A Mechanistic Foundation Model for Human Brain Dynamics Across Lifespan

Jiaqi Wu **Shuheng Cao** **Tingting Dan**
Ziquan Wei **Zhixuan Zhou** **Guorong Wu**

Abstract

Foundation models for functional magnetic resonance imaging (fMRI) show great potential in recognizing cognitive states and supporting disease diagnosis. However, current approaches predominantly focus on modeling extrinsic functional fluctuations, with limited attention to the holistic modeling of intrinsic brain state dynamics. This gap constrains both interpretability and generalization to neuroscience-driven applications. To address this limitation, we introduce BrainMech, a Koopman-structured fMRI foundation model designed to learn an analyzable stochastic latent dynamical representation of fMRI signals, enabling explicit analysis of latent spectra, temporal evolution, and region-level dynamical patterns. The architecture integrates a population-level Koopman operator, which maps functional dynamics into a linear latent state space, with a subject-specific stochastic component to account for individual variability. By leveraging intrinsic spatiotemporal dependencies, BrainMech learns stable spatiotemporal latent dynamics that transfer across heterogeneous lifespan cohorts and downstream neuroimaging tasks. Experimental results on multiple downstream neuroimaging tasks (including $>$130K fMRI scans) show that BrainMech improves predictive performance under the evaluated protocols while providing spectral, temporal, and region-level analyses of the learned latent dynamics. Code is provided in Supplementary Materials.

1 Introduction

Understanding brain dynamics from functional Magnetic Resonance Imaging (fMRI) data is critical in computational neuroscience, yet remains challenging due to the complex mapping between high-dimensional neuroimaging signals and heterogeneous behavioral phenotypes across individuals. In parallel, sample sizes in clinical neuroimaging studies are often limited (Marek et al., 2022). For example, many pilot studies may recruit only a small number of patients. This data scarcity makes it difficult to reliably test hypotheses or assess the diagnostic utility of fMRI using conventional learning-based approaches. In this context, there is a critical need for foundation models trained on large-scale public datasets to facilitate current neuroscience studies and clinical applications by providing a generalizable understanding of functional dynamics for human brains.

Recent advances in foundation models have substantially promoted the modeling of brain function using an unprecedented amount of existing fMRI data (Caro et al., 2023; Dong et al., 2024; Yang, 2024). However, existing approaches predominantly cast brain dynamics as time series modeling problems, providing limited access to explicit dynamical quantities such as latent spectra, time constants, and region-level mode loadings that can be compared across cohorts and clinical groups. This disconnect undermines predictive accuracy for phenotypical traits while also limiting model interpretability, since the learned fMRI representations lack an explicit dynamical parameterization that supports spectral, temporal, and region-level analyses of learned fMRI representations.

Preprint.

To address this limitation, we propose a Koopman-structured stochastic latent dynamics foundation model that represents fMRI signals in an analyzable latent state space. Since human brain function operates in a nonlinear manner, the Koopman operator (Koopman, 1931) is employed to transform nonlinear dynamical systems into linear latent spaces, thereby simplifying the learning of brain-state dynamics. We specifically devise a diagonal-plus-low-rank (DPLR) decomposition on the Koopman operator to enable computationally efficient linearization. In addition, we model the temporal evolution of latent brain states with a stochastic state-space formulation. The deterministic drift captures structured latent evolution, while the diffusion term accounts for intrinsic uncertainty and inter-subject heterogeneity. The diffusion term in SDE explicitly captures uncertainty and latent, unmodeled dynamics, enabling principled uncertainty quantification within the dynamical system. To that end, the learned diffusion process captures time-varying uncertainty in latent neural trajectories, improving robustness over conventional approaches that treat the variation as unstructured noise.

Furthermore, we train BRAINMECH across cohorts spanning childhood to late adulthood and use age as a continuous control covariate to reduce cohort heterogeneity. We do not claim to identify systematic aging laws; instead, we evaluate whether a unified latent dynamics model transfers across age-diverse cohorts (Fig. 1).

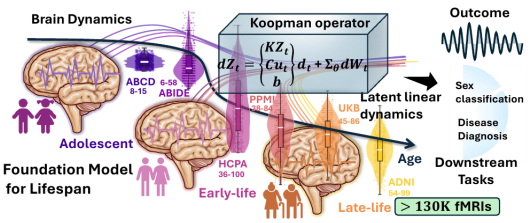

Figure 1: BRAINMECH learns Koopman-structured stochastic latent dynamics from large-scale lifespan fMRI data. Explicit latent dynamics modeling supports downstream prediction and enables spectral, temporal, and region-level analyses.

Together, BRAINMECH combines explicit latent dynamics with the scalability of modern foundation models. A distinctive advantage is analyzability: the learned latent operator can be empirically examined through spectra, eigenmodes, rollout behavior, and region-level patterns, enabling comparisons across cohorts and clinical groups. Through self-supervised learning on over 33,000 hours of resting-state fMRI data, from ages 8 to 100 years across multiple cohorts, our BRAINMECH achieves strong downstream performance on sex classification and disease diagnosis while providing analyzable latent dynamics.

2 Related Work

Brain dynamics modeling. Classical approaches include dynamic causal modeling (Friston et al., 2003) and data-driven decompositions such as tensorial/group ICA (Beckmann and Smith, 2005), as well as time-varying (dynamic) functional connectivity analyses (Hutchison et al., 2013; Preti et al., 2017; Lurie et al., 2020). Deep learning methods, including RNNs/LSTMs (Dvornek et al., 2017) and continuous-time models based on neural ODEs, have been applied to fMRI, but often provide limited access to explicit spectral, stability, and region-level dynamical quantities that can be compared across cohorts and clinical groups (Han et al., 2024).

Foundation models for learning functional dynamics. Recent foundation models have demonstrated strong representation learning from large-scale brain imaging data. BrainLM (Caro et al., 2023) uses transformer-based masked prediction on fMRI recordings but models signals as discrete tokens without a continuous-time structure. Brain-JEPA (Dong et al., 2024) learns latent representations through future state forecasting but relies on implicit objectives without explicit dynamics. BrainMass (Yang, 2024) applies masked auto-encoding to functional connectivity matrices but operates on static representations. BrainHarmonix (Dong et al., 2025) explores multimodal brain foundation modeling by unifying structural MRI morphology and fMRI functional dynamics into compact one-dimensional brain tokens. These approaches share common limitations: they treat brain dynamics as sequence or matrix prediction problems without explicit continuous-time dynamical structure. Although graph-based approaches such as BrainGNN (Li et al., 2021) were not originally designed as foundation models, their success underscores the importance of incorporating connectivity topology when modeling functional dynamics. However, the inductive biases implied by connectivity topology remain underexplored. We address these challenges by integrating Koopman operator theory with graph-based representations and stochastic differential equations, as described below.

Koopman operator for brain dynamics. The Koopman operator framework provides a principled approach to analyzing nonlinear dynamical systems by lifting them into an infinite-dimensional linear space (Koopman, 1931; Mezić, 2005). For a deterministic nonlinear system $\dot{z} = f(z)$ with state $z \in \mathbb{R}^d$, where d represents latent dimension, the Koopman operator $\mathcal{K}^t$ acts on observables $g : \mathbb{R}^d \to \mathbb{R}$ via

$$(\mathcal{K}^t g)(z) = g(\phi^t(z)), \tag{1}$$

where ϕ^t is the state transition function. The key insight is that while the original dynamics f are nonlinear, the Koopman operator is *linear* in the space of observables, enabling efficient system identification and behavior prediction.

On the other hand, neural dynamics are inherently noisy and subject to complex physiological fluctuations, rendering strict deterministic models insufficient. In this context, for stochastic systems, we apply the *stochastic Koopman semigroup* (Klus et al., 2020), which acts on observables via conditional expectations over stochastic trajectories.

3 Method

We first introduce the Koopman-structured stochastic latent dynamics used by BRAINMECH (Sec. 3.1). We then describe the network architecture that instantiates these dynamics, including the topology-aware spatiotemporal embedding network (TASEN), the DPLR-Koopman core, stochastic decoding, and stability-regularized self-supervised objectives (Secs. 3.2–3.3). Standard stochastic Koopman and SDE derivations are provided in Appendix B.

Let $x_\tau \in \mathbb{R}^N$ denote the observed regional blood-oxygen-level-dependent (BOLD) signal from fMRI at the τ-th time point, where N is the number of brain regions, for $\tau = 1, \ldots, T$. Although neural activity evolves in continuous time, fMRI provides **discrete** observations $\{x_\tau\}_{\tau=1}^T$ obtained by sampling the underlying process at regular acquisition intervals. We denote **continuous** sampling times by $\{t_\tau\}$ and assume approximately uniform sampling, i.e., $t_\tau = \tau \Delta t$. Here, Δt acts as a time-scale factor [1]. We then construct the functional brain network by calculating Pearson's correlation coefficients between regional time series, yielding the functional connectivity (FC) matrix A.

3.1 Koopman-Structured Stochastic Latent Dynamics

BRAINMECH uses a Koopman-inspired stochastic latent state-space model rather than claiming to recover exact finite-dimensional Koopman observables. Given an fMRI sequence $\{x_\tau\}_{\tau=1}^T$ and its functional connectivity matrix A, TASEN maps the observed BOLD signals into latent states $z_\tau \in \mathbb{R}^d$. We model the continuous-time evolution of the corresponding latent process $Z_t \in \mathbb{R}^d$ as

$$dZ_t \;=\; \big(K Z_t + C\, u_t + \beta\big)\, dt \;+\; \Sigma_\theta\big(Z_t, u_t\big)\, dW_t, \tag{2}$$

where $K \in \mathbb{R}^{d \times d}$ is a learnable continuous-time latent drift/generator, $u_t \in \mathbb{R}^{d_c}$ is a continuous-time control process, $C \in \mathbb{R}^{d \times d_c}$ maps the control signal into the latent dynamics, $\beta \in \mathbb{R}^d$ is a bias vector, W_t is Brownian motion, and $\Sigma_\theta(\cdot, \cdot)$ parameterizes stochastic diffusion. For a fixed control signal over $[t_\tau, t_{\tau+1})$, the corresponding sampled-time deterministic transition is $\Phi_{\Delta t} = \exp(\Delta t K)$, while Euler–Maruyama uses $I + \Delta t K$ only as a first-order numerical approximation for training on the fMRI sampling grid. Therefore, stability and spectral analyses of K are interpreted through the real parts $\mathrm{Re}(\lambda_i(K))$; spectral-radius criteria apply only to the sampled transition $\Phi_{\Delta t}$. A detailed discussion is provided in Appendix B.6. The central inductive bias is that nonlinear fMRI dynamics can be represented by approximately linear evolution in a learned latent coordinate system, while the diffusion term captures residual stochasticity and subject-level variability.

For an SDE-driven Markov process, the associated stochastic Koopman semigroup acts on observables through conditional expectation,

$$(\mathcal{K}^t g)(z) \;:=\; \mathbb{E}\big[g(Z_t) \mid Z_0 = z\big], \tag{3}$$

for any observable $g : \mathbb{R}^d \to \mathbb{R}$. We therefore use the term **Koopman-structured** to describe the operator-theoretic inductive bias imposed on the learned latent dynamics, not to assert that the encoder

[1] We set $\Delta t = 1$ and report normalized frame-time dynamics; TR calibration is discussed in Sec. F.

