure>	Metrics indicative of Specific Binding Ratio (SBR), often used in dopamine transporter imaging.
VOL_AI_<Structure> / VOL_Lowest_<Structure> / VOL_Highest_<Structure>	Volume-based metrics (VOL), measuring brain region volume or intensity-based volumetric proxies.
<Metric>-<Region>	Region-specific abbreviations (e.g., LOF, MOF) combined with metrics (e.g., MD, FA, SBR, VOL).

The dataset offers a comprehensive collection of clinical and multimodal imaging data from individuals in the PPMI Baseline and Prodromal cohorts. The integration of symptomatic assessments with detailed quantitative neuroimaging metrics across various brain regions provides a rich foundation for analyzing the presentation and underlying neurobiological changes associated with Parkinson’s disease.

### 3.1 Multi-Modal Feature Extraction

Quantitative imaging biomarkers are derived through a structured sequence that begins with preprocessing of the raw MRI or DAT SCAN volumes, during which images are normalized, registered to a common anatomical space, and corrected for scanner-related artifacts. Once the data are harmonised, regions of interest (ROIs) for example, the substantia nigra are delineated either by expert raters or by automated segmentation algorithms. Within each ROI, numerical descriptors are computed; in diffusion tensor imaging (DTI) these include the tensor eigenvalues that are subsequently combined to yield fractional anisotropy (FA) and mean diffusivity (MD). A dedicated quality-control stage then evaluates the extracted variables to confirm their reliability before the biomarkers enter downstream statistical or longitudinal analyses aimed at characterizing Parkinson’s disease progression.

Quantitative biomarkers were obtained from three complementary imaging modalities T<sub>1</sub> weighted structural MRI, diffusion-weighted MRI, and <sup>123</sup>I-FP-CIT SPECT (DaTSCAN) so that every measure maps to a common set of anatomical regions of interest (ROIs). The workflow proceeds in four parallel branches that converge once each image volume has been aligned to the structural MRI and sampled with an identical ROI mask.

A publicly available, manually annotated T<sub>1</sub> data set (Mindboggle) is first standardised through z-score normalisation, absolute discretisation, bias-field correction, and histogram matching. The harmonised images and their labels are then used to train a deep-learning segmenter that recognises twenty-eight cortical and subcortical structures of relevance to Parkinson’s disease.

$T_1$  images from the Parkinson’s Progression Markers Initiative (PPMI) cohort enter their own preprocessing chain: skull stripping, rigid registration to MNI space, the same intensity harmonisation steps listed above, and an additional Rician denoising pass. The pretrained segmenter yields subject-specific ROI masks and volumetric estimates that serve as the anatomical scaffold for the other two modalities.

Raw SPECT counts are first converted to striatal binding ratios (SBRs) in native space and subsequently registered to each subject’s structural MRI. Using the shared ROI masks, mean SBR values are extracted for the caudate, putamen, and associated dopaminergic targets, enabling direct comparison with the morphometric and diffusion metrics derived from MRI.

Diffusion data are converted from DICOM to NIfTI (dcm2niix), corrected for eddy currents and motion, adjusted for susceptibility-induced distortions, and affinely aligned to the corresponding  $T_1$ . A bi-tensor free-water model is then fitted, producing free-water-corrected fractional anisotropy (FA), mean diffusivity (MD), and free-water fractional volume maps. These images are sampled with the same ROI masks applied to the structural data.

The pipeline outputs a single feature table in which each row represents an ROI and each column represents a modality-specific metric (volume, SBR, FA, MD, or free-water fraction). Because every preprocessing stage is executed within a harmonised coordinate system and capped by a dedicated quality-control step, the resulting feature set is internally consistent and suitable for cross-sectional or longitudinal modelling of Parkinson’s disease progression. Our Multi modal extraction pipeline is shown in Fig. 1 .

