Supplementary Material

Supplementary Results and Discussion

Differences in neuroanatomical changes between two subgroups

We also performed visualization analyses for each subgroup, uncovering divergent neuroanatomical changes between the two subgroups. Specifically, in Subgroup 1, the subcortical regions, frontal, and temporal lobes displayed volumetric compression (Supplementary Figure 14). Beyond these regions, Subgroup 2 also exhibited atrophy in the parietal lobes (Supplementary Figure 15). Moreover, Subgroup 1 exhibited volume enlargement within the cerebellar white matter and areas adjacent to the ventricles and occipital lobes. Beyond these regions, Subgroup 2 showed an expansion in brain volume in the frontal areas. The voxel-wise statistical comparison (two-sided t-test) between these two subgroups is illustrated in Supplementary Figure 16.

A possible explanation for the faster cognitive decline in subgroup 1 may be due to severe atrophy in the hippocampus and its surrounding regions (Supplementary Figure 16). Previous cross-sectional studies have found reduced hippocampal volumes in PD with mild cognitive impairment[1,2]. On the other hand, severe atrophy in the parietal lobe and subcortical regions of patients in subgroup 2 may explain why they have faster motor progression. Due to the vital role of the parietal lobe in regulating movements[3,4]. And there exists a correlation between the macrostructural and microstructural properties of the subcortical regions and motor progression in PD[5–7]. Overall, clustering based on PD-specific features can provide insights into subgroups of patients with different patterns of disease progression.

Independence of PD-specific features from shared features

To test the independence of PD-specific features from shared ones, we conducted a quantitative analysis using the cosine similarity metric. Initially, we sampled PD-specific and shared feature vectors for each subject from the PPMI dataset. This resulted in two separate lists of vectors: one for PD-specific features and the other for shared features. Next, we calculated the cosine similarity between each pair of vectors from these two lists. This enabled us to quantify

the similarity between PD-specific and shared features. To ensure the reliability of our findings, we repeated this sampling process 10 times for each subject. For each iteration, we computed the average cosine similarity between the PD-specific and shared feature vectors. The results of these 10 iterations are presented in Supplementary Figure 4. Our findings reveal that the maximum average cosine similarity across all iterations was 0.056, while the minimum was 0.031 (Supplementary Figure 4). Additionally, we visualized the latent feature spaces using UMAP, which demonstrates a distinction between PD-specific and shared features (Supplementary Figure 5). In summary, our results indicate that PD-specific features are distinct from the shared features.

Sensitivity and specificity analyses

In our sensitivity analysis, we deliberately introduced varying degrees of noise into the background dataset. The background dataset, consisting of brain images from control participants, was perturbed with standardized normalized distribution noise at ratios of 0.5, 2, 4, and 8. Subsequently, we trained CVAE models using these noise-perturbed datasets and evaluated their performance. To quantitatively assess the sensitivity of our models, we computed the Kendall rank correlation (Kendall τ) coefficient between the PD-specific latent features extracted from models trained with different noise ratios and clinical as well as non-clinical measurements. This metric allowed us to measure the consistency of feature extraction across varying noise levels. To ensure the reproducibility and robustness of our findings, we replicated the latent feature sampling process for each patient ten times. Additionally, we conducted a one-sample t-test to statistically assess the significance of the correlations between the RDMs derived from clinical and non-clinical measurements. Remarkably, our analysis revealed that even as the noise-to-background ratio increased, there was no significant reduction in the performance of our models (Supplementary Figure 6). This observation underscores the robustness and sensitivity of our CVAE models.

In our specificity analysis, we visualized the PD-specific and shared features using UMAP. The results are presented in Supplementary Figure 5. From the scatter plot, we observed that in six of the ten samplings, the PD-specific latent features were clearly distinct from the shared ones. In two cases, the points representing these two sets of features were nearby but did not

overlap. Only in two instances did we note a minor overlap. These findings firmly demonstrate the specificity of our CVAE model.

Supplementary methods

The training process of the CVAE model

We defined the control participants as X, where $X=\{x_i\}_{i=1}^{N}$, and PD participants were defined as Y, where $Y=\{y_j\}_{j=0}^{M}$. We inputted the control participants into the shared encoder q_{ϕ_s} to generate the latent feature S_x, while the PD participants were inputted into both shared encoder q_{ϕ_s}, and specific encoder q_{ϕ_z} to generate the corresponding latent features S_y and Z_y. Both the shared and specific encoders are probabilistic encoders that model a multivariate high-dimensional Gaussian distribution.

We defined the decoder as $f(\theta)$, the reconstructed synthetic control brain is $\hat{x}=f_\theta([s_x,0])$, and the input to the decoder is a concatenation of S_x and a zero vector of the same dimension. The reconstructed PD brain was $\hat{y}=f([S_y,Z_y])$, generated from the concatenation of the corresponding latent features $[S_y,Z_y]$. We assumed that in the ideal scenario, the distributions of PD-specific feature and shared feature are mutually independent. To maximize the independence between these two distributions, a KL (Kullback-Leibler divergence[8])-based distribution correlation (DC) term was employed to quantify the relationship between the two and their joint conditional distribution.

$$\mathrm{DC} = -\mathrm{KL}\left(\bar{q} \,\middle\|\, q_{\phi_s}(s|y_i)\cdot q_{\phi_z}(z|y_i)\right) \tag{1}$$

The joint conditional probability distribution is defined as $\bar{\mathrm{q}}=q_{\phi_s}\cdot q_{\phi_z}(s,z|y_i)$. If the distributions S and Z are independent, the DC term equals zero.

The objective of the CVAE is a combination of reconstruction loss, distribution loss, disentangle loss, and discriminator loss. The reconstruction loss aims at matching the reconstructed data to the original data, and we used Mean Square Error (MSE) to measure the difference between the reconstructed data $\hat{x}_i$, $\hat{y}_j$ and original data x_i, y_j.

$$\text{Recon\_loss}\left(q_{\phi_s}, q_{\phi_z}, f\right) = \sum_{i=1}^{N} \text{MSE}(x_i, \hat{x_i}) + \sum_{j=1}^{M} \text{MSE}\left(y_j, \hat{y_j}\right) \quad (2)$$

Secondly, the distribution loss is computed using the KL divergence, which quantifies the dissimilarity between two distributions: the shared encoder q_{ϕ_s} and the PD-specific encoder q_{ϕ_z}. In the case of PD participants, the distributions $q_{\phi_s}(s|y_i)$ and $q_{\phi_z}(z|y_i)$ are used as approximations to the intractable posteriors over the latent variables S and Z, respectively. Similarly, the distribution $q_{\phi_s}(s|x_i)$ approximates the conditional posterior distribution of *S* for HC participants. In total, the whole distribution loss is defined like:

$$\text{KL\_loss}\left(q_{\phi_s}, q_{\phi_z}\right) = -\text{KL}\left(q_{\phi_s}(z|x) \| p(s)\right) - \text{KL}\left(q_{\phi_z}(z|y) \| p(z)\right) - \text{KL}\left(q_{\phi_s}(z|y) \| p(s)\right) \quad (3)$$

The third term in our model is the DC loss, which aims to maximize the independence between the PD-specific and shared features. However, directly calculating the DC term becomes challenging when using stochastic gradient descent on mini-batches of PD and HC datasets. Therefore, we employ an approximate term introduced previously to substitute for the DC term in our training process. The approximate term is defined like:

$$\text{DC\_loss(D)} = \log\left(D_\varphi(V_{(x,y)}) / \left(1 - D_\varphi(V_{(x,y)})\right)\right) \quad (4)$$

$V_{(x,y)} = [S_y, Z_y]$ is a vector concatenated by latent feature S_y and Z_y.

The fourth term is the discriminator loss, which serves a similar purpose to the DC_loss, but is specifically designed to differentiate whether the concatenated feature [s, z] originates from $\bar{q}$ or $q_{\phi_s}(s|y_i) \cdot q_{\phi_z}(z|y_i)$. On the other hand, the DC loss aims to encourage the distributions of q_{ϕ_s} and q_{ϕ_z} to converge towards independent distributions. Throughout the training process, we utilize a discriminator D_φ to effectively distinguish the target, and employ a density-ratio trick inspired by the method proposed by *Kim, H. and Mnih A*[9] to approximate the DC term. We use the binary cross-entropy loss function to distinguish whether a sample is from the decoder or the encoder. The discriminator is trained to minimize the loss function, while the encoder and decoder are trained to maximize it, resulting in the separation of the latent features. The final objective function can be expressed as:

$$\text{Totall\_loss} = \text{Recon\_loss}\left(q_{\phi_s}, q_{\phi_z}, f\right) + \lambda \text{KL\_loss}\left(q_{\phi_s}, q_{\phi_z}\right) + \mu \text{DC\_loss(D)} + \text{Discri\_loss(D)} \quad (5)$$

Where μ and λ are two hyperparameters that control the relative importance of each loss function during the training process. We aim to obtain the optimal encoders q_{ϕ_s}, q_{ϕ_z} decoder f, and discriminator D through this total loss function. Hyperparameters μ and λ are used to regulate the relative importance of each loss function throughout the training process. Our objective is to derive the optimal encoders q_{ϕ_s}, q_{ϕ_z}, decoder f, and discriminator D by minimizing the total loss function.

$$q_{\phi_s}, q_{\phi_z}, f, D = \arg \min_{\mathrm{f}} \max_{q_{\phi_s}, q_{\phi_z}, D} \text{Total\_loss}\left(q_{\phi_s}, q_{\phi_z}, f, D\right) \tag{6}$$

We configured the CVAE model with μ = 10 and λ = 0.9, using a learning rate of 0.001. The model achieved convergence after 300 epochs. For the VAE, we set μ = 100, λ = 0.9, learning rate= 0.001, and trained it for 300 epochs as well. Both models were trained until the total loss was minimized and showed no further changes.