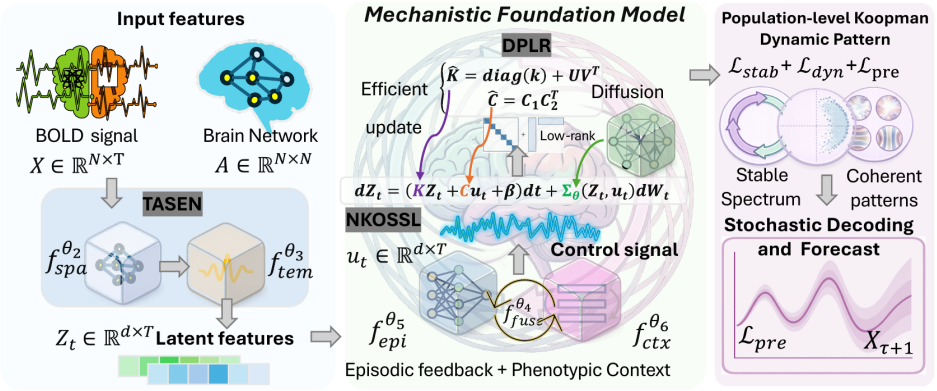

Figure 2: BRAINMECH includes three learning components. *Left.* TASEN maps BOLD signals and functional connectivity into latent features. *Middle.* A Koopman-structured stochastic latent dynamics module models state transitions through linear drift, control, and diffusion. *Right.* Stability-regularized self-supervised objectives support forecasting and downstream adaptation.

exactly recovers Koopman eigenfunctions. Under standard Lipschitz and growth conditions, Eq. 2 induces a well-defined stochastic process and a valid stochastic Koopman semigroup; these standard results and proofs are provided in Appendix B.

Our model-specific contributions are the learnable control-aware latent SDE in Eq. 2, the diagonal-plus-low-rank (DPLR) parameterization of K for scalable brain-network modeling, and a stability regularizer motivated by mean-square dissipativity of the latent process. These components make the learned dynamics analyzable through spectra, eigenmodes, and mediation effects.

3.2 Network architecture: Mechanistic Foundation Model of Functional Dynamics

Latent dynamics instantiation. Eq. (2) defines latent stochastic dynamics with drift (K, C, β) and diffusion Σ_θ, and Eq. (4) connects latent states to fMRI observations. In our implementation, TASEN encodes $(x_{1:T}, A) \mapsto z_{1:T}$ (Sec. 3.2.1), the control head maps latent trajectories and context variables into $u_{1:T}$ (Eq. 6), and the low-rank parameterizations realize (K, C) (Eqs. 9 and 10). Euler–Maruyama discretization yields a tractable Gaussian transition density $p(z_{\tau+1} \mid z_\tau, u_\tau)$, enabling likelihood-based dynamics learning; the full derivation is provided in Appendix B.

Observation model. We connect latent states z_τ to discrete fMRI observations $\{x_\tau\}$ via a stochastic measurement model:

$$x_\tau \mid z_\tau \sim \mathcal{N}\Big(\mu_{\theta_1}(z_\tau),\ \sigma^2_{\theta_1}(z_\tau) I\Big), \quad z_\tau := Z_{t_\tau}. \tag{4}$$

where the stochastic decoder outputs a mean $\mu_{\theta_1}(z_\tau)$ and a variance $\sigma^2_{\theta_1}(z_\tau)$ in the observation space.

3.2.1 Spatial Brain Mapping: Topology-Aware Spatiotemporal Embedding Network

The spatial brain mapping module TASEN projects discrete fMRI observations onto the latent dynamical manifold, decomposing the encoding process into spatial feature extraction and temporal integration. Given the inputs $\{x_\tau\}_{\tau=1}^T$, TASEN computes embeddings through a graph spatial encoder followed by a temporal model:

$$z_\tau = f_{\text{tem}}^{\theta_3}(f_{\text{spa}}^{\theta_2}(x_\tau; A)), \tag{5}$$

where the spatial encoder $f_{\text{spa}}^{\theta_2}$ (e.g., ChebNet (Defferrard et al., 2016) or GAT (Veličković et al., 2018)) leverages the functional connectivity matrix A to aggregate local topological information. Subsequently, the temporal encoder $f_{\text{tem}}^{\theta_3}$, implemented as a Transformer-based module, captures long-range dependencies across the sequence of spatial embeddings (as shown in Fig. 2, left).

We construct the control signal by combining short-range trajectory history with global subject-level context (Fig. 2, middle):

$$u_\tau \;=\; f_{\text{fuse}}^{\theta_4}\Big(f_{\text{epi}}^{\theta_5}(z_{<\tau}),\; f_{\text{ctx}}^{\theta_6}(v)\Big), \tag{6}$$

where v denotes phenotypic covariates that condition the dynamics. We instantiate v with age, as it constitutes a major variation in functional brain organization throughout development and aging (Betzel et al., 2014). The episodic encoder $f_{\text{epi}}^{\theta_5}$ summarizes the recent latent trajectory before time τ, capturing transient dynamical state and short-term history. In contrast, the contextual encoder $f_{\text{ctx}}^{\theta_6}$ uses the covariates v to estimate subject-level factors. The fused control u_τ therefore modulates the latent dynamics using both local temporal evidence and global contextual information.

For discrete fMRI samples, we use a piecewise-constant control interpolation, $u_t = u_\tau, t \in [t_\tau, t_{\tau+1})$, so the continuous-time dynamics evolve under a fixed control between adjacent frames.

3.2.2 Parameterization of Functional Dynamic: Euler-Maruyama on Sampling Times

To learn from discrete sequences, we discretize Eq. (2) at $\{t_\tau\}$ using Euler-Maruyama:

$$z_{\tau+1} = z_\tau + \big(Kz_\tau + Cu_\tau + \beta\big)\Delta t + \Sigma_\theta(z_\tau, u_\tau)\sqrt{\Delta t}\,\xi_\tau, \quad \xi_\tau \sim \mathcal{N}(0, I). \tag{7}$$

With normalized frame time $\Delta t = 1$, Eq. 7 defines the Euler–Maruyama training transition used to evaluate the one-step likelihood on discrete fMRI samples. This discretization does not change the stability notion of the underlying model: K remains a continuous-time drift/generator, while $I + \Delta t K$ is only the first-order Euler approximation to the sampled transition $\exp(\Delta t K)$.

This discretization gives the Gaussian one-step transition density

$$p(z_{\tau+1} \mid z_\tau, u_\tau) = \mathcal{N}\Big(z_\tau + (Kz_\tau + Cu_\tau + \beta)\Delta t,\; \Delta t\, Q_\theta(z_\tau, u_\tau)\Big), \tag{8}$$

where $Q_\theta(z_\tau, u_\tau) = \Sigma_\theta(z_\tau, u_\tau)\Sigma_\theta(z_\tau, u_\tau)^\top$. We use this transition density to define the dynamics-consistency loss $\mathcal{L}_{\text{dyn}}$ as a negative log-likelihood. The derivation is provided in Appendix B.

3.2.3 DPLR-Koopman Core: Structured Parameterization for Efficiency

The DPLR-Koopman core is our efficient parameterization of the latent drift operator, combining Koopman-structured stochastic dynamics with a diagonal-plus-low-rank (DPLR) representation. To reduce parameter complexity, we parameterize the drift matrix with a DPLR form:

$$\hat{K} \;=\; \operatorname{diag}(k) + UV^\top, \tag{9}$$

where $k \in \mathbb{R}^d$ specifies the diagonal component $\operatorname{diag}(k) \in \mathbb{R}^{d\times d}$, and $U, V \in \mathbb{R}^{d\times r}$ with $r \ll d$ form the low-rank correction. This parameterization reduces the parameter complexity from $\mathcal{O}(d^2)$ to $\mathcal{O}(dr)$. Equivalently, $UV^\top$ can be interpreted as a rank-r low-rank correction to the residual $R := \hat{K} - \operatorname{diag}(k)$, i.e., $R \approx P_r \Sigma_r J_r^\top$ with $U := P_r\Sigma_r^{1/2}$ and $V := J_r\Sigma_r^{1/2}$.

For stable, data-agnostic initialization, we initialize the diagonal drift k with small negative values so that $\operatorname{diag}(k)$ is initially contractive, while U and V are initialized with small zero-mean random entries so that $UV^\top$ starts as a small perturbation. During training, the low-rank correction is amplified only when supported by the data.

Low-rank control. The control signal contributes the term Cu_t with $C \in \mathbb{R}^{d\times d_c}$, mapping the d_c-dimensional continuous-time control process into the d-dimensional latent dynamics. Directly learning C requires $\mathcal{O}(d\,d_c)$ parameters. To reduce complexity and encourage a shared low-dimensional control subspace, we factorize C at rank $r_c = \min(r, d_c)$:

$$\hat{C} \;=\; C_1 C_2^\top, \qquad C_1 \in \mathbb{R}^{d\times r_c},\;\; C_2 \in \mathbb{R}^{d_c\times r_c}, \tag{10}$$

so that $\operatorname{rank}(\hat{C}) \le r_c$ and the parameter count becomes $\mathcal{O}((d + d_c)r_c)$. Equivalently, $C_2^\top u_t \in \mathbb{R}^{r_c}$ produces compact control coordinates and C_1 projects them into latent-state directions, meaning the control influences the dynamics through r_c shared latent modes. The factors C_1 and C_2 are initialized with small random values and are optimized end-to-end during training. See Algorithm 1 in the appendix for a high-level summary.

3.3 Loss Function: Stability-Regularized Dynamics Learning

To learn Koopman-governed latent dynamics from discrete fMRI sequences, we optimize objectives that directly supervise Koopman transitions and their observable forecasting behavior, as shown in Fig. 2-right.

Prediction loss ($\mathcal{L}_{\text{pre}}$). Let $\hat{x}_{\tau+1}$ be the one-step forecast decoded from the Koopman-predicted latent state $\hat{z}_{\tau+1}$. We minimize the one-step forecasting negative log-likelihood:

$$\mathcal{L}_{\text{pre}} = -\frac{1}{T-1}\sum_{\tau=1}^{T-1} \log p\big(x_{\tau+1} \mid \hat{z}_{\tau+1}\big),$$

with Gaussian output parameterization $p(x_{\tau+1} \mid \hat{z}_{\tau+1}) = \mathcal{N}(\mu_{\theta_1}(\hat{z}_{\tau+1}), \sigma^2_{\theta_1}(\hat{z}_{\tau+1})I)$.

Dynamics consistency ($\mathcal{L}_{\text{dyn}}$). We require the Koopman-predicted next latent to match the temporal encoder's next-step estimate via mean-squared error $\mathcal{L}_{\text{dyn}} = \frac{1}{T-1}\sum_{\tau} \|\hat{z}_{\tau+1} - z_{\tau+1}\|_2^2$, where $\hat{z}_{\tau+1} = z_\tau + (Kz_\tau + Cu_\tau + b)\,\Delta t$ is the Euler step in Eq. 7 of the DPLR-Koopman drift.

Spectral stability regularization ($\mathcal{L}_{\text{stab}}$). For discrete-time stability of the Euler iteration we penalize the spectral radius of K exceeding a target $\rho < 1$: $\mathcal{L}_{\text{stab}} = \big(\,\text{ReLU}(\,|\lambda(K)|_{\max} - \rho\,)\big)^2$, with $\rho = 0.95$ in our experiments. We additionally apply a hard projection $k \leftarrow s\,k$, $U, V \leftarrow \sqrt{s}\,U, \sqrt{s}\,V$ with $s = \rho/|\lambda(K)|_{\max}$ whenever the bound is violated during training.