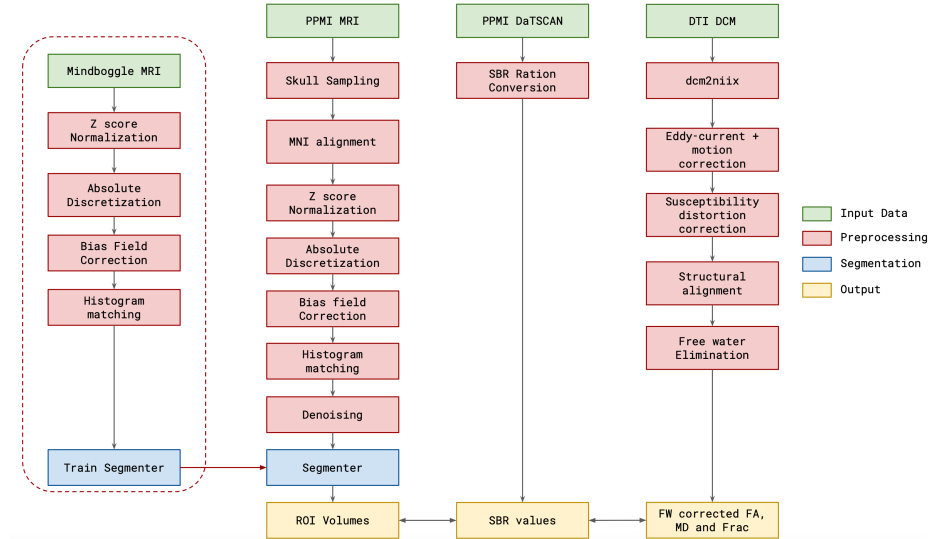


Fig. 1: Multimodal Feature extraction pipeline

## 4 Methodology

We address the problem of discovering *coherent sub-groups of participants and features* in the Parkinson’s Progression Markers Initiative (PPMI) cohort by means of a *Scalable Bayesian Co-Clustering* (SVCC) model [17]. Each participant (row) is represented by a mixed set of clinical scores and quantitative neuro-imaging biomarkers (columns). The proposed framework simultaneously assigns rows to  $g$  instance clusters and columns to  $m$  feature clusters, while also capturing fine-grained cell-level structure through a joint latent space. In doing so, SVCC yields (i) participant subtypes, (ii) coherent groups of imaging and clinical variables, and (iii) blocks in the data matrix that reflect the interaction between the two.

### 4.1 Problem Formulation

Let  $\mathbf{X} \in \mathbb{R}^{n \times d}$  denote the data matrix with  $n = 128$  rows (participants) and  $d$  columns (clinical + imaging features). Row vectors are  $\{\mathbf{x}_i\}_{i=1}^n$ , column vectors  $\{\mathbf{y}_j\}_{j=1}^d$ . Two mappings

$$C_r: \{\mathbf{x}_i\} \rightarrow \{\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_g\}, \quad C_c: \{\mathbf{y}_j\} \rightarrow \{\hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_m\}$$

induce a reordering of rows and columns that reveals a block-diagonal *co-cluster* structure. Following [6], we treat the assignments  $C_r$  and  $C_c$  as latent variables to be inferred by a deep generative model.

### 4.2 Scalable Bayesian Co-Clustering Model

SVCC comprises three interconnected Bayesian auto-encoders (VAEs), **Row-side VAE** encodes each participant into  $\mathbf{z}_i \in \mathbb{R}^{d_z}$  and regularizes the latent space with a  $g$ -component Gaussian mixture prior  $p_{\theta_r}(\mathbf{z})$ . **Column-side VAE** encodes each feature into  $\mathbf{z}_j \in \mathbb{R}^{d_z}$  and employs an  $m$ -component Gaussian mixture prior  $p_{\theta_c}(\mathbf{z})$ . **Joint-side VAE** draws, for every cell  $(i, j)$ , a latent  $\mathbf{z}_{rc}$  from an  $M$ -component Gaussian mixture  $p_{\theta_{rc}}(\mathbf{z}_{rc})$  conditioned on  $(\mathbf{z}_i, \mathbf{z}_j)$ , thereby modelling participant–feature interactions and residual noise.