Supplementary Tables

Supplementary Table 1 Comparison of the similarity of the extracted features with various nonclinical and clinical characteristics

Property	N	Specific > VAE	Shared > VAE
PPMI			
Site	321	$\Delta\tau$= -0.01, t_9 = -1.72, p = -0.119	$\Delta\tau$= 0.01, t_9 = 1.86, p = 0.095
Age	481	$\Delta\tau$= 0.02, t_9 = 4.21, p = 0.002	$\Delta\tau$= 0.00, t_9 = 0.78 p = 0.454
Gender	485	$\Delta\tau$= 0.06, t_9 = 20.21, p <0.001	$\Delta\tau$= -0.00, t_9 = -0.82, p = 0.432
Education	484	$\Delta\tau$= -0.01, t_9 = -1.14, p =0.283	$\Delta\tau$= -0.00, t_9 = -0.01, p = 0.989
Putamen SBR	318	$\Delta\tau$= 0.02, t_9 = 6.28, p <0.001	$\Delta\tau$= -0.01, t_9 = -1.60, p = 0.145
Serum NfL	425	$\Delta\tau$= 0.03, t_9 = 5.23, p < 0.001	$\Delta\tau$= 0.01, t_9 = 1.34, p = 0.215
MDS-UPDRS III	456	$\Delta\tau$= 0.01, t_9 = 2.52, p = 0.033	$\Delta\tau$= -0.00, t_9 = -0.40, p = 0.701
MoCA	483	$\Delta\tau$= -0.01, t_9 = -2.20, p = 0.056	$\Delta\tau$= 0.00, t_9 = 0.08, p = 0.941
α-Synuclein	362	$\Delta\tau$= -0.02, t_9 = -4.47, p = 0.002	$\Delta\tau$= 0.01, t_9 = 0.72, p = 0.490
T-Tau	419	$\Delta\tau$= 0.02, t_9 = 4.53 p = 0.002	$\Delta\tau$= -0.00, t_9 = -0.06 p = 0.956
P-Tau	397	$\Delta\tau$= 0.02, t_9 = 4.94, p < 0.001	$\Delta\tau$= 0.01, t_9 = 0.78, p = 0.457
Aβ42	358	$\Delta\tau$= 0.01, t_9 = 2.02, p = 0.075	$\Delta\tau$= -0.00, t_9 = -0.13, p = 0.896
Race	485	$\Delta\tau$= 0.03, t_9 = 1.07, p = 0.312	$\Delta\tau$= 0.00, t_9 = 0.02, p = 0.982
Duration	480	$\Delta\tau$= -0.02, t_9 = -3.30, p = 0.990	$\Delta\tau$=- 0.00, t_9 = -0.65, p = 0.531
Striatum SBR	318	$\Delta\tau$= 0.00, t_9 = 0.34, p = 0.742	$\Delta\tau$= -0.01, t_9 = -0.88, p = 0.401
Caudate SBR	318	$\Delta\tau$= -0.01, t_9 = -1.58, p = 0.149	$\Delta\tau$= -0.00, t_9 = -0.46, p = 0.5
SAHZJU			
Gender	373	$\Delta\tau$= 0.13, t_9 = 53.68, p < 0.001	$\Delta\tau$= -0.00, t_9 = -0.09, p = 0.928
Age	373	$\Delta\tau$= 0.03, t_9 = 9.73, p < 0.001	$\Delta\tau$= 0.00, t_9 = 0.69, p = 0.508
Education	373	$\Delta\tau$= 0.02, t_9 = 7.11, p < 0.001	$\Delta\tau$= -0.00, t_9 = -0.67, p = 0.517
MoCA	336	$\Delta\tau$= 0.01, t_9 = 1.68, p = 0.127	$\Delta\tau$= -0.00, t_9 = -0.29, p = 0.777
Duration	373	$\Delta\tau$= 0.01, t_9 = 1.37, p = 0.205	$\Delta\tau$= 0.00, t_9 = 0.18 p = 0.860
SAHSU			
Age	71	$\Delta\tau$= 0.03, t_9 = 2.57, p = 0.030	$\Delta\tau$= -0.01, t_9 = -1.22, p = 0.252
Gender	71	$\Delta\tau$= 0.08, t_9 = 17.21, p < 0.001	$\Delta\tau$= 0.01, t_9 = 2.00, p = 0.077
Education	71	$\Delta\tau$= 0.02, t_9 = 1.87, p = 0.094	$\Delta\tau$= 0.01, t_9 = 0.76, p = 0.466
Durtaion	71	$\Delta\tau$= -0.00, t_9 = -0.15, p = 0.885	$\Delta\tau$= 0.00, t_9 = 0.23, p = 0.820
MDS-UPDRS III	71	$\Delta\tau$= 0.04, t_9 = 3.17, p = 0.011	$\Delta\tau$= -0.00, t_9 = -0.06, p = 0.95

Supplementary Table 2 Demographics and clinical characteristics of two PD subgroups in PPMI

	Subgroup 1	Subgroup 2	*P*-value
Age	63.25 (56.22,70.20)	61.50 (54.20,68.60)	0.024
Gender (female%)	46.24%	23.28%	0.000
Education	16.00 (14.00,18.00)	16.00 (14.00,18.00)	0.133
Putamen SBR (median, [Q1, Q3])	0.83 (0.63,0.98)	0.77 (0.66,0.93)	0.321
Serum NfL (median, [Q1, Q3])	12.55 (9.19,17.25)	10.90 (8.25,14.30)	0.015
MDS-UPDRS III (median, [Q1, Q3])	20.00 (14.00,27.00)	120.00 (15.00,28.00)	0.419
MoCA (median, [Q1, Q3])	27.00 (25.00,29.00)	27.00 (26.00,29.00)	0.837
α-Synuclein (median, [Q1, Q3])	1489.80 (1124.48,1920.93)	1263.50 (978.48,1640.78)	0.002
T-Tau (median, [Q1, Q3])	157.80 (127.50,206.45)	150.80(122.15,196.83)	0.099
P-Tau (median, [Q1, Q3])	13.67 (11.04,17.83)	12.72 (10.52,16.77)	0.060
Aβ42 (median, [Q1, Q3])	840.30 (640.95,1131.75)	825.25 (615.13,1046.50)	0.405
Race (White%)	94.70%	96.17%	0.603
Duration	213.00(61.00,699.00)	152.00(61.00,457.50)	0.024
Striatum SBR	1.38(1.14,1,64)	1.36(1.20,1.62)	0.965
Caudate SBR	2.01(1.62,2,34)	2.01(1.71,2.36)	0.633

For categorial variables, chi-square test was used to identify differences between subgroups. For other quantitative variables, a one-way ANOVA analysis was performed for comparison.

Supplementary Table 3 RSA results of the study site and properties of PD patients

Variables	T-value	*P* -value
Age	0.00	0.724
Gen	-0.00	0.699
Education	0.13	<0.000
Putamen SBR	0.01	0.027
Serum_NfL	0.01	<0.000
UPDRS III	0.01	0.005
MoCA	0.01	0.029
α-Synuclein	-0.02	<0.000
T-Tau	0.00	0.441
P-Tau	0.00	0.443
Aβ42	0.02	<0.000
Race	-0.04	<0.000
Duration	-0.01	<0.000
Striatum SBR	-0.01	0.002
Caudate SBR	-0.00	0.170

Supplementary Table 4 The architecture and parameters of the CVAE model

Layer (type)	Output Shape	Param #
Shared Encoder		
Conv3d	[-1, 64, 32, 32, 32]	1,792
ReLU	[-1, 64, 32, 32, 32]	0
Conv3d	[-1, 128, 16, 16, 16]	221,312
ReLU	[-1, 128, 16, 16, 16]	0
Linear	[-1, 128]	67,108,992
ReLU	[1,128]	0
Linear	[-1, 16]	2,064
Linear	[-1,16]	2,064
Specific Encoder		
Conv3d	[-1, 64, 32, 32, 32]	1,792
ReLU	[-1, 64, 32, 32, 32]	0
Conv3d	[-1, 128, 16, 16, 16]	221,312
ReLU	[-1, 128, 16, 16, 16]	0
Linear	[-1, 128]	67,108,992
ReLU	[1,128]	0
Linear	[-1, 16]	2,064
Linear	[-1,16]	2,064
Decoder		
Linear	[-1, 128]	4,224
ReLU	[-1, 128]	0
Linear	[-1, 524288]	67,633,152
ReLU	[-1, 524288]	0
ConvTranspose3d	[-1, 32, 32, 32, 32]	262,176
ReLU	[-1, 32, 32, 32, 32]	0
ConvTranspose3d	[-1, 16, 64, 64, 64]	32,784
ReLU	[-1, 16, 64, 64, 64]	0
ConvTranspose3d	[-1, 1, 64, 64, 64]	433
Sigmoid	[-1, 1, 64, 64, 64]	0