Total objective. The final loss function is a weighted sum of the three terms:

$$\mathcal{L} = \mathcal{L}_{\text{pre}} + \lambda_1 \mathcal{L}_{\text{dyn}} + \lambda_2 \mathcal{L}_{\text{stab}},$$

where λ_1 and λ_2 are hyperparameters balancing dynamical consistency and system stability.

4 Experiments

4.1 Datasets and Setup

We evaluate BRAINMECH on six large-scale fMRI cohorts spanning development, aging, population neuroscience, and neurological disorders (age distributions in Appendix D.1). The Adolescent Brain Cognitive Development study (**ABCD**; (Casey et al., 2018)) provides a large longitudinal neurodevelopmental cohort ($N = 49{,}147$ fMRIs) and is used primarily for pretraining and sex classification. The Human Connectome Project Aging (**HCPA**; (Elam et al., 2021)) contains high-quality fMRI scans from $N = 720$ middle-aged and older adults and is used to evaluate transfer to aging-related brain dynamics and sex prediction in a non-clinical population. The UK Biobank (**UKB**; (Miller et al., 2016)) provides population-level fMRI data ($N = 76{,}603$ fMRIs) with rich phenotypic annotations, serving as both diverse pretraining data and a benchmark for demographic prediction. For clinical evaluation, the Autism Brain Imaging Data Exchange (**ABIDE**; (Heinsfeld et al., 2018)) includes multi-site fMRI data from 2,072 young adults and is used for binary typical-control versus ASD classification; the Alzheimer's Disease Neuroimaging Initiative (**ADNI**; (Petersen et al., 2010)) contains 2,126 older-adult fMRI scans and is used for cognitively normal versus mild cognitive impairment classification; and the Parkinson's Progression Markers Initiative (**PPMI**; (Marek et al., 2018)) includes 418 unique subjects across normal control, SWEDD, prodromal, and Parkinson's disease groups, which we formulate as a four-class classification task.

Experimental setup. BRAINMECH is pretrained on a pooled, lifespan cohort combining existing fMRI data of healthy subjects from HCPA, UKB, and ABCD. For each dataset, we use the publicly released preprocessed fMRI (Xu et al., 2023; Dan et al., 2024) and extract fixed-length segments of length $T = 300$ using a sliding window with stride $S = 300$ (see Appendix D.2). All experiments default to the Automated Anatomical Labeling (AAL-116) atlas (Tzourio-Mazoyer et al., 2002), yielding 116 brain regions. For each window, we construct both the BOLD time series $x_{1:T} \in \mathbb{R}^{T\times 116}$ and its FC matrix $A \in \mathbb{R}^{116\times 116}$ as input. To bridge the developmental gap across the lifespan, we incorporate age labels to condition the model's dynamics on developmental stage. Specifically, we normalize each subject's age to $[0, 1]$ based on the full age range (8-100 years) and embed it as the first dimension of context vectors v (i.e., external control input in Eq. 6), which are then processed by the context encoder $f^{\theta_6}_{\text{ctx}}$ to produce age-conditioned context embeddings, which are fused with episodic feedback via $f^{\theta_4}_{\text{fuse}}$ to generate control signals u_τ that modulate the Koopman dynamics.

Table 1: Disease diagnosis performance. Mean ± std across folds. Bold denotes best; * denotes significant improvement over competing methods ($p < 0.05$).

Method	ABIDE		ADNI		PPMI	
	ACC	F1	ACC	F1	ACC	F1
BrainGNN	62.83±2.77	61.73±3.83	83.48±6.99	78.69±8.25	76.92±15.60	75.24±16.43
BrainLM	53.25±4.00	51.51±5.67	83.30±4.71	75.79±6.63	50.53±19.59	45.83±23.41
BrainMass-SVM	62.44±2.71	62.51±2.67	82.96±5.02	75.32±7.06	63.55±14.00	55.67±16.50
BrainMass-MLP	67.22±3.88	66.97±4.04	83.70±6.02	77.59±8.41	67.38±10.97	61.37±13.41
Brain-JEPA	63.84±0.82	61.53±1.35	76.84±1.00	**86.32±0.50**	82.09±9.58	77.16±12.34
BrainHarmonix	55.96±3.68	52.07±4.15	76.75±2.97	44.04±0.60	58.33±3.41	24.19±3.58
BRAINMECH (Ours)	**72.83*±6.61**	**72.70*±6.51**	**87.85*±2.87**	73.22±7.80	**84.07±29.55**	**87.63±22.34**

Table 2: Biological sex classification accuracy across four cohorts. Results are mean ± standard deviation across folds. Bold denotes the best result.

Method	HCPA	ADNI	ABIDE	PPMI
BrainLM	40.68±4.03	41.72±7.19	78.59±4.98	42.03±9.21
BrainMass-SVM	69.44±1.69	45.40±12.80	73.84±3.49	48.83±6.27
BrainMass-MLP	69.93±1.64	65.74±7.66	74.73±4.20	67.73±12.60
Brain-JEPA	43.53±0.65	62.34±6.51	78.11±6.41	48.46±7.09
BrainHarmonix	54.65±1.28	52.35±5.65	78.18±2.74	57.14±4.39
BRAINMECH (Ours)	**77.88±2.81**	**66.24±2.04**	**82.21±1.88**	**68.48±4.80**

Pretraining optimizes a composite objective that combines forecasting (see section 3.3), along with dynamics consistency and stability regularization over resting-state (rs)-fMRI segments. The graph-based spatial encoder $f_{\text{spa}}^{\theta_2}$ consumes BOLD signals and FC matrices, the temporal module $f_{\text{tem}}^{\theta_3}$ summarizes dynamics, and the DPLR-Koopman core imposes a structured stochastic latent transition while predicting next-step latent states under an uncertainty-aware decoder. We train BRAINMECH with AdamW (Loshchilov and Hutter, 2017), ReduceLROnPlateau scheduling; detailed hyperparameters are listed in Appendix D.2. During pretraining, BRAINMECH achieves an R^2 as high as 98.3% on held-out validation data, demonstrating strong self-supervised learning of functional dynamics. By doing so, BRAINMECH can learn a latent dynamical representation that exhibits interpretable and structured behavior across both task-agnostic and task-specific analyses.

4.2 Downstream Task Performance Comparisons

We pretrained BRAINMECH using a forecasting-based objective to emphasize latent dynamical prediction. BRAINMECH contains approximately 110M parameters with training hours ∼12 on 8 H100 NVL, making it substantially more compact than existing brain foundation models. For sex classification, we freeze the backbone and train only an MLP head on pooled latent features, where the MLP consists of fully-connected layers with LayerNorm, GELU activation, and dropout. For disease diagnosis, we perform full fine-tuning of the entire model with stratified cross-validation to capture intricate pathological patterns while ensuring comparability with previous work (see Appendix D.3 for details).

As shown in Tables 1 and 2, BRAINMECH achieves strong downstream performance on both sex prediction and disease diagnosis under the evaluated protocols, outperforming the compared graph-based baselines and recent fMRI foundation models. For sex classification, BRAINMECH consistently exceeds the best competing methods across all cohorts. For disease diagnosis, our BRAINMECH achieves 72.83% accuracy on ABIDE, improving over both graph-based and foundation model baselines. For ADNI, BRAINMECH delivers competitive performance while complementary models (e.g., Brain-JEPA). Finally, for PPMI, BRAINMECH reaches 84.07% accuracy and 87.63% F1, showing favorable performance among the evaluated approaches and suggesting that the learned latent dynamics transfer across the evaluated non-clinical and clinical cohorts (see Sec. 4.3).

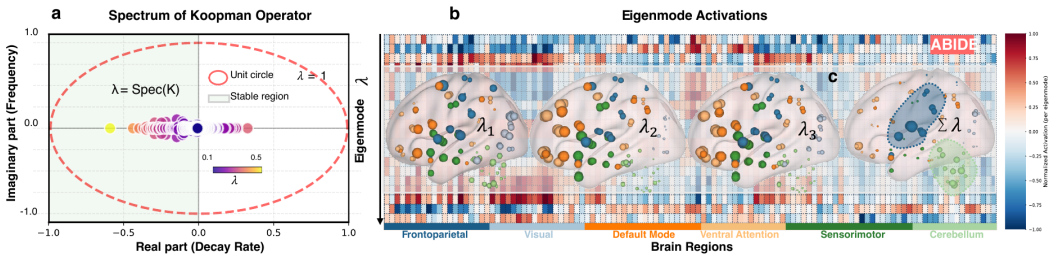

Figure 3: ABIDE eigen-spectrum and decoded eigenmode analysis. **(a)** Complex eigenvalue spectrum of the learned DPLR-Koopman latent operator. Points denote eigenvalues; the stable region is shown for reference. The spectrum is predominantly real-dominated with small imaginary components, suggesting bounded latent evolution. **(b)** Decoded eigenmode heatmap in the AAL-116 ROI space. Rows correspond to the top discrete eigenmodes sorted by decreasing eigenvalue magnitude; columns correspond to AAL-116 ROIs grouped by functional subnetworks; colors denote row-normalized signed loadings. **(c)** Anatomical projections of representative eigenmodes and aggregate mode-density summaries, where node size denotes absolute loading magnitude and color denotes signed contribution. Additional datasets are shown in Appendix E.5.

4.3 Interpretable Analysis on Eigenmode Dynamics

A key advantage of BRAINMECH is that the learned DPLR-Koopman core enables direct eigenmode analysis, decomposing brain dynamics into interpretable spectral components. We analyze the learned Koopman matrix K via eigenvalue decomposition (Fig. 3a) for ASD since Autism is strongly associated with altered functional dynamics. On the ABIDE dataset, the learned spectrum is predominantly real-dominated with relatively small imaginary components. This pattern suggests bounded and largely convergent latent dynamics with modest oscillatory structure, rather than strongly oscillatory dynamics. We therefore interpret the spectrum conservatively as evidence of stable latent evolution, not as direct evidence of a biological oscillatory mechanism. Each Koopman eigenmode corresponds to a distinct dynamical component characterized by a specific timescale and frequency. The associated spatial activation patterns (Fig. 3b) exhibit clear network-level structure, with differential engagement of default mode, frontoparietal, ventral attention, and sensorimotor networks, indicating that BRAINMECH decomposes whole-brain dynamics into interpretable functional building blocks. Ordering eigenmodes from low to high frequency and projecting them onto the brain surface (Fig. 3c) reveals a hierarchical organization, where low-frequency modes predominantly involve default mode, frontoparietal, and cerebellar regions associated with global integration and internally oriented processing (Raichle et al., 2001; Vincent et al., 2008), whereas higher-frequency modes are increasingly localized to sensorimotor and ventral attention cortices, consistent with faster and more transient functional processes (Cordes et al., 2001; Corbetta and Shulman, 2002). Finally, the aggregated eigenmode contribution map highlights frontoparietal and cerebellar regions as major contributors to the overall latent dynamics in ABIDE, consistent with converging clinical and neurodevelopmental evidence of atypical frontoparietal control and cerebellar involvement in autism spectrum disorder (Courchesne et al., 2001; Vasa et al., 2016).