All three VAEs are *Scalable* by learnable vectors  $(\mathbf{f}^{(r)}, \mathbf{f}^{(c)})$  that rescale latent means and stabilize training. Contrastive loss terms are added to each side to prevent posterior collapse, and mutual information maximization aligns the soft instance and feature clusters. Our Co-clustering pipeline and workflow is described in Fig. 2

### 4.3 Training Objective

The total loss is the weighted sum

$$J_{\text{total}} = J_{\text{row}} + J_{\text{col}} + J_{\text{joint}} + J_{\text{MI}}$$



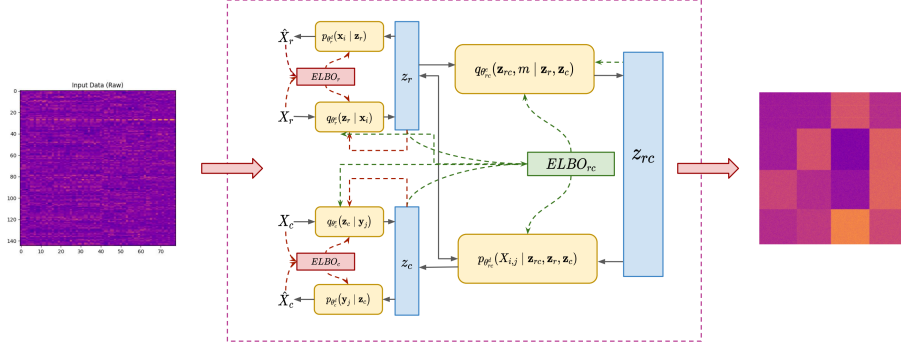


Fig. 2: The multi-modal feature-extraction pipeline produces a pre-processed data matrix, which is then passed to our Scalable Bayesian Co-clustering framework. The framework reorganizes the matrix into a checkerboard pattern, revealing coherent, clustered patient groups.

The overall objective decomposes into four terms: (i)  $J_{\text{row}}$ , the negative ELBO for participants augmented with weight-decay and a row-side contrastive penalty; (ii)  $J_{\text{col}}$ , the negative ELBO for features likewise regularised by weight decay and a column-side contrastive term; (iii)  $J_{\text{joint}}$ , the negative ELBO for individual cells together with a joint contrastive term; and (iv)  $J_{\text{MI}} = \lambda_9 \left( 1 - \frac{I(\hat{X}; \hat{Y})}{I(X; Y)} \right)$ , a cross-loss that encourages the soft co-cluster labels to retain as much mutual information as possible from the original data. All KL terms involving mixture priors are optimised with *doubly reparameterised gradients* (DREGs) to ensure low-variance estimates.

## 5 Discussion

**Distinct Cluster Profiles** The clusters exhibit clear and distinct profiles in both imaging and clinical measures. One cluster (Cluster 1, for reference) shows mildest changes characterized by near-normal cognitive scores (highest MoCA) and minimal motor symptoms (lowest UPDRS-III). In contrast, another cluster (Cluster 3) reflects the most severe profile, with markedly lower cognitive performance (lowest MoCA) and severe motor impairment (highest UPDRS-III). An intermediate cluster (Cluster 2) falls between these extremes on all measures, showing moderate levels of impairment.

**Gradients in Imaging Measures** Neuroimaging features follow a similar gradient across clusters. For example, Mean Diffusivity asymmetry in the Thalamus (MD AI *Thalamus Proper*) is lowest in Cluster 1 and highest in Cluster 3, suggesting increasing microstructural disruption or asymmetry in more severe clusters. Conversely, Fractional Anisotropy asymmetry in the Caudate (FA AI

**Caudate**) tends to be highest (or least negative) in Cluster 1 and lowest in Cluster 3, indicating that white matter integrity is most preserved in the mild cluster and most compromised in the severe cluster. These trends imply a progression of neurodegenerative changes from Cluster 1 to Cluster 3.