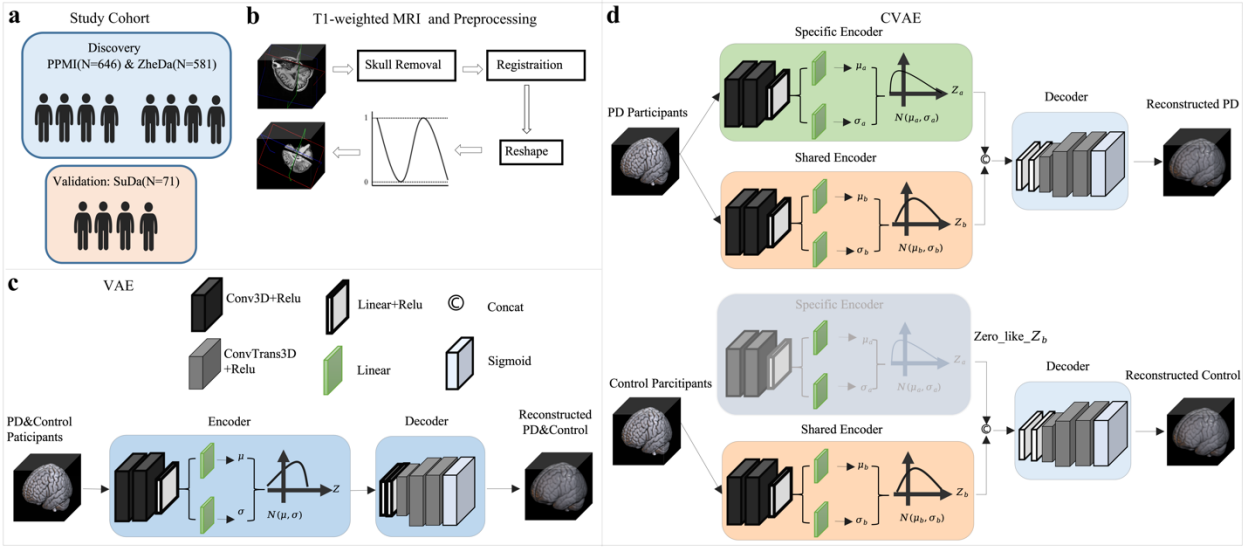
Supplementary Table 5 The architecture and parameters of the VAE model

Layer (type)	Output Shape	Param #
Encoder		
Conv3d	[-1, 96, 32, 32, 32]	2,688
ReLU	[-1, 96, 32, 32, 32]	0
Conv3d	[-1, 192, 16, 16, 16]	497,856
ReLU	[-1, 192, 16, 16, 16]	0
Linear	[-1, 128]	100,663,424
ReLU	[1,128]	0
Linear	[-1, 32]	4,128
Linear	[-1,32]	4,128
Decoder		
Linear	[-1, 128]	4,224
ReLU	[-1, 128]	0
Linear	[-1, 786432]	101,449,728
ReLU	[-1, 786432]	0
ConvTranspose3d	[-1, 192, 32, 32, 32]	2,359,488
ReLU	[-1, 192, 32, 32, 32]	0
ConvTranspose3d	[-1, 96, 64, 64, 64]	1,179,744
ReLU	[-1, 96, 64, 64, 64]	0
ConvTranspose3d	[-1, 1, 64, 64, 64]	2,593
Sigmoid	[-1, 1, 64, 64, 64]	0

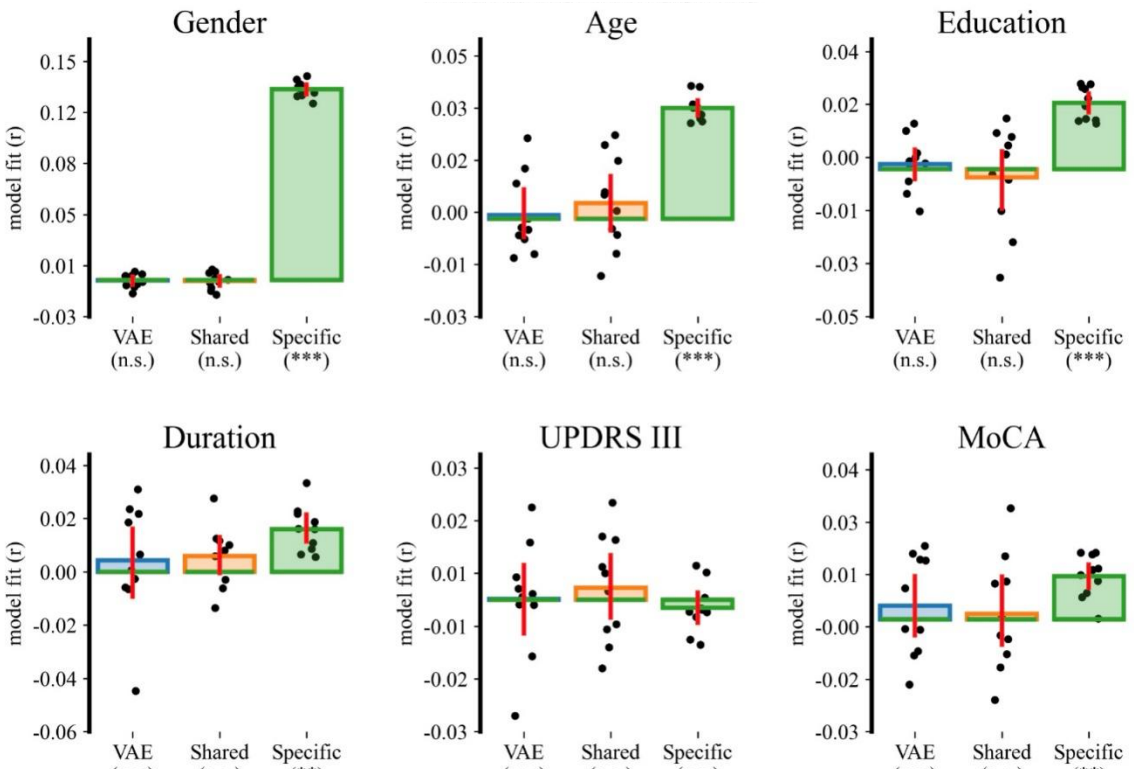
Supplementary Table 6 Abbreviation and full name list in the paper.

Abbreviation	Full name
Aβ42	Amyloid-β 1-42
ANTs	Advanced Normalization Tools
CVAE	Contrastive Variational Autoencoders
CSF	cerebrospinal fluid
DAT	dopamine transporter
IQR	interquartile range
LME	linear mixed-effects
MDS-UPDRS III	Movement Disorder Society-Unified Parkinson Disease Rating Scale Part-III
MoCA	Montreal Cognitive Assessment
NfL	Neurofilament light chain protein
PD	Parkinson's disease
PPMI	Parkinson's Progression Markers Initiative
P-tau	Phosphorylated tau
RDMs	representational dissimilarity matrices
RSA	representational similarity analysis
SD	standard deviation
SBR	specific binding ratios
SPECT	single photon emission computed tomography
T-tau	total tau