4.4 Ablation Studies

Architectural component. To quantify the contribution of each architectural component, we conduct systematic ablations across multiple datasets. The results are summarized in Fig. 4. Removing temporal components (No Temporal) causes the largest performance drop on sex classification (67.1% on HCPA), indicating that temporal modeling is critical. For disease diagnosis, the impact varies by dataset, removing TASEN components causes the most severe degradation on ADNI (41.9%), while spatial components are critical for ABIDE.

Koopman rank. Increasing the rank r of DPLR-Koopman core gradually improves performance, with rank $r = 16$ achieving the best disease classification accuracy (74.86%) for latent dimension $d = 256$ (see Table 5 for details). As shown in Fig. 5, our model benefits from the Koopman core dynamics flexibility provided by higher-rank parameterization, with $r = 16$ representing a sweet spot for this latent dimension. Beyond this spot, performance plateaus, indicating diminishing returns. In

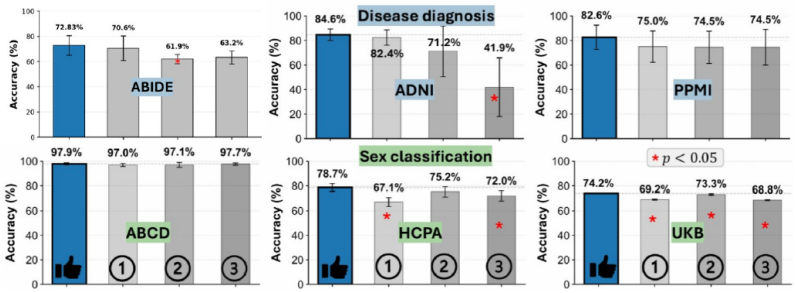

Figure 4: Ablation study results across disease diagnosis (top) and sex classification (bottom) tasks. Each plot shows the effect of removing different components (①Temporal $f^{\theta_3}_{\text{tem}}$, ②Spatial $f^{\theta_2}_{\text{spa}}$, ③TASEN $f^{\theta_3}_{\text{tem}} + f^{\theta_2}_{\text{spa}}$) compared to the full model 👍.

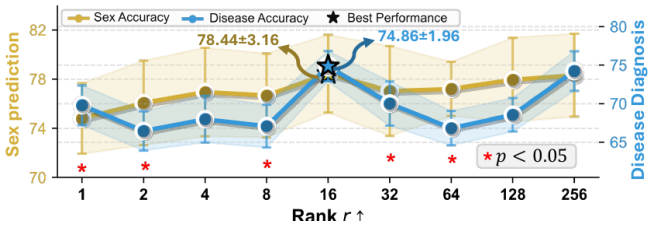

Figure 5: Effect of Koopman rank r on downstream accuracy. Disease performance peaks at $r = 16$.

practice, we scale the rank proportionally with the hidden dimension size to maintain a consistent ratio between rank and latent dimensionality.

4.5 Mediation Analysis

Region-level mediation analysis. To understand how the learned latent representations capture disease-relevant information, we perform region-level mediation analysis to examine whether model features mediate the relationship between BOLD time series and disease diagnosis. For each region, we test the indirect pathway: BOLD $\rightarrow$ learned features $\rightarrow$ disease, controlling for age and sex as confounders. This analysis reveals which brain region contributes most significantly to disease prediction through the learned features. Using the learned BRAINMECH representations, the mediation analysis provides ROI-level evidence of disease-associated signals encoded by the latent features across ABIDE, PPMI, and ADNI (Fig. 6): (1) *Significant Mediation*: We observe robust indirect effects in specific regions (dashed contours, $p < 0.05$), indicating the model successfully encodes spatially localized, disease-relevant signals. Compared with BrainJEPA and BrainMass, BRAINMECH produces stronger and more spatially organized indirect effects across datasets. (2) *Biological Alignment*: The spatial distribution of mediated effects aligns with known disease mechanisms. For ABIDE, the features map to the default mode network, a hallmark of ASD dysfunction (Monk et al., 2009). For PPMI, the model captures widespread involvement in the sensorimotor and visual networks, reflecting the global functional disruptions characteristic of PD (Fiorenzato et al., 2019). For ADNI, the mediated regions involve ventral attention and default-mode systems, consistent with large-scale network disruption in Alzheimer's disease (Li et al., 2012). (3) *Signal Enhancement*: The indirect effects are consistently more widespread and significant than direct effects, suggesting that the learned features encode disease-associated information beyond mean ROI-level BOLD signals under the specified mediation model. More experimental results are shown in Appendix E.

5 Conclusion

We presented BRAINMECH, a Koopman-structured stochastic latent dynamics foundation model for fMRI. Rather than treating fMRI solely as an implicit sequence representation problem, BRAINMECH combines topology-aware spatiotemporal encoding, a DPLR latent operator, age-conditioned control, and learned diffusion to obtain stable and analyzable latent dynamics. Across large-scale pretraining

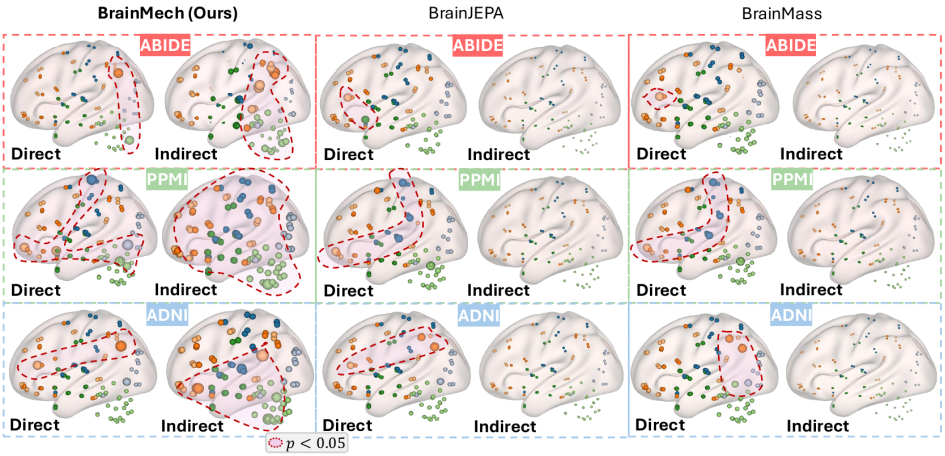

Figure 6: Mediation analysis across three disease datasets and three foundation models. For each dataset, we compare direct effects from ROI-level BOLD signals to disease and indirect effects mediated through learned features. Dashed red contours indicate ROIs with significant mediation effects ($p < 0.05$).

and multiple downstream evaluations, BRAINMECH improves predictive performance under the evaluated protocols. More importantly, eigenmode analysis, component ablations, and region-level mediation results suggest that the learned representation captures structured temporal and spatial patterns relevant to clinical cohorts. The current evidence supports BRAINMECH as an interpretable latent dynamics model, not as a causal discovery tool for neurobiological mechanisms. Future work will focus on stronger cross-site validation, multimodal covariates, intervention-aware modeling, and prospective clinical evaluation.

References

Christian F. Beckmann and Stephen M. Smith. Tensorial extensions of independent component analysis for multisubject fmri analysis. NeuroImage, 25(1):294–311, 2005. doi: 10.1016/j.neuroimage.2004.10.043.

Richard F. Betzel, Lisa Byrge, Ye He, Joaquín Goñi, Xi-Nian Zuo, and Olaf Sporns. Changes in structural and functional connectivity among resting-state networks across the human lifespan. NeuroImage, 102:345–357, 2014. doi: 10.1016/j.neuroimage.2014.07.067.

Josue Ortega Caro, Antonio H. de O. Fonseca, Christopher Averill, Syed A. Rizvi, Matteo Rosati, James L. Cross, Prateek Mittal, Emanuele Zappala, Daniel Levine, Rahul M. Dhodapkar, Chadi G. Abdallah, and David van Dijk. Brainlm: A foundation model for brain activity recordings. bioRxiv, 2023. doi: 10.1101/2023.09.12.557460.

B. J. Casey, T. Cannonier, M. I. Conley, A. O. Cohen, D. M. Barch, M. M. Heitzeg, M. E. Soules, T. Teslovich, D. V. Dellarco, H. Garavan, et al. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. Developmental Cognitive Neuroscience, 32: 43–54, 2018. doi: 10.1016/j.dcn.2018.03.001.

Maurizio Corbetta and Gordon L Shulman. Control of goal-directed and stimulus-driven attention in the brain. Nature reviews neuroscience, 3(3):201–215, 2002.

Dietmar Cordes, Victor M Haughton, Konstantinos Arfanakis, John D Carew, Patrick A Turski, Chad H Moritz, Michelle A Quigley, and M Elizabeth Meyerand. Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. American journal of neuroradiology, 22(7):1326–1333, 2001.

Eric Courchesne, CM Karns, HR Davis, R Ziccardi, RA Carper, ZD Tigue, HJ Chisum, P Moses, K Pierce, C Lord, et al. Unusual brain growth patterns in early life in patients with autistic disorder: an mri study. Neurology, 57(2):245–254, 2001.

Tingting Dan, Ziquan Wei, Won Hwa Kim, and Guorong Wu. Exploring the enigma of neural dynamics through a scattering-transform mixer landscape for Riemannian manifold. In Proceedings of the 41st International Conference on Machine Learning, volume 235, pages 9976–9990, 2024.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering, 2016. URL https://arxiv.org/abs/1606.09375.

Zijian Dong, Ruilin Li, Yilei Wu, Thuan Tinh Nguyen, Joanna Su Xian Chong, Fang Ji, Nathanael Ren Jie Tong, Christopher Li Hsian Chen, and Juan Helen Zhou. Brain-JEPA: Brain dynamics foundation model with gradient positioning and spatiotemporal masking. In Advances in Neural Information Processing Systems, volume 37, 2024. doi: 10.48550/arXiv.2409.19407. URL https://arxiv.org/abs/2409.19407. NeurIPS 2024 Spotlight.

Zijian Dong, Ruilin Li, Joanna Su Xian Chong, Niousha Dehestani, Yinghui Teng, Yi Lin, Zhizhou Li, Yichi Zhang, Yapei Xie, Leon Qi Rong Ooi, B. T. Thomas Yeo, and Juan Helen Zhou. Brain harmony: A multimodal foundation model unifying morphology and function into 1d tokens, 2025.

Nicha C. Dvornek, Pamela Ventola, Kevin A. Pelphrey, and James S. Duncan. Identifying autism from resting-state fmri using long short-term memory networks. In Machine Learning in Medical Imaging (MLMI), MICCAI Workshop, volume 10541 of Lecture Notes in Computer Science, pages 362–370, 2017. doi: 10.1007/978-3-319-67389-9_42.

Jennifer Stine Elam, Matthew F. Glasser, Michael P. Harms, Stamatios N. Sotiropoulos, Jesper L. R. Andersson, Gregory C. Burgess, Sandra W. Curtiss, Robert Oostenveld, Linda J. Larson-Prior, Jan-Mathijs Schoffelen, et al. The Human Connectome Project: A retrospective. NeuroImage, 244: 118543, 2021. doi: 10.1016/j.neuroimage.2021.118543.

Lingzhong Fan, Hai Li, J Zhuo, Yu Zhang, J Wang, L Chen, Z Yang, C Chu, S Xie, Angela R. Laird, Peter T. Fox, Simon B. Eickhoff, Chunshui Yu, and Tianzi Jiang. The human Brainnetome Atlas: A new brain atlas based on connectional architecture. Cerebral Cortex, 26(8):3508–3526, 2016. ISSN 1047-3211. doi: 10.1093/cercor/bhw157. URL https://doi.org/10.1093/cercor/bhw157.