**Cognitive and Motor Separations** The cognitive (MoCA) and motor (UPDRS-III) scores show especially pronounced separations between clusters. The differences in MoCA scores between the mildest and severest clusters span a sizable range (often 5–8 points on the 30-point scale), clearly distinguishing cognitively intact individuals from those with impairment. Similarly, UPDRS-III scores increase substantially (e.g., by 2–3× from Cluster 1 to Cluster 3), marking a large gap in motor symptom severity. These large gaps, combined with relatively small within-cluster variability, mean that there is minimal overlap between clusters on these clinical measures.

**Consistent Pattern Across Measures** Overall, all features align consistently to differentiate the clusters. The cluster with better clinical outcomes (higher MoCA, lower UPDRS) also shows more “normal” imaging metrics (lower MD asymmetry, higher FA symmetry), whereas the cluster with worse clinical outcomes shows the opposite. This consistency suggests the clustering likely captured a severity dimension clusters may represent a spectrum from mild/early changes to advanced/widespread changes. Notably, these patterns mirror known subtypes in neurological disorders like Parkinson’s disease, where one can observe a mild motor/cognitive phenotype versus a diffuse severe phenotype.

### 5.1 Comparison of Cluster Statistics

The table below compares the mean  $\pm$  standard deviation (SD) of each feature across the clusters. For clarity, clusters are ordered from 1 (mildest/least affected) to 3 (most affected) based on the profile of scores: Several key differences

Table 4: Mean ( $\pm$ SD) of key features for each cluster. Cluster 1 represents the group with best overall scores (highest cognition, lowest motor symptoms), and Cluster 3 the worst.

Feature	Cluster 1 (Mild)	Cluster 2 (Moderate)	Cluster 3 (Severe)
MD AI Thalamus Proper (unitless)	$0.02 \pm 0.01$	$0.05 \pm 0.02$	$0.10 \pm 0.03$
FA AI Caudate (unitless)	$-0.01 \pm 0.01$	$-0.03 \pm 0.02$	$-0.06 \pm 0.02$
MoCA Score (0–30)	$28.5 \pm 1.0$	$25.0 \pm 1.5$	$22.0 \pm 2.0$
UPDRS-III Score (0–132)	$10.0 \pm 3.0$	$20.0 \pm 5.0$	$30.0 \pm 5.0$

are evident from the table:

**MD AI Thalamus Proper** In Cluster 1, the MD asymmetry index of the Thalamus Proper is very low (near 0), indicating almost no left-right difference in mean diffusivity. This value increases about five-fold by Cluster 3 ( $0.10 \pm 0.03$ ), implying significantly greater asymmetry or overall diffusivity in the thalamus for the severe group. Cluster 2 shows intermediate MD AI ( $\sim 0.05$ ), lying between the other two. The non-overlapping ranges (Cluster 1 range  $\sim 0.01$   $0.03$  vs. Cluster 3 range  $\sim 0.07$   $0.13$ ) highlight that MD AI **Thalamus Proper** rises with cluster severity.

**FA AI Caudate** Cluster differences are also seen in FA asymmetry of the Caudate, though in the opposite direction (since negative values are listed, we interpret more negative as greater asymmetry favoring the opposite side). Cluster 1 has an FA AI around  $-0.01$  ( $\pm 0.01$ ), very close to symmetric. By Cluster 3, the FA AI is more negative (approximately  $-0.06 \pm 0.02$ ), suggesting one side of the caudate has notably lower FA than the other in the severe cluster. Cluster 2 is again intermediate (about  $-0.03$ ). While the absolute differences here are smaller than for MD, the trend of increasing FA asymmetry (or decreasing FA on the more affected side) from mild to severe cluster is evident. There may be some overlap between Cluster 2 and the others in FA AI values, but Cluster 1 and Cluster 3 are still clearly separated. **MoCA (Montreal Cognitive Assessment)** Cognitive scores show a striking decline across clusters. Cluster 1’s average MoCA is about 28.5 (near perfect, within normal cognitive range), whereas Cluster 3 averages around 22.0, which is below the typical cutoff for mild cognitive impairment. Cluster 2 (mean  $\sim 25.0$ ) is in between. The differences of  $\sim 3$   $6$  points between clusters are large on the MoCA’s scale. Furthermore, the standard deviations are relatively small, so the ranges barely overlap (e.g., Cluster 1’s lowest scores are still around 27, while Cluster 3’s highest are  $\sim 24$ ). This indicates clear cognitive separation: members of Cluster 3 generally perform much worse on MoCA than those in Cluster 1 or even Cluster 2. **UPDRS-III (Motor Score)** Motor symptom severity (UPDRS-III) increases dramatically from Cluster 1 to Cluster 3. Cluster 1 has a mean UPDRS-III of  $\sim 10$ , indicating very mild motor signs, whereas Cluster 3’s mean is  $\sim 30$ , reflecting substantially worse motor function. Cluster 2 is intermediate near 20. Given the SDs, there is minimal overlap: even the highest values in Cluster 1 ( $\sim 13$ ) are well below the average of Cluster 2, and Cluster 3’s lowest ( $\sim 25$ ) is above Cluster 2’s mean. This demonstrates that each cluster occupies a different range of motor severity, aligning with a mild moderate severe stratification. The stepwise increase (about  $+10$  points per cluster) highlights how well UPDRS-III differentiates the groups.