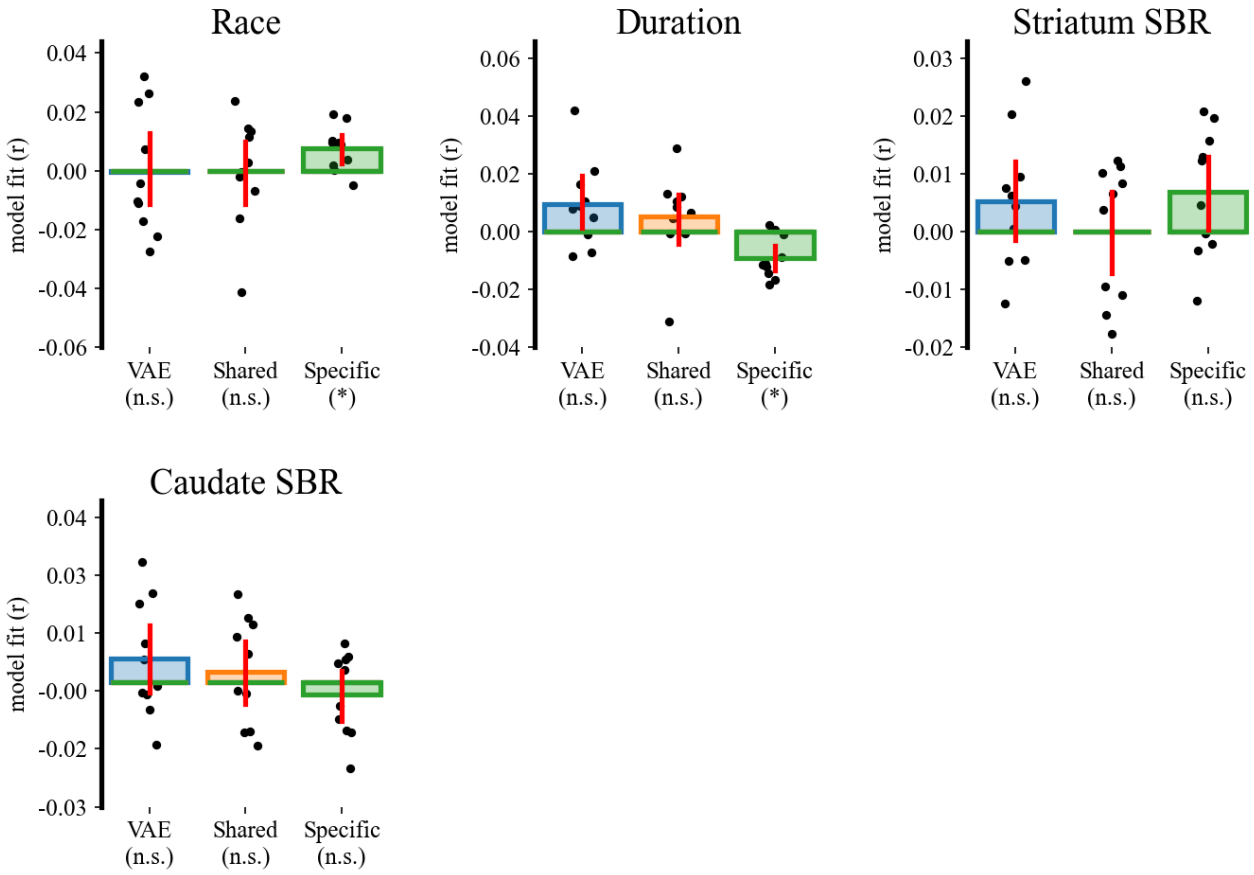
Supplementary Figures



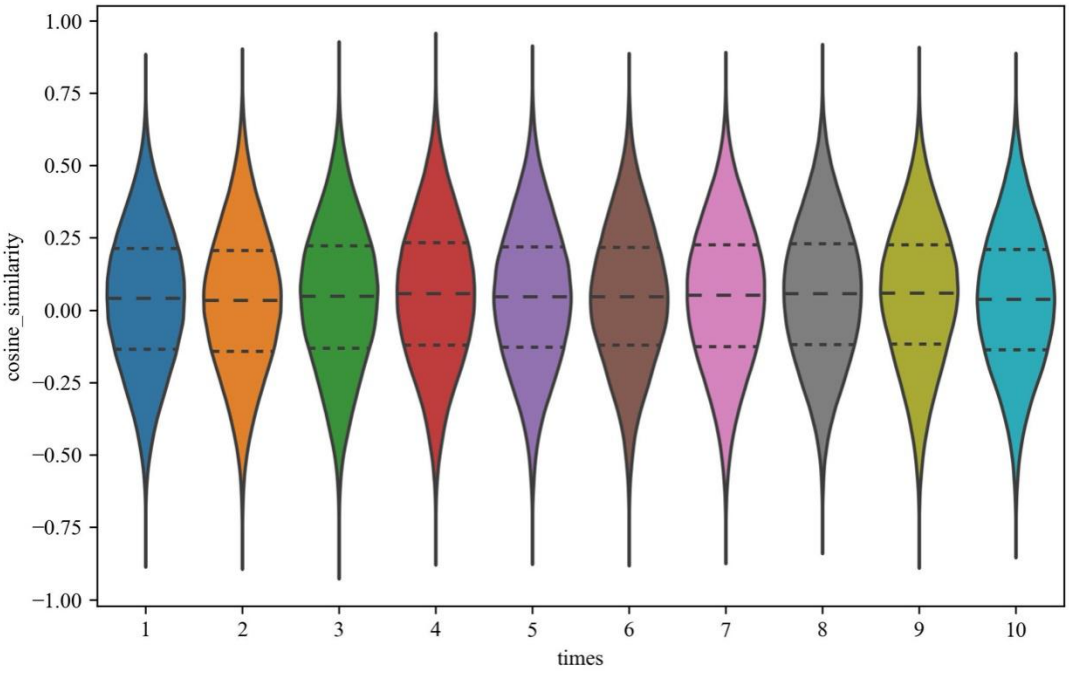
Supplementary Figure 1. **Overview of VAE and CVAE models and the preprocessing pipeline. a** The study cohort: The PPMI dataset (n=646) and SAHZJU dataset (n=581) were used to train a CVAE and VAE model, while the SAHSU dataset (n=71) was used for model validation. **b** The preprocessing pipeline involved several steps in FreeSurfer, including motion correction, intensity normalization to address inhomogeneity, and non-brain tissue removal. After FreeSurfer processing, the MRI images were registered to the standard MNI template (MNI152-2009c) and resampled to the size of 64×64×64. **c** The network architecture of the VAE: The VAE model consists of encoder and decoder modules. The encoder module encodes input data into a latent space representation, while the decoder module reconstructs the input data from the latent space. **d** The network architecture of the CVAE: The CVAE model includes similar modules to the VAE. However, it also incorporates additional modules to separate PD-specific variation from the shared variation with the Control participants.



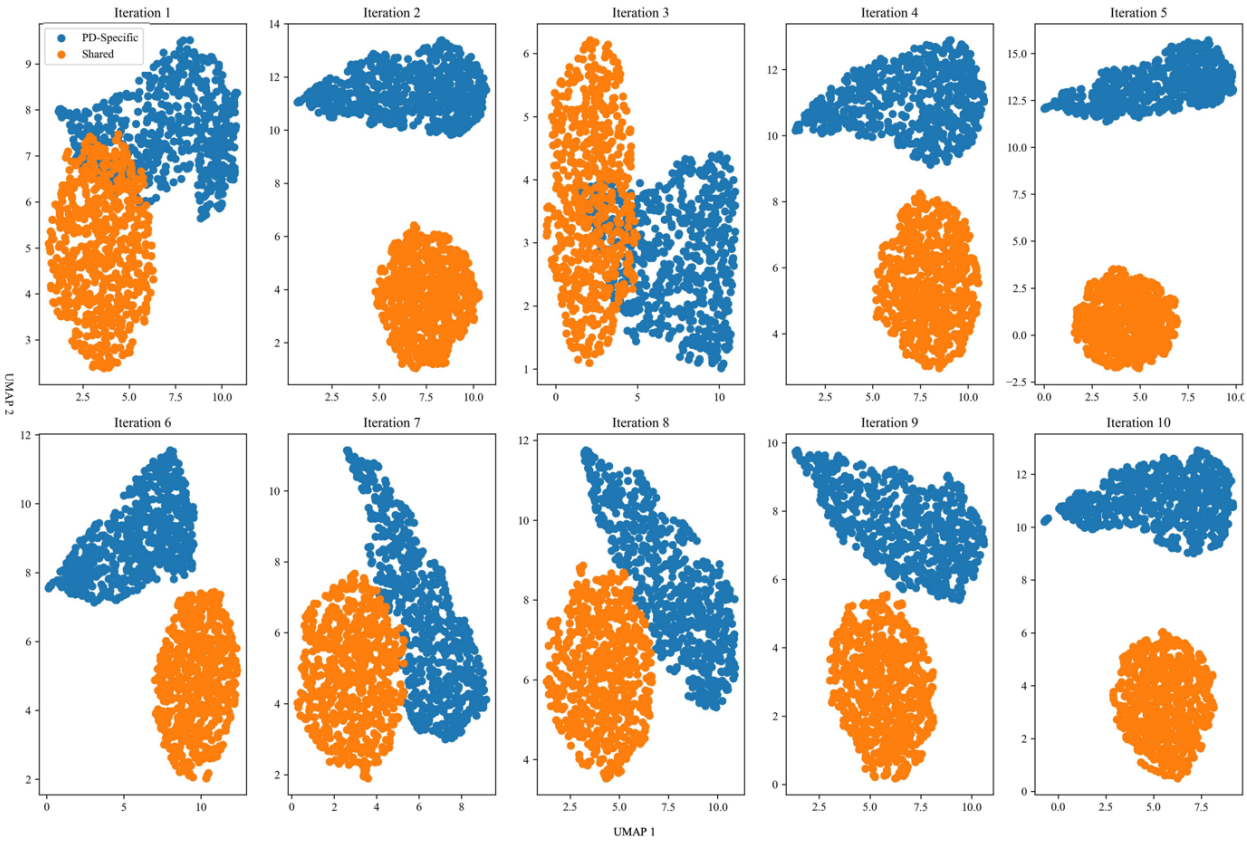
Supplementary Figure 2. **Representational Similarity Analysis (RSA) results of SAHZJU dataset.** Bar plots show representational similarity between neuroanatomical features (VAE's, CVAE's shared, and PD-specific) and patient properties. (* $p <$ .05, ** $p <$.001, *** $p <$ .0001).



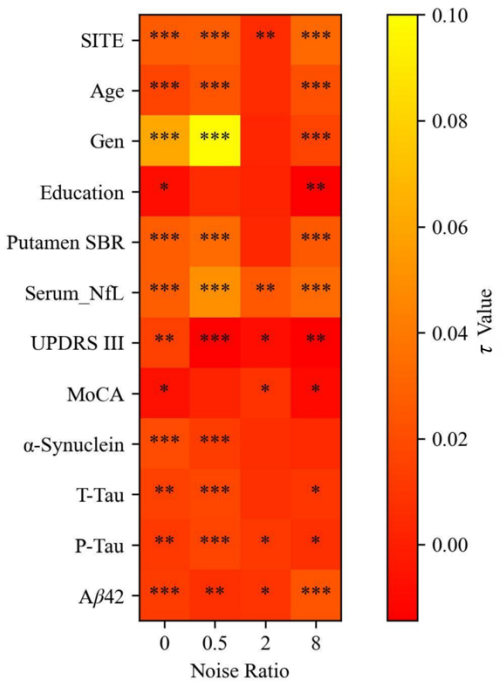
Supplementary Figure 3. **Representational Similarity Analysis (RSA) results of PPMI dataset.** Bar plots show representational similarity between neuroanatomical features (VAE's, CVAE's shared, and CVAE PD-specific) and patient properties. (* $p <$ .05, ** $p <$.001, *** $p <$ .0001).