Eleonora Fiorenzato, Antonio P Strafella, Jinhee Kim, Roberta Schifano, Luca Weis, Angelo Antonini, and Roberta Biundo. Dynamic functional connectivity changes associated with dementia in parkinson's disease. Brain, 142(9):2860–2872, 2019.

Karl J. Friston, Lee Harrison, and Will Penny. Dynamic causal modelling. NeuroImage, 19(4): 1273–1302, 2003. doi: 10.1016/S1053-8119(03)00202-7.

Kaiqiao Han, Yi Yang, Zijie Huang, Xuan Kan, Yang Yang, Ying Guo, Lifang He, Liang Zhan, Yizhou Sun, Wei Wang, and Carl Yang. Brainode: Dynamic brain signal analysis via graph-aided neural ordinary differential equations. arXiv preprint arXiv:2405.00077, 2024.

Anibal Sólon Heinsfeld, Alexandre Rosa Franco, R. Cameron Craddock, Augusto Buchweitz, and Felipe Meneguzzi. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage: Clinical, 17:16–23, 2018. doi: 10.1016/j.nicl.2017.08.017.

R. Matthew Hutchison, Thilo Womelsdorf, Elena A. Allen, Peter A. Bandettini, Vince D. Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H. Duyn, Gary H. Glover, Javier Gonzalez-Castillo, Daniel A. Handwerker, Shella Keilholz, Vesa Kiviniemi, David A. Leopold, Francesco de Pasquale, Olaf Sporns, Marcus Walter, and Catie Chang. Dynamic functional connectivity: Promise, issues, and interpretations. NeuroImage, 80:360–378, 2013. doi: 10.1016/j.neuroimage.2013.05.079.

Stefan Klus, Feliks Nüske, Sebastian Peitz, Jan-Hendrik Niemann, Cecilia Clementi, and Christof Schütte. Data-driven approximation of the Koopman generator: Model reduction, system identification, and control. Physica D: Nonlinear Phenomena, 406:132416, 2020.

Bernard O. Koopman. Hamiltonian systems and transformation in hilbert space. Proceedings of the National Academy of Sciences, 17(5):315–318, 1931. doi: 10.1073/pnas.17.5.315.

Rui Li, Xia Wu, Adam S. Fleisher, Eric M. Reiman, Kewei Chen, and Li Yao. Attention-related networks in alzheimer's disease: a resting functional mri study. Human Brain Mapping, 33(5): 1076–1088, 2012. doi: 10.1002/hbm.21269.

Xiaoxiao Li, Yuan Zhou, Nicha C. Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H. Staib, Pamela Ventola, and James S. Duncan. BrainGNN: Interpretable brain graph neural network for fMRI analysis. In Medical Image Analysis, volume 74, page 102233, 2021.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2017. URL https://arxiv.org/abs/1711.05101.

Daniel J. Lurie, Daniel Kessler, Danielle S. Bassett, Richard F. Betzel, Michael Breakspear, Shella Kheilholz, Aaron Kucyi, Raphaël Liégeois, Martin A. Lindquist, Anthony R. McIntosh, Russell A. Poldrack, James M. Shine, William H. Thompson, Natalia Z. Bielczyk, Linda Douw, Robert T. Krafty, Robin L. Miller, Muthuraman Muthuraman, Leonardo Pasquini, Abhishek Rajan, Diego Vidaurre, Hongming Xie, and Vince D. Calhoun. Questions and controversies in the study of time-varying functional connectivity in resting fmri. Network Neuroscience, 4(1):30–69, 2020. doi: 10.1162/netn_a_00116.

Joliot Marc, Gael Jobard, Mikaël Naveau, Nicolas Delcroix, Laurent Petit, Laure Zago, Fabrice Crivello, Emmanuel Mellet, Bernard Mazoyer, and Nathalie Tzourio-Mazoyer. Aicha: An atlas of intrinsic connectivity of homotopic areas. Journal of neuroscience methods, 254, 07 2015. doi: 10.1016/j.jneumeth.2015.07.013.

Kenneth Marek, S. Chowdhury, A. Siderowf, S. Lasch, C. S. Coffey, C. Caspell-Garcia, T. Simuni, D. Jennings, C. M. Tanner, J. Q. Trojanowski, L. M. Shaw, J. Seibyl, N. Schuff, A. Singleton, K. Kieburtz, A. W. Toga, B. Mollenhauer, D. Galasko, L. M. Chahine, D. Weintraub, et al. The Parkinson's Progression Markers Initiative (PPMI) – establishing a PD biomarker cohort. Annals of Clinical and Translational Neurology, 5(12):1460–1477, 2018. doi: 10.1002/acn3.644.

Scott Marek, Brenden Tervo-Clemmens, Finnegan J Calabro, David F Montez, Benjamin P Kay, Alexander S Hatoum, Meghan Rose Donohue, William Foran, Ryland L Miller, Timothy J Hendrickson, et al. Reproducible brain-wide association studies require thousands of individuals. Nature, 603(7902):654–660, 2022.

Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. Nonlinear Dynamics, 41(1-3):309–325, 2005. doi: 10.1007/s11071-005-2824-x.

Karla L. Miller, Fidel Alfaro-Almagro, Neal K. Bangerter, Derek L. Thomas, Essa Yacoub, Junqian Xu, Andreas J. Bartsch, Saad Jbabdi, Stamatios N. Sotiropoulos, Jesper L. R. Andersson, Ludovica Griffanti, Gaël Douaud, Thomas W. Okell, Peter Weale, Irina Dragonu, Sean Garratt, Simon Hudson, Rory Collins, Mark Jenkinson, Paul M. Matthews, and Stephen M. Smith. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nature Neuroscience, 19(11):1523–1536, 2016. doi: 10.1038/nn.4393.

Christopher S Monk, Scott J Peltier, Jillian Lee Wiggins, Shih-Jen Weng, Melisa Carrasco, Susan Risi, and Catherine Lord. Abnormalities of intrinsic functional connectivity in autism spectrum disorders. Neuroimage, 47(2):764–772, 2009.

Bernt Oksendal. Stochastic Differential Equations: An Introduction with Applications. Springer, 5 edition, 2013.

Grigorios A. Pavliotis. Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations. Springer, 2014.

Ronald C. Petersen, Paul S. Aisen, Laurel A. Beckett, Michael C. Donohue, Anthony C. Gamst, Danielle J. Harvey, Clifford R. Jack, William J. Jagust, Leslie M. Shaw, Arthur W. Toga, John Q. Trojanowski, and Michael W. Weiner. Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. Neurology, 74(3):201–209, 2010. doi: 10.1212/WNL.0b013e3181cb3e25.

Maria Giulia Preti, Thomas A. W. Bolton, and Dimitri Van De Ville. The dynamic functional connectome: State-of-the-art and perspectives. NeuroImage, 160:41–54, 2017. doi: 10.1016/j.neuroimage.2016.12.061.

Marcus E Raichle, Ann Mary MacLeod, Abraham Z Snyder, William J Powers, Debra A Gusnard, and Gordon L Shulman. A default mode of brain function. Proceedings of the national academy of sciences, 98(2):676–682, 2001.

X. Shen, F. Tokoglu, X. Papademetris, and R.T. Constable. Groupwise whole-brain parcellation from resting-state fmri data for network node identification. NeuroImage, 82:403–415, 2013. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2013.05.081. URL https://www.sciencedirect.com/science/article/pii/S1053811913005818.

N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. NeuroImage, 15(1):273–289, 2002. ISSN 10538119. doi: https://doi.org/10.1006/nimg.2001.0978. URL https://www.sciencedirect.com/science/article/pii/S1053811901909784.

Roma A Vasa, Stewart H Mostofsky, and Joshua B Ewen. The disrupted connectivity hypothesis of autism spectrum disorders: time for the next phase in research. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 1(3):245–252, 2016.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks, 2018. URL https://arxiv.org/abs/1710.10903.

Justin L Vincent, Itamar Kahn, Abraham Z Snyder, Marcus E Raichle, and Randy L Buckner. Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. Journal of neurophysiology, 100(6):3328–3342, 2008.

Jiaxing Xu, Yunhan Yang, David Huang, Sophi Shilpa Gururajapathy, Yiping Ke, Miao Qiao, Alan Wang, Haribalan Kumar, Josh McGeown, and Eryn Kwon. Data-driven network neuroscience: On data collection and benchmark. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, Advances in Neural Information Processing Systems, volume 36, pages 21841–21856. Curran Associates, Inc., 2023.

Ye C. Su G. Zhang Z. Chang Z. Chen H. ... Ma T. Yang, Y. Brainmass: Advancing brain network analysis for diagnosis with large-scale self-supervised learning. In IEEE transactions on medical imaging, 2024.

A Appendix Overview

This appendix provides additional details that complement the main paper, including theory details for the Koopman-structured stochastic latent dynamics, network architecture specifications, experimental setup information, and further analyses that support the claims in the main text.

B Theory Details for Koopman-Structured Stochastic Latent Dynamics

This section provides the standard stochastic Koopman and SDE details underlying the compact formulation in Sec. 3.1. These results are included for completeness and to clarify the assumptions behind the latent stochastic process. They are not claimed as new mathematical contributions. The model-specific contributions of BRAINMECH are the control-aware latent SDE, the DPLR parameterization of the latent drift operator, and the stability-regularized learning objective.

B.1 Latent SDE and Stochastic Koopman Semigroup

We model the latent process as

$$dZ_t \;=\; \big(K Z_t + C\, u_t + \beta\big)\, dt \;+\; \Sigma_\theta\big(Z_t, u_t\big)\, dW_t. \tag{11}$$

For SDE-driven Markov processes, the stochastic Koopman semigroup acts on observables through conditional expectation:

$$(\mathcal{K}^t g)(z) \;:=\; \mathbb{E}\big[g(Z_t) \mid Z_0 = z\big]. \tag{12}$$

This provides an operator-theoretic view of the learned latent dynamics. Importantly, we do not assume that the neural encoder recovers exact Koopman observables; instead, the Koopman structure is used as an inductive bias for learning stable and analyzable latent dynamics.

B.2 Well-posedness and Semigroup Property

Assumption 1 (Well-posed latent SDE)**.** *Assume the context process u_t is progressively measurable and bounded on finite horizons, and the diffusion $\Sigma_\theta(z, u)$ is globally Lipschitz in z uniformly over u and satisfies linear growth:*

$$\begin{cases} \|\Sigma_\theta(z,u) - \Sigma_\theta(z',u)\|_F \le L\|z - z'\|, \\ \|\Sigma_\theta(z,u)\|_F^2 \le c_0 + c_1\|z\|^2. \end{cases} \tag{13}$$

Then the latent SDE admits a unique non-explosive strong solution $(Z_t)_{t\ge 0}$ for any initial condition $Z_0 = z$.

Assumption 1 ensures that the stochastic process used by BRAINMECH is mathematically well-defined and that the stochastic Koopman semigroup is a valid object.