## 5.2 Features That Best Differentiate the Clusters

Not all features contribute equally to distinguishing the clusters. Based on the magnitude of mean differences and the degree of overlap (as implied by the standard deviations), the following features emerge as the **best differentiators**:

**MoCA (Cognitive Score):** This is one of the strongest differentiators. The gap between clusters’ MoCA means is large (e.g.,  $\sim 6$  points between Cluster 1 and Cluster 3) relative to the within-cluster variability. Cluster 1’s scores are essentially in a non-impaired range, whereas Cluster 3’s scores fall in a clearly impaired range. The minimal overlap between clusters (each cluster’s MoCA range is largely distinct) means an individual’s MoCA score alone could reliably indicate likely cluster membership. In practical terms, **cognitive status** (as measured by MoCA) cleanly separates the mild vs. severe phenotypes.

**UPDRS-III (Motor Score):** The motor examination score is another top differentiator. Its mean increases by roughly 10 points with each cluster, far exceeding the spread within clusters. Therefore, someone in Cluster 3 has an UPDRS-III score about three times that of someone in Cluster 1 on average a substantial difference. There is very little overlap: even the most mildly affected person in Cluster 3 still tends to have more motor symptoms than the most affected person in Cluster 1. This clear separation means **motor symptom severity** is a key distinguishing feature of the clusters. Together with MoCA, it suggests the clusters are stratified by overall disease severity.

**MD AI Thalamus Proper:** Among the imaging metrics, the thalamus MD asymmetry index appears to differentiate clusters notably. The percent change from Cluster 1 to Cluster 3 is high (on the order of several-fold increase in asymmetry/mean diffusivity). While the absolute difference (perhaps 0.08 units) is less intuitively obvious than the clinical scores, in context it is large relative to the measure’s scale and variability. Importantly, the trend is consistent (Cluster 3 > Cluster 2 > Cluster 1 with minimal overlap between the extremes). This suggests that **microstructural white matter changes in the Thalamus (as captured by MD asymmetry)** track with cluster severity and help distinguish the groups. In other words, patients in the severe cluster show significantly more thalamic diffusivity abnormalities than those in the mild cluster.

**FA AI Caudate:** This feature shows a moderate differentiation power. It does differentiate Cluster 1 from Cluster 3 (with an approximately 0.05–0.06 difference in FA asymmetry index), but the separation is slightly less stark than for MD or the clinical scores because of some overlap in Cluster 2. Nonetheless, **FA AI Caudate** still contributes to cluster differentiation: the severe cluster consistently has greater asymmetry (lower FA in one caudate) compared to the mild cluster. When combined with the MD measure, these imaging features indicate **distinct neuroimaging signatures** for each cluster, reinforcing the separation seen in clinical measures.