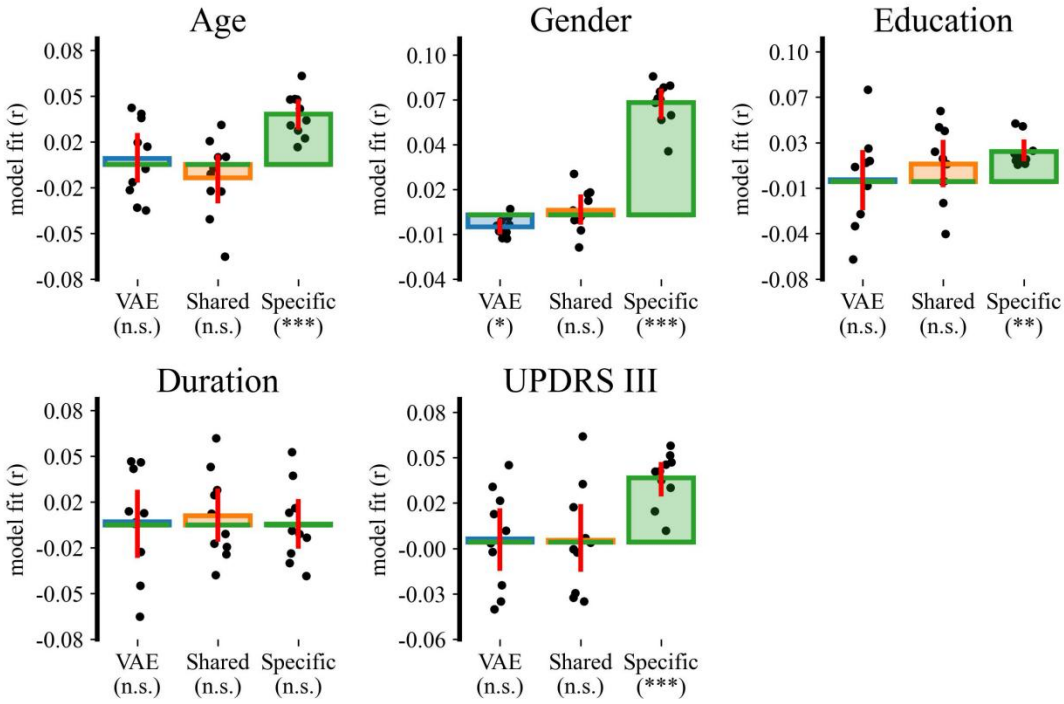
Supplementary Figure 4. The cosine similarity between the paired PD-specific and shared features of each PD patients in PPMI, each patient is sampled 10 times repeatedly.



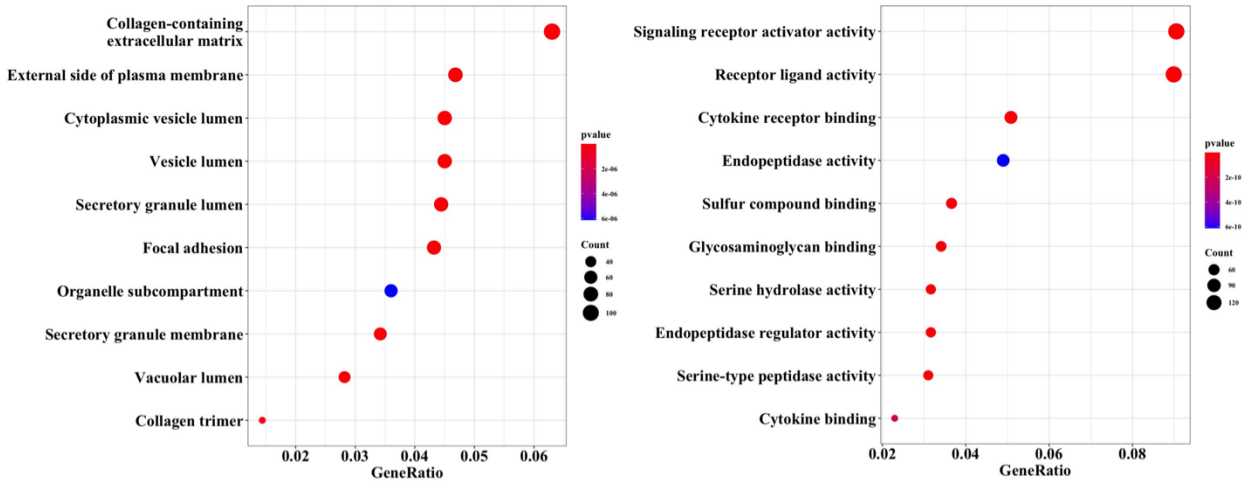
Supplementary Figure 5. The visualization of PD-Specific and Shared latent features generated by the PD-specific encoder and Shared encoder through UMAP during 10 times sampling.



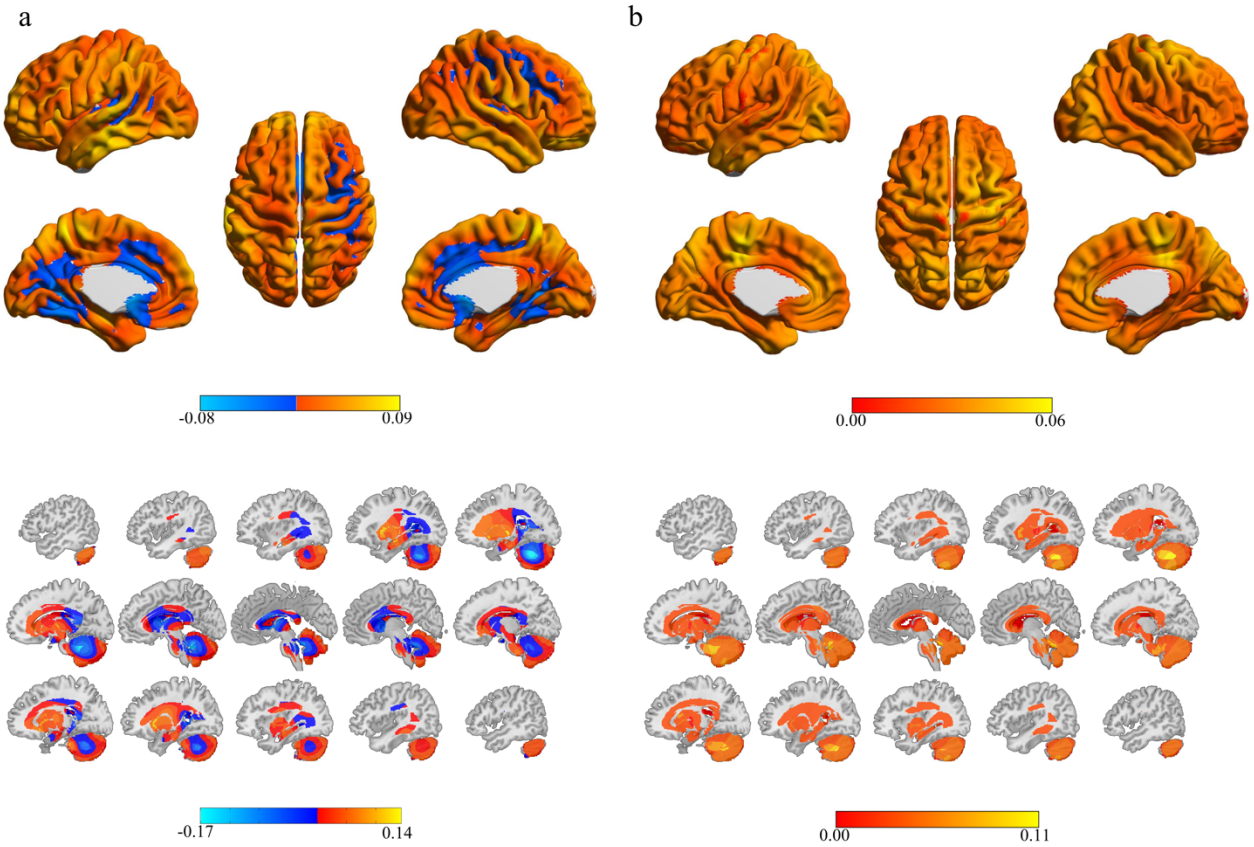
Supplementary Figure 6. Sensitivity analysis of the CVAE model. We conducted a comprehensive evaluation by introducing varying ratios (0.5, 2, 8) of standard normalized distribution noise to the pre-processed brain images of control participants and subsequently trained separate CVAE models for each noise level. To assess the reliability of the PD-specific latent features generated by these models, we performed a Kendall rank correlation analysis, comparing them to clinical and non-clinical measurements. For robustness, the PD-specific latent feature of each subject was sampled ten times. Finally, a one-sample t-test was employed to evaluate the consistency of our findings across different samples, ensuring that the observed Kendall rank correlations were statistically significant. This comprehensive sensitivity analysis demonstrates the robustness and reliability of our CVAE model in generating PD-specific latent features, regardless of the noise level introduced.