Proposition 1 (Linearity and semigroup property)**.** *Let $(Z_t)_{t\ge 0}$ be the Markov process induced by the latent SDE under Assumption 1. Define $(\mathcal{K}^t g)(z) := \mathbb{E}[g(Z_t) \mid Z_0 = z]$ for any bounded measurable observable g. Then:*

$$\mathcal{K}^t(ag_1 + bg_2) = a\mathcal{K}^t g_1 + b\mathcal{K}^t g_2,$$

for $a, b \in \mathbb{R}$, and

$$\mathcal{K}^{t+s} = \mathcal{K}^t \circ \mathcal{K}^s, \qquad \mathcal{K}^0 = \mathrm{Id}.$$

Proof. Linearity follows from linearity of expectation. For semigroup, apply the tower property and Markov property:

$$\begin{aligned} (\mathcal{K}^{t+s} g)(z) &= \mathbb{E}[g(Z_{t+s}) \mid Z_0 = z] \\ &= \mathbb{E}[\mathbb{E}[g(Z_{t+s}) \mid Z_t] \mid Z_0 = z] \\ &= \mathbb{E}[(\mathcal{K}^s g)(Z_t) \mid Z_0 = z] = \big(\mathcal{K}^t(\mathcal{K}^s g)\big)(z). \end{aligned} \tag{14}$$

B.3 Generator View and PDE Interpretation

Theorem 1 (Infinitesimal generator of the stochastic Koopman semigroup)**.** *Let the drift be $f(z,u) := Kz + Cu + \beta$ and let $Q_\theta(z,u) := \Sigma_\theta(z,u)\Sigma_\theta(z,u)^\top$. For any $g \in C_b^2(\mathbb{R}^d)$, the generator $\mathcal{G}$ of $(\mathcal{K}^t)_{t\geq 0}$ is*

$$(\mathcal{G}g)(z) = \nabla g(z)^\top f(z,u) + \frac{1}{2}\mathrm{Tr}\Big(Q_\theta(z,u)\nabla^2 g(z)\Big), \tag{15}$$

and $v(t,z) := (\mathcal{K}^t g)(z)$ satisfies the backward Kolmogorov equation

$$\partial_t v(t,z) = (\mathcal{G}v)(t,z), \qquad v(0,z) = g(z).$$

Proof. Apply Itô's formula to $g(Z_t)$, substitute the latent SDE, and take conditional expectation given $Z_0 = z$. The Itô integral term has zero mean, yielding the generator form and the backward equation via standard semigroup arguments (Oksendal, 2013; Pavliotis, 2014).

This provides the operator-generator interpretation behind the PDE-based terminology: the learned drift and diffusion define a generator for expected observables, although the model should not be interpreted as solving a biologically specified PDE of the brain.

B.4 Linear Observables and Eigenmode Analysis

Proposition 2 (Closure of linear observables under linear drift)**.** *For the class of linear observables*

$$\mathcal{G}_{\mathrm{lin}} = \{g_\alpha(z) = \alpha^\top z : \ \alpha \in \mathbb{R}^d\},$$

the stochastic Koopman semigroup maps $\mathcal{G}_{\mathrm{lin}}$ to affine functions: there exist $\alpha_t \in \mathbb{R}^d$ and $\beta_t \in \mathbb{R}$ such that

$$(\mathcal{K}^t g_\alpha)(z) = \alpha_t^\top z + \beta_t.$$

Proof. Since the drift is affine in Z_t, $\mathbb{E}[Z_t \mid Z_0 = z]$ solves a linear ODE in z, with affine terms from β and u_t, and therefore remains affine in z. Then $(\mathcal{K}^t g_\alpha)(z) = \alpha^\top \mathbb{E}[Z_t \mid Z_0 = z]$ is affine.

This closure property motivates analyzing the learned latent drift through spectra and eigenmodes, while keeping the interpretation limited to the learned latent coordinate system.

B.5 Mean-Square Stability Motivation

Theorem 2 (Mean-square dissipativity under negative symmetric drift)**.** *Assume $u_t \equiv 0$ and $\beta = 0$. Suppose there exists $\alpha > 0$ such that*

$$\frac{1}{2}(K + K^\top) \preceq -\alpha I.$$

Further assume $\|\Sigma_\theta(z,0)\|_F^2 \leq a + b\|z\|^2$ with $b < 2\alpha$. Then

$$\mathbb{E}\|Z_t\|^2 \leq e^{-(2\alpha - b)t}\|Z_0\|^2 + \frac{a}{2\alpha - b}\big(1 - e^{-(2\alpha-b)t}\big). \tag{16}$$

Proof. Let $V(z) = \|z\|^2$. By Itô's formula,

$$\frac{d}{dt}\mathbb{E}V(Z_t) = \mathbb{E}\big[2Z_t^\top K Z_t + \|\Sigma_\theta(Z_t,0)\|_F^2\big].$$

Using $Z^\top K Z = Z^\top \frac{1}{2}(K + K^\top)Z \leq -\alpha\|Z\|^2$ and the growth bound on Σ_θ, we obtain

$$\frac{d}{dt}\mathbb{E}\|Z_t\|^2 \leq -(2\alpha - b)\mathbb{E}\|Z_t\|^2 + a.$$

Applying Grönwall's inequality gives the result.

This bound motivates regularizing the symmetric part $\frac{1}{2}(K + K^\top)$ as a practical surrogate for promoting stable continuous-time latent dynamics during training.

B.6 Continuous-Time Stability and Sampled-Time Spectra

The latent dynamics in Eq. 2 are formulated in continuous time:

$$dZ_t = (KZ_t + Cu_t + \beta)\,dt + \Sigma_\theta(Z_t, u_t)\,dW_t.$$

Thus, K is a continuous-time drift/generator, not a discrete-time transition matrix. For a fixed control signal over a sampling interval $[t_\tau, t_{\tau+1})$, the deterministic homogeneous part induces the exact sampled-time transition

$$\Phi_{\Delta t} = \exp(\Delta t K).$$

Euler–Maruyama instead uses the first-order approximation $I+\Delta t K$ of $\exp(\Delta t K)$ when constructing the likelihood on discretely sampled fMRI observations.

Accordingly, continuous-time stability of K is characterized by the spectral abscissa

$$\alpha(K) = \max_i \mathrm{Re}(\lambda_i(K)).$$

A sufficient dissipativity condition is

$$\mu_2(K) = \lambda_{\max}\left(\frac{K + K^\top}{2}\right) < 0,$$

which implies $\alpha(K) < 0$. In contrast, spectral-radius stability is a discrete-time criterion and should be applied to the sampled transition:

$$\rho(\Phi_{\Delta t}) = \rho(\exp(\Delta t K)) < 1.$$

Since the eigenvalues of $\exp(\Delta t K)$ are $\exp(\Delta t \lambda_i(K))$, we have

$$\rho(\exp(\Delta t K)) < 1 \quad \Longleftrightarrow \quad \max_i \mathrm{Re}(\lambda_i(K)) < 0.$$

Therefore, eigenvalue plots of K should be interpreted in the continuous-time complex plane, with stability determined by $\mathrm{Re}(\lambda) < 0$. If sampled-time eigenvalues are plotted, they should be explicitly labeled as $\phi_i = \exp(\Delta t \lambda_i(K))$, for which the unit-circle criterion applies.

B.7 Euler–Maruyama Discretization

Using Euler–Maruyama discretization, the latent SDE becomes

$$z_{\tau+1} = z_\tau + \big(Kz_\tau + Cu_\tau + \beta\big)\Delta t + \Sigma_\theta(z_\tau, u_\tau)\sqrt{\Delta t}\,\xi_\tau, \qquad \xi_\tau \sim \mathcal{N}(0, I). \tag{17}$$

Proposition 3 (Euler–Maruyama induces a Gaussian one-step conditional)**.** *Conditioned on fixed (z_τ, u_τ), the one-step transition is Gaussian:*

$$z_{\tau+1} \mid z_\tau, u_\tau \sim \mathcal{N}\Big(z_\tau + (Kz_\tau + Cu_\tau + \beta)\Delta t,\ \Delta t\, Q_\theta(z_\tau, u_\tau)\Big),$$

where $Q_\theta(z_\tau, u_\tau) = \Sigma_\theta(z_\tau, u_\tau)\Sigma_\theta(z_\tau, u_\tau)^\top$.

Proof. Conditioned on (z_τ, u_τ), the only random term is $\xi_\tau \sim \mathcal{N}(0, I)$. Therefore $\Sigma_\theta(z_\tau, u_\tau)\sqrt{\Delta t}\xi_\tau$ is Gaussian with covariance $\Delta t\, Q_\theta(z_\tau, u_\tau)$.

C Model Architecture Details

C.1 Network Architecture Overview

To complement the methodological description and the high-level architecture in Fig. 2, we summarize the encoder-side network flow and key hyperparameters of BRAINMECH in Fig. 7. The architecture first receives the ROI time series $x \in \mathbb{R}^{B\times T\times 116}$ and an optional functional-connectivity adjacency matrix `fc_matrix` $\in \mathbb{R}^{116\times 116}$ or $\mathbb{R}^{B\times 116\times 116}$. The Spatial Encoder $f_{\theta_2}^{\mathrm{spa}}$ embeds each scalar ROI value from $1 \to 128$, constructs or receives the graph adjacency, clips negative functional-connectivity edges to zero, normalizes the adjacency, applies DropEdge with $p = 0.1$, and processes the graph through the default ChebConv path with three Chebyshev graph-convolution layers of order $K = 3$, each followed by residual addition, LayerNorm, and dropout. The resulting 116×128 node

Algorithm 1 DPLR-Koopman Core

Require: Latent dimension d, control dimension d_c, drift rank $r \ll d$, target spectral radius ρ

$r_c \leftarrow \min(r, d_c)$

Parameters:

$k \in \mathbb{R}^d,\ U, V \in \mathbb{R}^{d \times r},\ C_1 \in \mathbb{R}^{d \times r_c},\ C_2 \in \mathbb{R}^{d_c \times r_c}$

Initialization:

$k_i \sim \mathcal{U}(-0.5\rho,\ 0.5\rho)$ for $i = 1, \ldots, d$

$U \leftarrow 0.1 \cdot \mathrm{Orth}(d, r),\ \ V \leftarrow 0.1 \cdot \mathrm{Orth}(d, r)$

$C_1 \leftarrow \mathrm{XavierUniform}(d, r_c; \text{gain} = 0.1),\ \ C_2 \leftarrow \mathrm{XavierUniform}(d_c, r_c; \text{gain} = 0.1)$

function DRIFT(z_t, u_t)

  $K \leftarrow \mathrm{diag}(k) + UV^\top,\ \ C \leftarrow C_1 C_2^\top$

  $\mu_t \leftarrow k \odot z_t + (z_t U)V^\top + (u_t C_2)C_1^\top$ ▷ matrix-free evaluation

  return μ_t ▷ used as deterministic drift in Eq. 2

end function

procedure SPECTRALPROJECT

  $\hat{K} \leftarrow \mathrm{diag}(k) + UV^\top,\ \ s \leftarrow \rho(\hat{K})$

  if $s > \rho$ **then**

    $\gamma \leftarrow \rho/(s + \varepsilon)$

    $k \leftarrow \gamma k,\ \ U \leftarrow \sqrt{\gamma}\, U,\ \ V \leftarrow \sqrt{\gamma}\, V$

  end if

end procedure

for each training iteration **do**

  Compute loss using latent SDE rollout with DRIFT($\cdot, \cdot$)

  Update $\{k, U, V, C_1, C_2\}$ by gradient descent

  SPECTRALPROJECT

end for

representation is flattened into a 14848-dimensional vector, projected back to 128 dimensions with Linear + GELU + LayerNorm, and combined with a learned temporal projection $128 \to 128$ to produce $z_{\text{spatial}} \in \mathbb{R}^{B \times T \times 128}$.