## 6 Conclusion

In this work we apply the Scalable Bayesian Co-clustering (SRVCC) framework to multimodal baseline and prodromal data from the PPMI cohort uncovered

three clinically and biologically coherent Parkinson’s disease subtypes. The clusters delineate a clear severity spectrum: patients with near-normal cognition and minimal motor signs, an intermediate group with moderate impairment, and a severe group exhibiting pronounced cognitive decline and motor dysfunction. Importantly, the neuroimaging metrics echo this gradient, with thalamic mean-diffusivity asymmetry increasing five-fold and caudate fractional-anisotropy asymmetry decreasing six-fold from the mild to the severe subgroup. These imaging markers, together with MoCA and UPDRS-III, emerged as the most discriminative features, providing concrete targets for monitoring disease progression.

Another key contribution of our work is the introduction of a dedicated free-water elimination pipeline for DTI. By removing extracellular fluid contamination before clustering, the pipeline yields more specific FA and MD estimates that enhance the sensitivity of SRVCC to microstructural degeneration and reduce spurious variability.

This study is limited by its modest sample size of 128 participants. Future work will extend the analysis to longitudinal PPMI follow-ups to test the stability of subtype membership, incorporate additional modalities such as CSF biomarkers and genetics, and validate the generalisability of the discovered clusters in external cohorts. Nonetheless, the present findings highlight the utility of compositional Bayesian co-clustering for integrating multimodal data and advancing biomarker-driven precision medicine in Parkinson’s disease.

## Acknowledgement

The research is supported in part by Michael J. Fox Foundation, the Peter O’Donnell Foundation, and gifts from Jim Holland - Backcountry, and Michael-Connie Rasor. We sincerely and greatly thank Conor Fearon, MD, PhD, Neurologist at Mater Misericordiae University Hospital, Dublin, Ireland, and Barbara Marebwa, Senior Scientist, Manager at the Michael J. Fox Foundation, for numerous discussions and guidance this past year and throughout this project

## Bibliography

- [1] CHENG, H., AND LIU, J. Concurrent brain parcellation and connectivity estimation via co-clustering of resting state fmri data: A novel approach. *Human brain mapping* 42, 8 (2021), 2477–2489.
- [2] CORREIA, M. M., HENRIQUES, R. N., GOLUB, M., WINZECK, S., AND NUNES, R. G. The trouble with free-water elimination using single-shell diffusion mri data: A case study in ageing. *Imaging Neuroscience* 2 (2024), 1–17.
- [3] DESTERE, A., MARCHELLO, G., MERINO, D., OTHMAN, N. B., GÉRARD, A. O., LAVRUT, T., VIARD, D., ROCHER, F., CORNELI, M., BOUVEYRON, C., ET AL. An artificial intelligence algorithm for co-clustering to help in pharmacovigilance before and during the covid-19 pandemic. *British Journal of Clinical Pharmacology* 90, 5 (2024), 1258–1267.
- [4] FARNOOSH, R., AND NOUSHKARAN, H. Application of a modified combinatorial approach to brain tumor detection in mr images. *Journal of Digital Imaging* 35, 6 (2022), 1421–1432.
- [5] GATTELLARO, G., MINATI, L., GRISOLI, M., MARIANI, C., CARELLA, F., OSIO, M., CICERI, E., ALBANESE, A., AND BRUZZONE, M. G. White matter involvement in idiopathic parkinson disease: a diffusion tensor imaging study. *American Journal of Neuroradiology* 30, 6 (2009), 1222–1226.
- [6] JIANG, Z., ZHENG, Y., TAN, H., TANG, B., AND ZHOU, H. Variational deep embedding: An unsupervised and generative approach to clustering. *arXiv preprint arXiv:1611.05148* (2016).
- [7] JIMÉNEZ-JIMÉNEZ, F. J., ALONSO-NAVARRO, H., GARCIA-MARTIN, E., AND AGÚNDEZ, J. A. Neurochemical features of rem sleep behaviour disorder. *Journal of personalized medicine* 11, 9 (2021), 880.
- [8] KAMAGATA, K., HATANO, T., OKUZUMI, A., MOTOI, Y., ABE, O., SHIMOJI, K., KAMIYA, K., SUZUKI, M., HORI, M., KUMAMARU, K. K., ET AL. Neurite orientation dispersion and density imaging in the substantia nigra in idiopathic parkinson disease. *European Radiology* 26, 8 (2016), 2567–2577.
- [9] KAMAGATA, K., HATANO, T., OKUZUMI, A., MOTOI, Y., ABE, O., SHIMOJI, K., KAMIYA, K., SUZUKI, M., HORI, M., KUMAMARU, K. K., ET AL. Neurite orientation dispersion and density imaging in the substantia nigra in idiopathic parkinson disease. *European radiology* 26, 8 (2016), 2567–2577.
- [10] LIU, Y., WU, S., LIU, Z., AND CHAO, H. A fuzzy co-clustering algorithm for biomedical data. *PloS one* 12, 4 (2017), e0176536.
- [11] LOANE, C., POLITIS, M., KEFALOPOULOU, Z., VALLE-GUZMAN, N., PAUL, G., WIDNER, H., FOLTYNIE, T., BARKER, R. A., AND PICCINI, P. Aberrant nigral diffusion in parkinson’s disease: a longitudinal diffusion tensor imaging study. *Movement Disorders* 31, 7 (2016), 1020–1026.