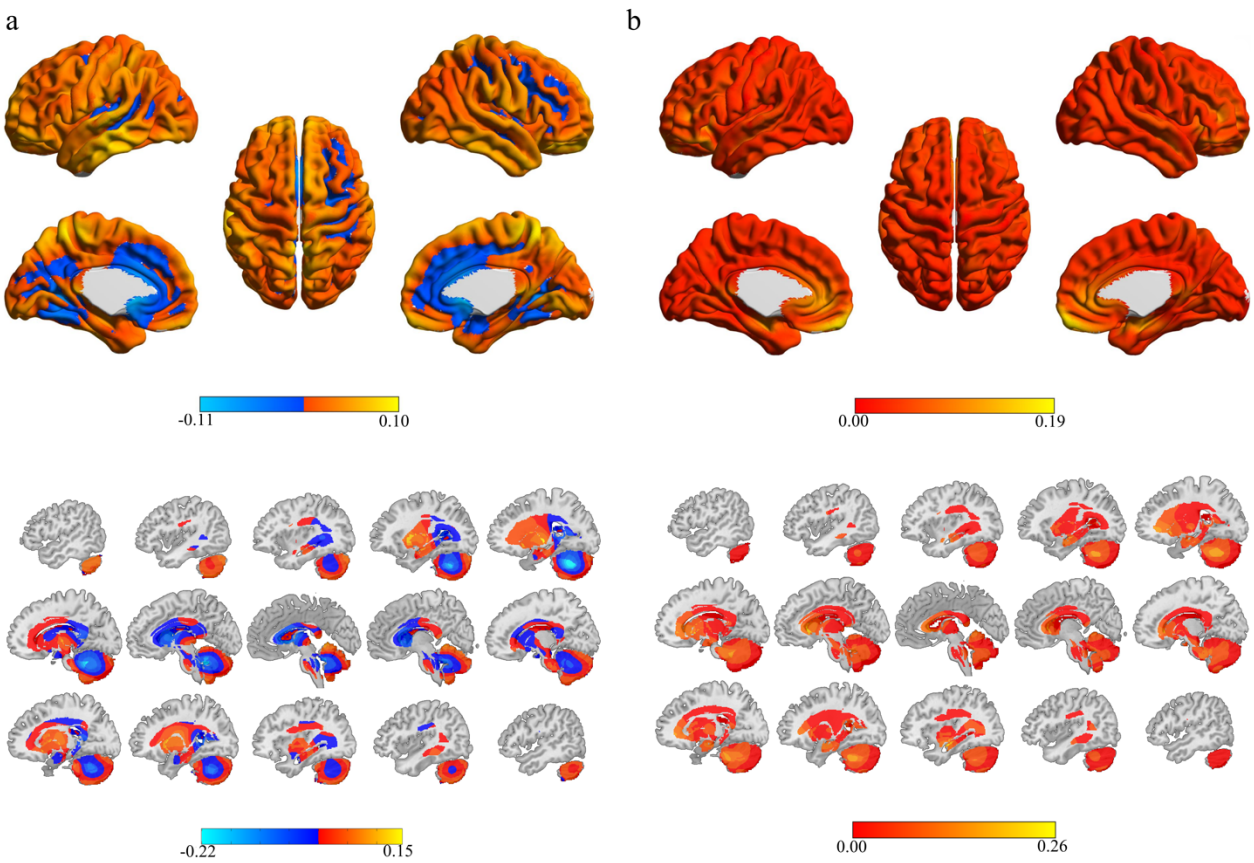
Supplementary Figure 7. Representational Similarity Analysis (RSA) results of SAHSU dataset (independent from discovery cohorts). Bar plots show representational similarity between neuroanatomical features (VAE's, CVAE's shared, and PD-specific) and patient properties. (* $p <$ .05, ** $p <$ .001, *** $p <$.0001).



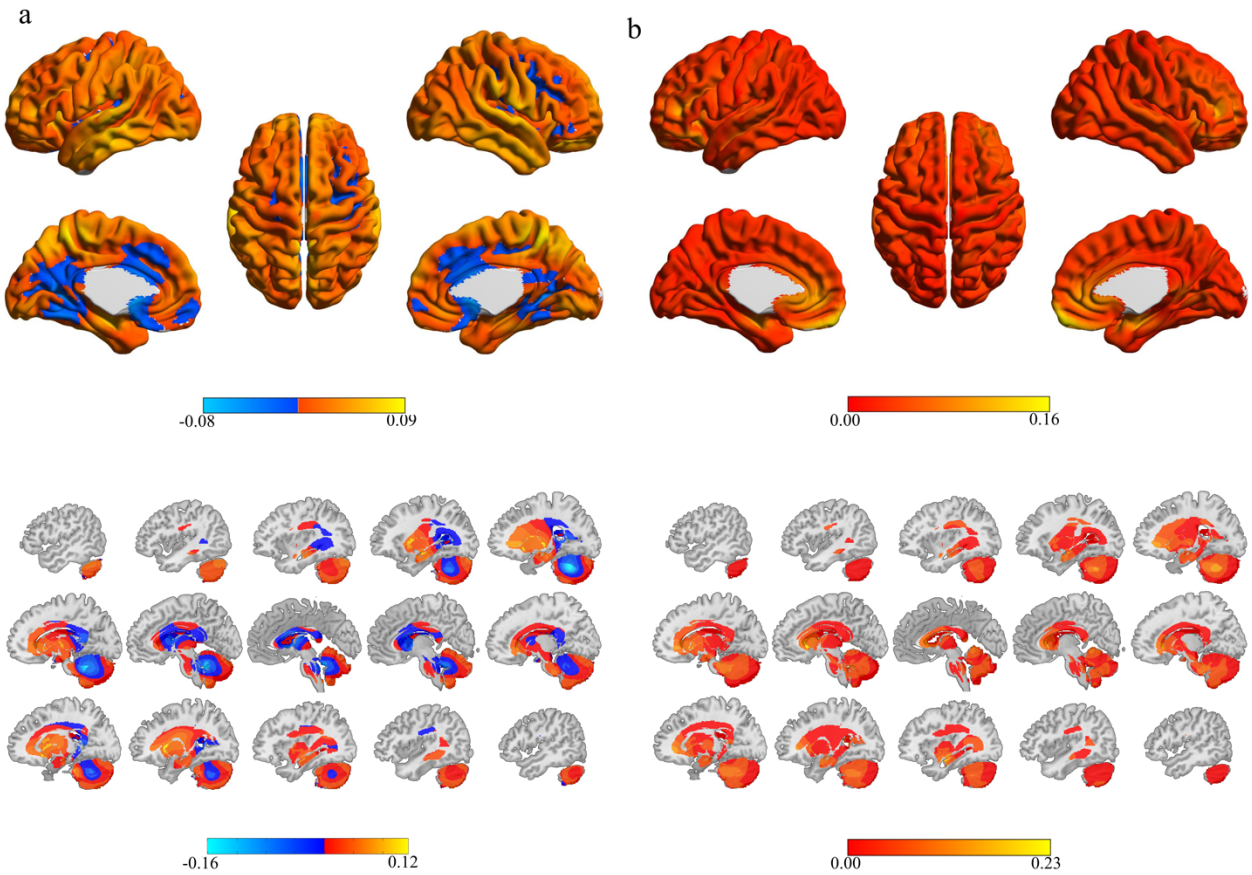
Supplementary Figure 8. The enrichment analysis results of the significant proteins that are associated with PD-specific features on cellular components (left) and molecular function (right). Statistical significance was defined as an FDR-corrected $P < 0.05$.



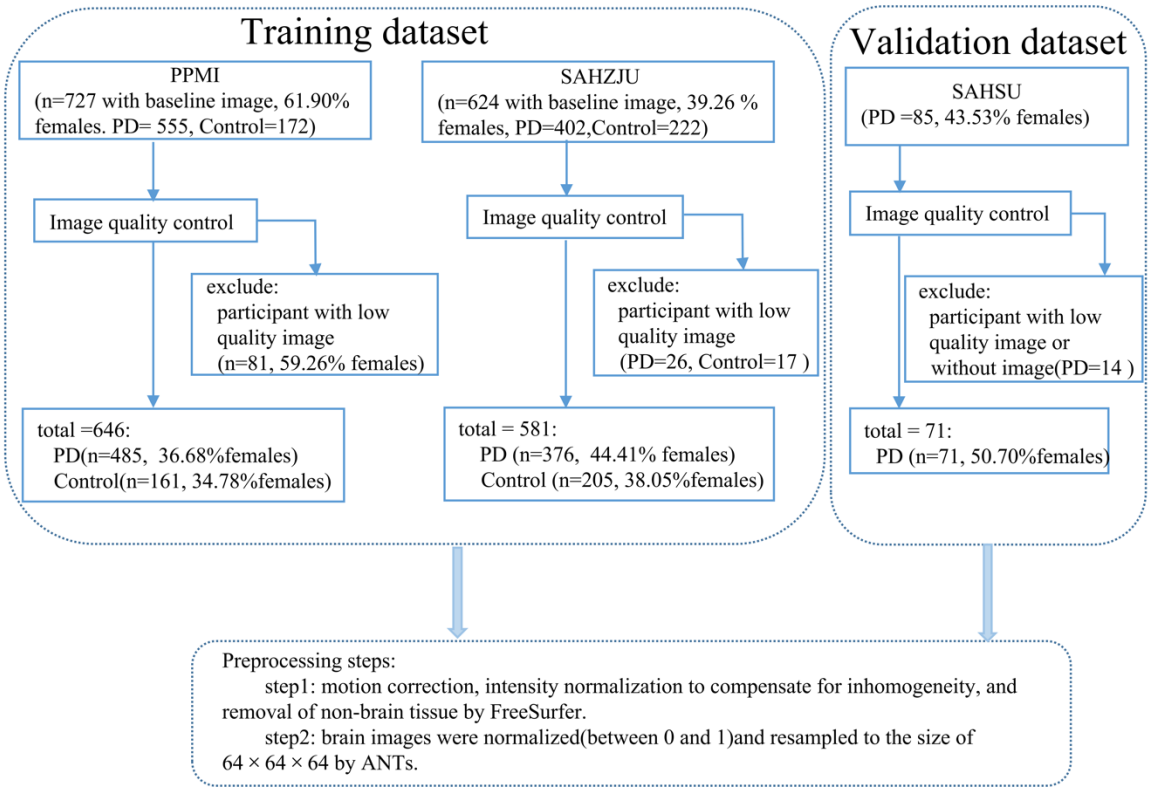
Supplementary Figure 9. The mean (a) and standard deviation (b) deformation field of synthesized case-control pairs for all PD patients in the PPMI dataset. On the top, the cortical region is displayed, whereas the underside showcases the subcortical, white matter, and cerebellar regions. Warm colors indicate decreased volumes, while cold colors denote increased volumes in synthetic PD brains when compared with synthetic control brains.