The Temporal Encoder $f_{\theta_3}^{\text{tem}}$ takes z_{spatial}, adds sinusoidal positional encoding with dropout, applies three Transformer encoder layers with eight attention heads and FFN dimension 512, and finishes with Linear $128 \to 128$ + LayerNorm to produce $z_{\text{temporal}} \in \mathbb{R}^{B \times T \times 128}$. In parallel, the Episodic Feedback Encoder $f_{\theta_5}^{\text{epi}}$ summarizes z_{temporal} by first applying an input projection $128 \to 64$, then passing the sequence through two Transformer encoder layers with four heads and FFN size 128, followed by global average pooling over time and Linear $64 \to 64$ + GELU + Dropout to produce `episodic_summary` $\in \mathbb{R}^{B \times 64}$. The Task Context Encoder $f_{\theta_6}^{\text{ctx}}$ maps `task_vectors` $\in \mathbb{R}^{B \times T \times 32}$ through two Linear + GELU + LayerNorm + Dropout blocks followed by Linear $64 \to 64$, yielding `context_embedding` $\in \mathbb{R}^{B \times T \times 64}$. Together, these encoders transform raw ROI activity, graph topology, temporal dependencies, episodic information, and task context into structured latent representations used by the downstream forecasting module.

C.2 Scientific Data Flow

As shown in Fig. 8, BRAINMECH follows a structured data flow from observed brain activity to latent dynamical evolution. The input BOLD signal $X \in \mathbb{R}^{T \times N}$ and brain network adjacency $A \in \mathbb{R}^{N \times N}$ are first processed by the spatial graph encoder $f_{\theta_2}^{\text{spa}}(x_\tau; A)$, which generates ROI-aware latent features $z_\tau \in \mathbb{R}^d$. These embeddings are then passed into the temporal encoder $f_{\theta_3}^{\text{tem}}(h_{0:T-1})$ to capture sequential dependencies across time.

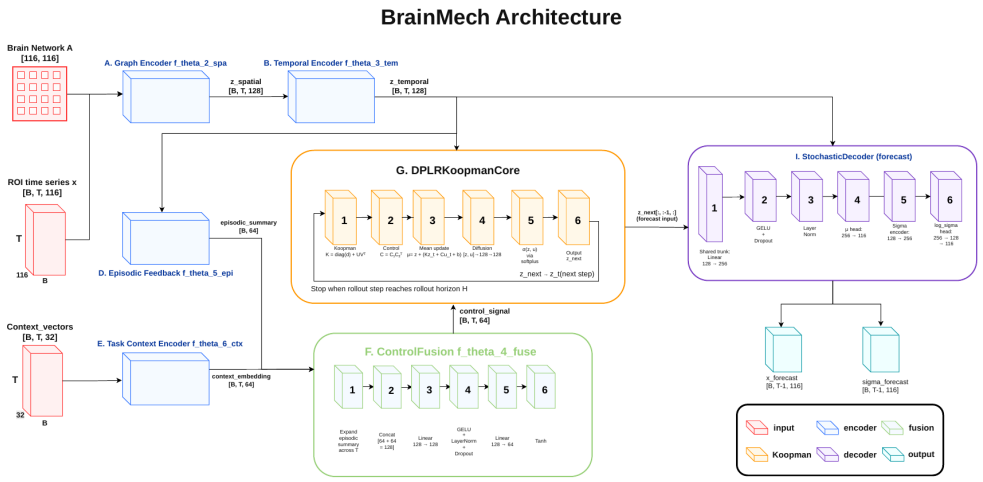

Figure 7: Encoder-side architecture of BRAINMECH. The model first extracts graph-aware spatial embeddings from ROI time series and functional connectivity, then models temporal dependencies, episodic summaries, and task-context embeddings for downstream latent forecasting.

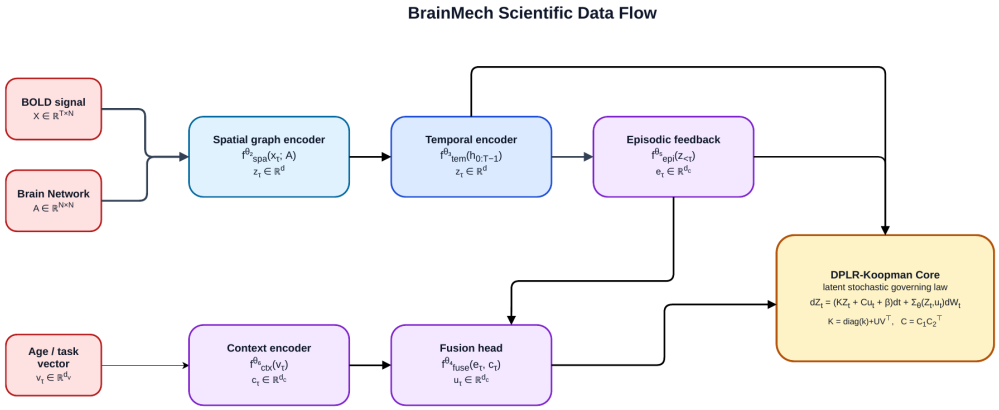

Figure 8: Scientific data flow of BRAINMECH. The model transforms BOLD signals, brain network structure, and age/task context into spatial-temporal latent representations, combines episodic and contextual information into a control signal, and uses the DPLR-Koopman core to govern latent stochastic brain dynamics.

In parallel, the episodic feedback module $f^{\text{epi}}_{\theta_5}(z_{<\tau})$ summarizes prior latent states into $e_\tau \in \mathbb{R}^{d_c}$, while the context encoder $f^{\text{ctx}}_{\theta_6}(v_\tau)$ maps age or task vectors into $c_\tau \in \mathbb{R}^{d_c}$. The fusion head $f^{\text{fuse}}_{\theta_4}(e_\tau, c_\tau)$ combines these signals into the control input $u_\tau \in \mathbb{R}^{d_c}$, which conditions the DPLR-Koopman core and defines the latent stochastic governing law

$$dZ_t = (KZ_t + Cu_t + \beta)\,dt + \Sigma_\theta(Z_t, u_t)\,dW_t, \quad K = \text{diag}(k) + UV^\top, \quad C = C_1C_2^\top.$$

D Experimental Setup Details

D.1 Dataset Statistics

To provide additional context for the main experiments, we summarize the demographic characteristics of each cohort used in this study. These statistics clarify the population coverage of our pooled pretraining data and highlight potential distribution shifts across downstream evaluation cohorts.

Table 3 summarizes the demographic characteristics of all datasets used in this study. The combined cohorts span a wide age range (8.9–100 years), covering key developmental and aging periods across

Table 3: Demographic characteristics of the datasets used in this study. Age is presented as Mean±STD in years. All datasets contain resting-state fMRI recordings preprocessed using standardized pipelines.

Dataset	# of fMRIs	Age (years)	Age Range
ABCD	49,147	11.65 ± 1.49	[8.9, 15.8]
HCPA	720	60.34 ± 15.73	[36.0, 100.0]
UKB	76,603	66.02 ± 7.85	[45.0, 86.0]
ABIDE	2,072	16.46 ± 7.37	[6.0, 58.0]
ADNI	2,126	75.95 ± 7.44	[54.4, 99.6]
PPMI	418	62.95 ± 9.51	[38.0, 84.0]
Total	131,086	—	[6.0, 100.0]

Table 4: Complete hyperparameter settings for model pretraining and downstream task evaluation. These configurations were determined through systematic hyperparameter search and validated across multiple random seeds.

Hyperparameter	Pretraining	Downstream
Optimizer	AdamW	AdamW
Initial Learning Rate	5×10^{-5}	1×10^{-3}
Weight Decay	2×10^{-4}	2×10^{-4}
Dropout Rate	0.2	0.1
LR Scheduler	ReduceLROnPlateau	ReduceLROnPlateau
Scheduler Patience	5 epochs	10 epochs
Scheduler Reduction Factor	0.8	0.5
Sequence Length (T)	300	100
Sliding Window Stride	300	50
Batch Size (per GPU)	8-32	Variable

$N = 131{,}086$ fMRIs. Acquisition parameters (e.g., repetition time) vary across cohorts; accordingly, we normalize the temporal scale in the model by working in normalized time units (Sec. 3.2.2).

D.2 Training Hyperparameters

For reproducibility and to facilitate future work building on BRAINMECH, we detail the complete set of optimization and scheduling hyperparameters used in both pretraining and downstream stages. The choices below were obtained via systematic hyperparameter search and were kept fixed across all reported experiments unless otherwise specified.

Table 4 details the complete hyperparameter configuration for both pretraining and downstream evaluation phases. We employ AdamW optimization with adaptive learning rate scheduling to ensure stable convergence across different dataset scales. The learning rate for pretraining is set lower (5×10^{-5}) to enable gradual learning of the complex Koopman dynamics, while downstream tasks use higher learning rates (1×10^{-3}) for faster adaptation to specific objectives. The code of BRAINMECH can be found in the supplementary.

D.3 Downstream Evaluation Protocol

We follow standard subject-level cross-validation protocols from prior work to ensure a fair comparison with existing fMRI foundation models. This section formalizes our splitting strategy, model heads, and optimization choices for both sex classification and disease diagnosis, complementing the high-level description in Sec. 4.

Following established protocols for comparability with the previous work (Xu et al., 2023), we employ subject-level cross-validation (CV) for all downstream evaluations. To ensure fair comparison with prior studies, ADNI uses 5-fold CV, while ABIDE and PPMI use 10-fold CV.

Sex classification. We evaluate the quality of learned representations by training a MLP head classifier on frozen backbone features. For each subject, we extract latent features $z \in \mathbb{R}^{T \times d}$ and

compute pooled statistics (temporal mean and standard deviation), yielding a fixed-dimensional feature vector. The MLP head consists of two hidden layers (512, 256 units) with LayerNorm, GELU activation, and dropout ($p = 0.3$). We report accuracy averaged over 30 runs with different random seeds.

Disease diagnosis. For clinical classification tasks (ASD vs. Control on ABIDE, CN vs. MCI on ADNI, and 4-class Parkinson's staging on PPMI), we perform full fine-tuning of both the backbone and classification head. We use 10-fold stratified cross-validation to ensure robust evaluation, with class-weighted cross-entropy loss and label smoothing ($\epsilon = 0.1$) to handle class imbalance. We employ differential learning rates (lower for the pretrained backbone, higher for the classification head) and early stopping based on validation F1 score. Results are reported as mean $\pm$ standard deviation across folds.