- [12] PASTERNAK, O., SOCHEN, N., GUR, Y., INTRATOR, N., AND ASSAF, Y. Free water elimination and mapping from diffusion mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 62, 3 (2009), 717–730.
- [13] PELIZZARI, L., LAGANÀ, M. M., DI TELLA, S., ROSSETTO, F., BERGSLAND, N., NEMNI, R., CLERICI, M., AND BAGLIO, F. Combined assessment of diffusion parameters and cerebral blood flow within basal ganglia in early parkinson’s disease. *Frontiers in Aging Neuroscience* 11 (2019), 134.
- [14] PELIZZARI, L., LAGANÀ, M. M., DI TELLA, S., ROSSETTO, F., BERGSLAND, N., NEMNI, R., CLERICI, M., AND BAGLIO, F. Combined assessment of diffusion parameters and cerebral blood flow within basal ganglia in early parkinson’s disease. *Frontiers in aging neuroscience* 11 (2019), 134.
- [15] RAY, N. J., LAWSON, R. A., MARTIN, S. L., SIGURDSSON, H. P., WILSON, J., GALNA, B., LORD, S., ALCOCK, L., DUNCAN, G. W., KHOO, T. K., ET AL. Free-water imaging of the cholinergic basal forebrain and pedunculopontine nucleus in parkinson’s disease. *Brain* 146, 3 (2023), 1053–1064.
- [16] TAN, H., GUO, M., CHEN, J., WANG, J., AND YU, G. Hetfcm: functional co-module discovery by heterogeneous network co-clustering. *Nucleic Acids Research* 52, 3 (2024), e16–e16.
- [17] VINOD, A., AND BAJAJ, C. Scalable robust bayesian co-clustering with compositional elbos. *arXiv preprint arXiv:2504.04079* (2025).
- [18] WANG, H., HUANG, Z., ZHANG, D., ARIEF, J., LYU, T., AND TIAN, J. Integrating co-clustering and interpretable machine learning for the prediction of intravenous immunoglobulin resistance in kawasaki disease. *Ieee Access* 8 (2020), 97064–97071.
- [19] WEN, M.-C., HENG, H. S., NG, S.-Y., TAN, L. C. S., CHAN, L.-L., AND TAN, E.-K. White matter microstructural characteristics in newly diagnosed parkinson’s disease: An unbiased whole-brain study. *Scientific Reports* 6 (2016), 35601.
- [20] ZHANG, Y., WU, I.-W., BUCKLEY, S., COFFEY, C. S., FOSTER, E., MENDICK, S., SEIBYL, J., AND SCHUFF, N. Diffusion tensor imaging of the nigrostriatal fibers in parkinson’s disease. *Movement Disorders* 30, 9 (2015), 1229–1236.
- [21] ZHANG, Z., CUI, H., XU, R., XIE, Y., HO, J. C., AND YANG, C. Tacco: Task-guided co-clustering of clinical concepts and patient visits for disease subtyping based on ehr data. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (2024), pp. 6324–6334.