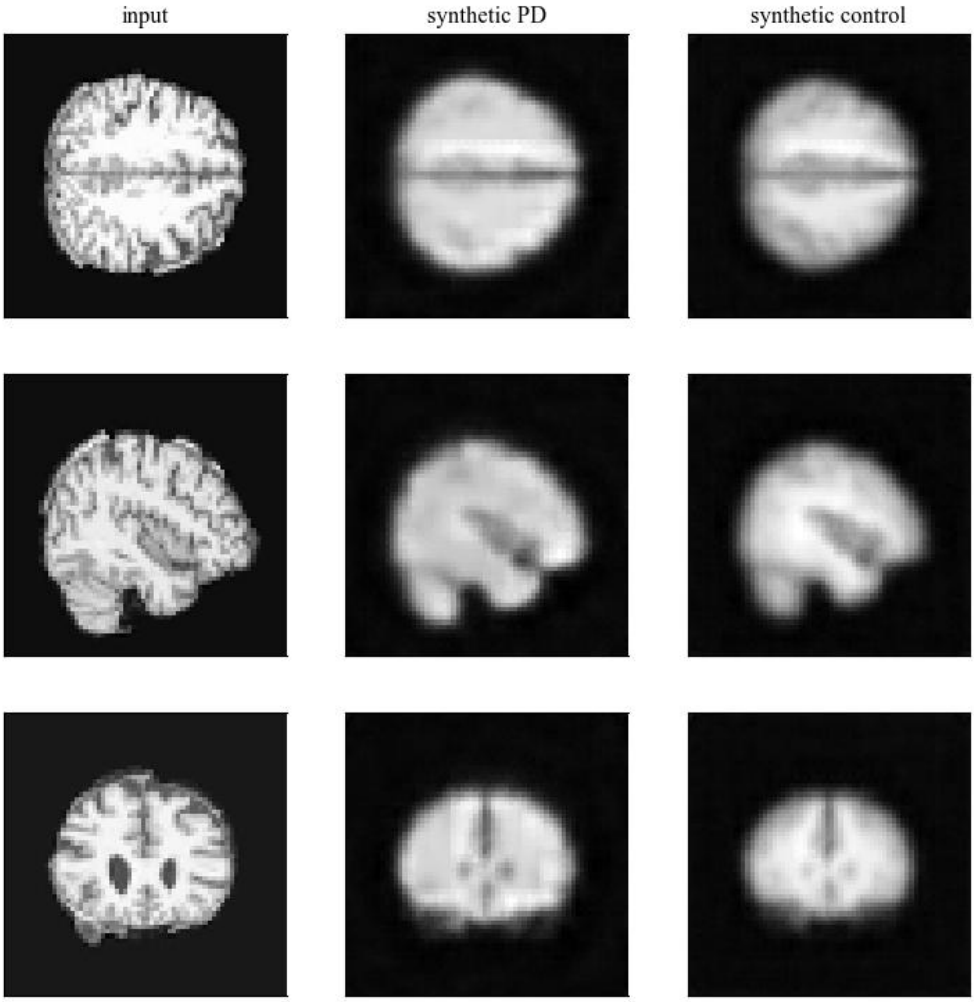
Supplementary Figure 10. The mean(a) and standard deviation (b) deformation field of synthesized case-control pairs for all PD patients in the SAHZJU dataset. On the top, the cortical region is displayed, whereas the underside showcases the subcortical, white matter, and cerebellar regions. Warm colors indicate decreased volumes, while cold colors denote increased volumes in synthetic PD brains when compared with synthetic control brains.



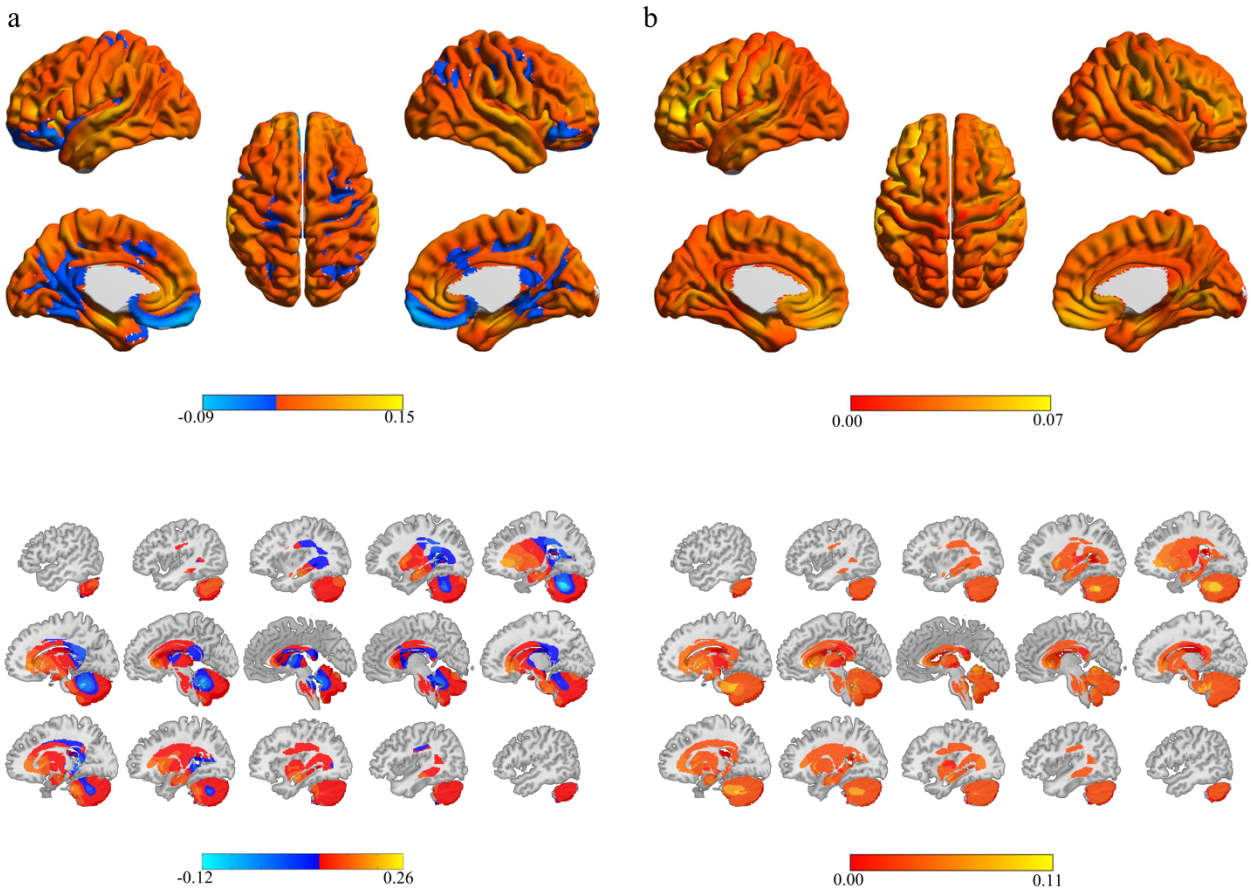
Supplementary Figure 11. The mean(a) and standard deviation (b) deformation field of synthesized case-control pairs for all PD patients in the SAHSU dataset. On the top, the cortical region is displayed, whereas the underside showcases the subcortical, white matter, and cerebellar regions. Warm colors indicate decreased volumes, while cold colors denote increased volumes in synthetic PD brains when compared with synthetic control brains.



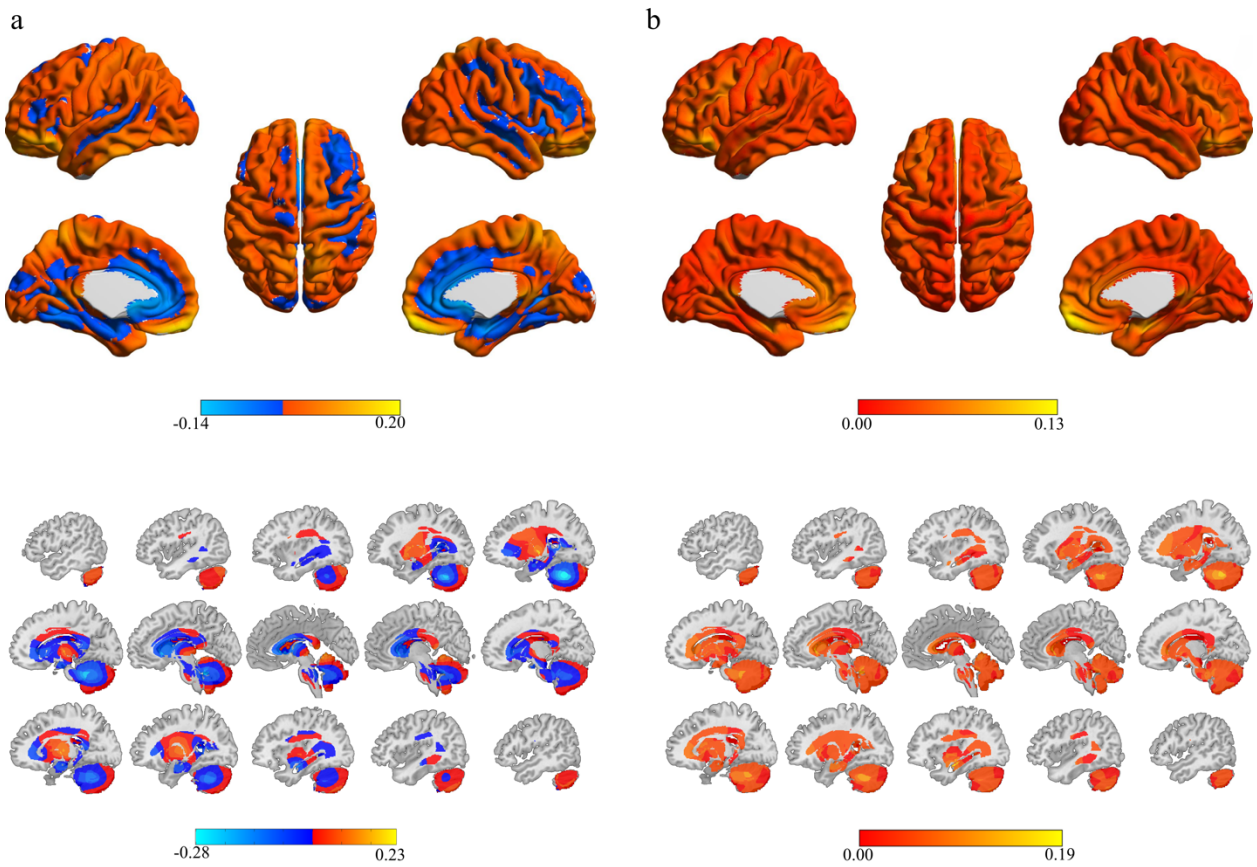
Supplementary Figure 12. Detailed flowchart outlining the participant selection process employed in the study. Images of substandard quality, specifically those with a CAT12 (https://neuro-jena.github.io/cat12-help/#intro) image quality rating below 75%, were excluded



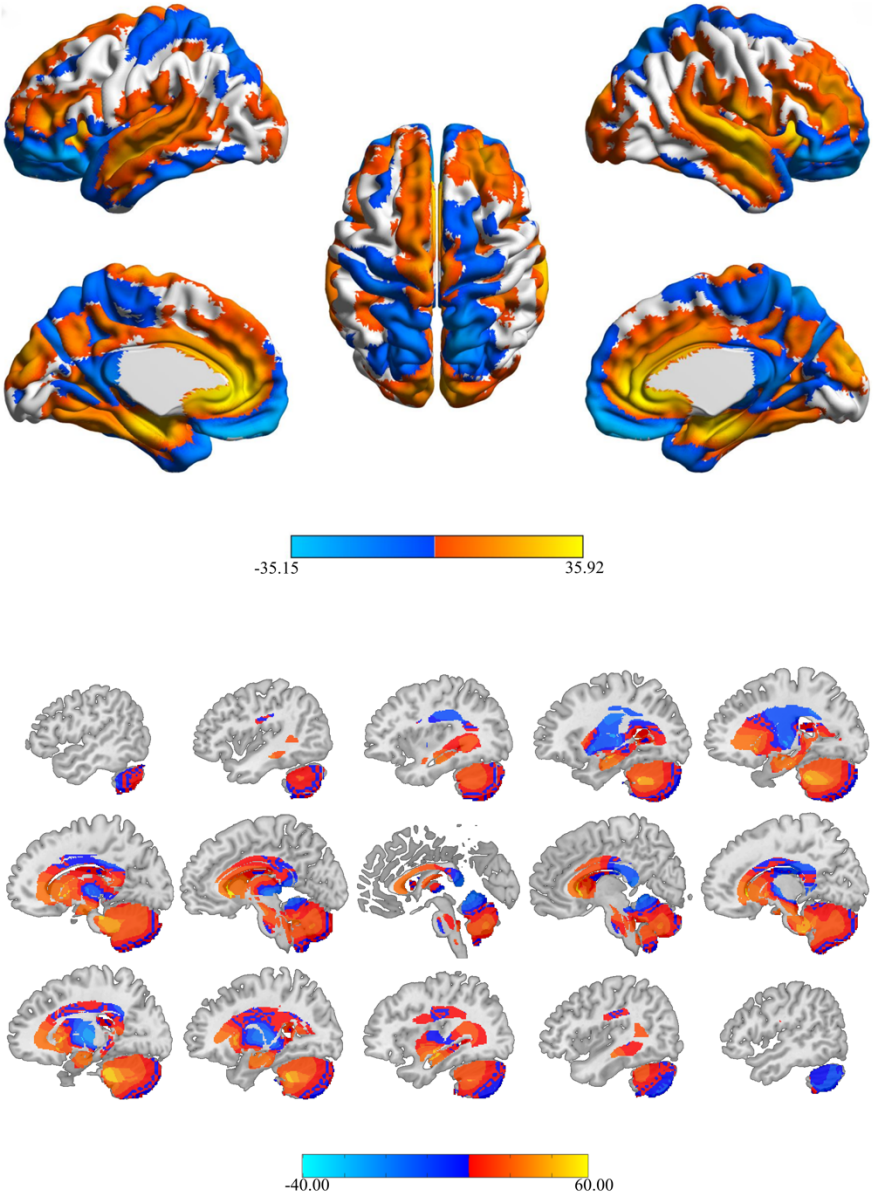
Supplementary Figure 13. Illustration of input images, synthetic PD brain images encompassing both PD-specific and shared variations, alongside their corresponding synthetic control brain images, which solely exhibit shared variations. This figure presents a representative example from a single subject, demonstrating the images in transverse, sagittal, and coronal planes.



Supplementary Figure 14. The mean(a) and standard deviation (b) deformation field of synthesized case-control pairs for all subgroup 1 patients in the PPMI dataset. On the top, the cortical region is displayed, whereas the underside showcases the subcortical, white matter, and cerebellar regions. Warm colors indicate decreased volumes, while cold colors denote increased volumes in synthetic PD brains when compared with synthetic control brains.



Supplementary Figure 15. The mean(a) and standard deviation (b) deformation field of synthesized case-control pairs for all subgroup 2 patients in the PPMI dataset. On the top, the cortical region is displayed, whereas the underside showcases the subcortical, white matter, and cerebellar regions. Warm colors indicate decreased volumes, while cold colors denote increased volumes in synthetic PD brains when compared with synthetic control brains.



Supplementary Figure 16. Voxel-wise statistical comparison (two-sided t-test) of deformation field between subgroup 1 and subgroup 2. The upper side displays the cortical region, whereas the lower side exhibits the subcortical, white matter, and cerebellar regions. Bonferroni correction for multiple comparisons with *p*-value threshold of 0.05 was applied.

Supplementary reference

1. Mak, E. *et al.* Baseline and longitudinal grey matter changes in newly diagnosed Parkinson's disease: ICICLE-PD study. *Brain* **138**, 2974–2986 (2015).
2. Becker, S. *et al.* Association of Hippocampal Subfields, CSF Biomarkers, and Cognition in Patients With Parkinson Disease Without Dementia. *Neurology* **96**, e904--e915 (2021).
3. Burciu, R. G. & Vaillancourt, D. E. Imaging of Motor Cortex Physiology in Parkinson's Disease. *Movement Disorders* **33**, 1688–1699 (2018).
4. Guerra, A. *et al.* Driving motor cortex oscillations modulates bradykinesia in Parkinson's disease. *Brain* **145**, 224–236 (2022).
5. Wang, L. *et al.* Association of Cortical and Subcortical Microstructure with Clinical Progression and Fluid Biomarkers in Patients With Parkinson Disease. *Neurology* (2023) doi:10.1212/WNL.0000000000207408.
6. Wang, L. *et al.* Association of Structural Measurements of Brain Reserve With Motor Progression in Patients With Parkinson Disease. *Neurology* **99**, E977–E988 (2022).
7. Abbasi, N. *et al.* Predicting severity and prognosis in Parkinson's disease from brain microstructure and connectivity. *Neuroimage Clin* **25**, 102111 (2020).
8. Joyce, J. M. Kullback-Leibler Divergence. in *International Encyclopedia of Statistical Science* (ed. Lovric, M.) 720–722 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011). doi:10.1007/978-3-642-04898-2_327.
9. Kim, H. & Mnih, A. Disentangling by Factorising. in *Proceedings of the 35th International Conference on Machine Learning* (eds. Dy, J. & Krause, A.) vol. 80 2649–2658 (PMLR, 2018).