E Additional Analyses

E.1 Mediation Analysis Details

To quantify how BRAINMECH's representations mediate the relationship between regional BOLD signals and disease diagnosis, we perform ROI-wise mediation analysis following the procedure implemented in our mediation script. For each subject, we first compute the mean BOLD signal per ROI by averaging over time, and extract subject-level latent features from the frozen foundation model by pooling the latent trajectories $z_{1:T}$ (default: temporal mean pooling, with optional variants that concatenate mean and standard deviation). In implementation, we reduce the multi-dimensional latent feature vector to a scalar mediator by averaging across feature dimensions and standardize X, this scalar M, and the confounders C; we reuse the symbols X, M, C below to denote these standardized variables. Let X denote the mean BOLD signal of a given ROI, M the scalar mediator derived from pooled latent features, Y the disease label, and C the confounders (age and sex). We estimate *path a* via a linear regression of M on (X, C),

$$M \;=\; \alpha_X X + \alpha_C^\top C + \varepsilon_a,$$

and *path b* via a regression of Y on (M, X, C) (logistic for binary Y, linear for continuous Y; in the continuous case, Y is standardized before fitting),

$$Y \;=\; \beta_M M + \beta_X X + \beta_C^\top C + \varepsilon_b.$$

The indirect (mediation) effect is defined as the product of the coefficient of X in the first regression and the coefficient of M in the second regression,

$$\text{Indirect effect} \;=\; \alpha_X \beta_M,$$

while the direct effect corresponds to the coefficient of X in the second regression, Direct effect $=$ β_X; their statistical significance is assessed using a bootstrap procedure (typically 5000 resamples) to obtain two-sided p-values. Finally, we correct ROI-wise p-values for multiple comparisons across all regions using standard false discovery rate control and additionally report Bonferroni-corrected values, yielding the ROI-level mediation maps shown in Fig. 6, where the node size encodes the magnitude of the direct or indirect effect coefficient for each region.

E.2 Effect of Koopman Operator Rank

We provide more detailed results in Table 5.

E.3 Parcellation Generalization

A practical requirement for large-scale fMRI foundation models is robustness to the choice of brain atlas. Although our main experiments use the AAL-116 parcellation, BRAINMECH is not tied to a specific ROI dimensionality: the encoder maps an input time series with N regions into a fixed-dimensional latent space, while the Koopman dynamics operate only in this latent space. This design allows the same modeling framework to be applied across atlases with different spatial granularities without modifying the operator-learning module.

To evaluate this property, we train BRAINMECH on UKB using four parcellation schemes: AAL-116, Brainnetome-246 (Fan et al., 2016), Shen-268 (Shen et al., 2013), and AICHA-384 (Marc et al.,

Table 5: Ablation study on the effect of Koopman operator rank r in the DPLR decomposition. Sex classification accuracy is evaluated on HCPA; disease classification accuracy is averaged across ABIDE, PPMI, and ADNI. Results are reported as mean $\pm$ standard deviation over 30 independent runs with different random seeds. The optimal rank $r = 16$ is highlighted.

Rank r	**Sex Accuracy (%)**	**Disease Accuracy (%)**
1	74.80 ± 2.86	69.79 ± 2.58
2	76.06 ± 3.44	66.46 ± 2.53
4	76.95 ± 3.60	67.99 ± 3.03
8	76.67 ± 3.41	67.08 ± 2.78
16	$\mathbf{78.44 \pm 3.16}$	$\mathbf{74.86 \pm 1.96}$
32	77.04 ± 3.65	70.01 ± 2.86
64	77.20 ± 2.21	66.82 ± 2.23
128	77.95 ± 3.42	68.55 ± 2.17
256	78.32 ± 3.37	74.23 ± 2.56

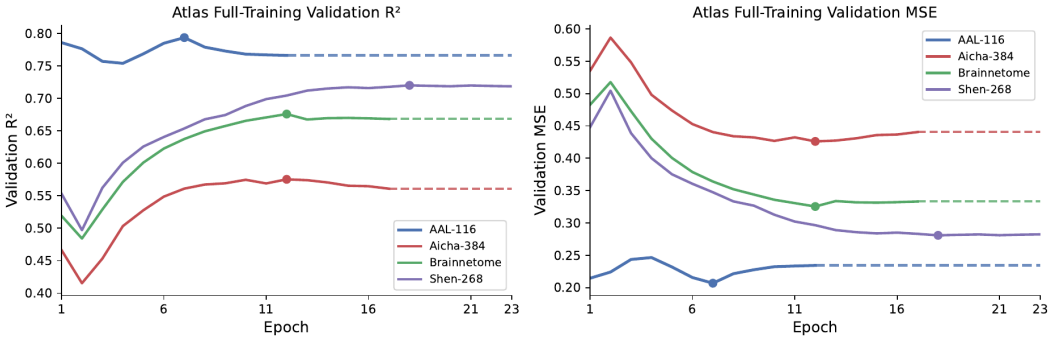

Figure 9: Full-training validation performance across parcellation atlases. We compare validation R^2 (higher is better) and validation MSE (lower is better) for AAL-116, Brainnetome-246, Shen-268, and AICHA-384. The four atlases show broadly consistent training behavior, indicating that BRAINMECH remains stable across different ROI granularities.

2015). All atlas variants use the same training and evaluation protocol, with latent dimension $d = 768$ unless otherwise specified. We compare both full-training validation behavior, measured by validation R^2 and MSE, and downstream transfer performance on disease and biological sex classification. As shown in Fig. 9, the different parcellations exhibit comparable validation trends, suggesting that the model is not overly sensitive to the input atlas resolution. Downstream results further support this conclusion: disease classification remains competitive across ADNI and PPMI (Table 6), while sex classification performance is stable across UKB and HCPA (Table 7). Overall, these results indicate that BRAINMECH generalizes across commonly used fMRI parcellations, with AAL-116 and Shen-268 often providing the strongest or near-strongest performance.

E.4 Model Scaling and Efficiency Analysis

Fig. 10 summarizes the pretraining efficiency of BRAINMECH. Compared with competing brain foundation models under their standard training recipes, BRAINMECH requires substantially less pretraining time while maintaining strong downstream performance. Across model sizes, pretraining remains practical, ranging from approximately 7.3 to 12.4 hours. Overall, BRAINMECH achieves an average 3.65$\times$ speedup over the compared baselines, suggesting a favorable accuracy–efficiency trade-off for scalable representation learning without prohibitive computational cost.

E.5 Additional Eigenmode Dynamics Analysis

Beyond the ABIDE-specific analysis in the main text, we further probe the learned eigenmode structure on ADNI and PPMI on Fig. 11 to assess whether the Koopman dynamics remain stable and interpretable across clinical cohorts. In both datasets, the left panel visualizes the full spectrum of learned eigenvalues, which lie within the stable region of the complex plane with negligible imaginary

Table 6: Disease classification accuracy (%, ↑) across parcellation atlases on ADNI and PPMI. Results are reported as mean ± standard deviation across folds. Bold denotes the best result.

Atlas	ADNI	PPMI
AAL-116	$\mathbf{82.2}_{\pm 3.2}$	$\mathbf{82.5}_{\pm 11.2}$
Brainnetome-246	$76.6_{\pm 4.5}$	$79.8_{\pm 13.5}$
Shen-268	$80.5_{\pm 3.6}$	$82.0_{\pm 11.6}$
AICHA-384	$77.4_{\pm 4.2}$	$80.8_{\pm 12.8}$

Table 7: Sex classification across parcellation atlases and datasets. All models are trained on UKB with latent dimension $d = 768$. Results are reported as mean ± standard deviation across folds. Bold denotes the best result.

Atlas	UKB Acc.	UKB AUC	HCPA Acc.	HCPA AUC
AAL-116	$\mathbf{74.2}_{\pm 0.8}$	$\mathbf{0.81}_{\pm 0.01}$	$77.5_{\pm 2.9}$	$0.85_{\pm 0.03}$
Brainnetome-246	$71.8_{\pm 1.0}$	$0.79_{\pm 0.01}$	$76.0_{\pm 2.6}$	$0.84_{\pm 0.03}$
Shen-268	$73.5_{\pm 0.8}$	$\mathbf{0.81}_{\pm 0.01}$	$\mathbf{79.0}_{\pm 3.2}$	$\mathbf{0.86}_{\pm 0.02}$
AICHA-384	$71.2_{\pm 1.0}$	$0.78_{\pm 0.01}$	$77.2_{\pm 2.6}$	$0.85_{\pm 0.03}$

components, indicating predominantly non-oscillatory but well-conditioned latent dynamics. The middle panel then focuses on a subset of eigenmodes corresponding to the most inhibitory and excitatory states, while the rightmost brain maps summarize the spatial density of these effects, highlighting regions that exhibit particularly strong responses under these dominant modes. Notably, as shown in panel (c), the regions most strongly associated with ADNI are primarily located within the default mode network (DMN) and the ventral attention network (VAN), both of which are critically involved in high-level cognitive processing. In contrast, the regions highlighted in the PPMI cohort (panel (f)) are predominantly concentrated in sensorimotor areas. Importantly, these network-level patterns are highly consistent with the distinct clinical phenotypes of the two cohorts, reflecting cognitive impairment in Alzheimer's disease and motor dysfunction in Parkinson's disease, respectively. This network-specific differentiation highlights the ability of the proposed framework to disentangle distinct neurodegenerative signatures across cohorts. Taken together, these patterns suggest that BRAINMECH consistently learns stable dynamical primitives whose spatial footprints concentrate on disease-relevant networks, supporting the robustness and interpretability of the governing-law formulation across disorders and cohorts.

F Limitations and Future Works

First, the Koopman-structured dynamics are defined in a learned neural latent space and should not be interpreted as exact Koopman observables, causal neurobiological mechanisms, or biologically specified PDEs of the brain. Second, normalizing the fMRI repetition time (TR) to $\Delta t = 1$ simplifies physical sampling differences across cohorts; future work should explicitly model acquisition-specific time scales and scanner effects. Although TR metadata are typically available in fMRI datasets, our current experiments set $\Delta t = 1$ for all cohorts and therefore report normalized frame-time dynamics. A TR-calibrated extension could use $K = \exp(\mathrm{TR} \cdot A_{\mathrm{phys}})$, or equivalently rescale $A_{\mathrm{phys}} = A_{\mathrm{norm}}/\mathrm{TR}$ for physical-time interpretation.

Third, although we evaluate multiple cohorts, multi-site fMRI remains vulnerable to scanner, preprocessing, and site-related domain shifts, and stronger leave-one-site-out and prospective external validation are needed. Fourth, age is used as the primary continuous control variable because it is consistently available across cohorts, but other demographic, clinical, and acquisition covariates may also influence functional dynamics. Finally, the model is intended for research support and hypothesis generation rather than standalone clinical diagnosis.

Future work will extend the control/context pathway to incorporate multimodal covariates such as genetic risk, environment, medication, therapy, or task stimuli. Such intervention-aware modeling may enable controlled "what-if" analyses, but will require careful treatment of confounding, missing data, uncertainty calibration, fairness, and robustness under cohort shifts.

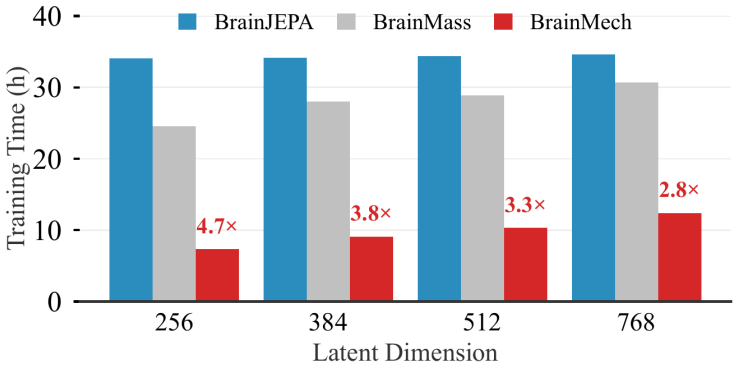

Figure 10: Pretraining efficiency analysis. BRAINMECH achieves substantially shorter pretraining time than competing foundation models, with an average 3.65× speedup.

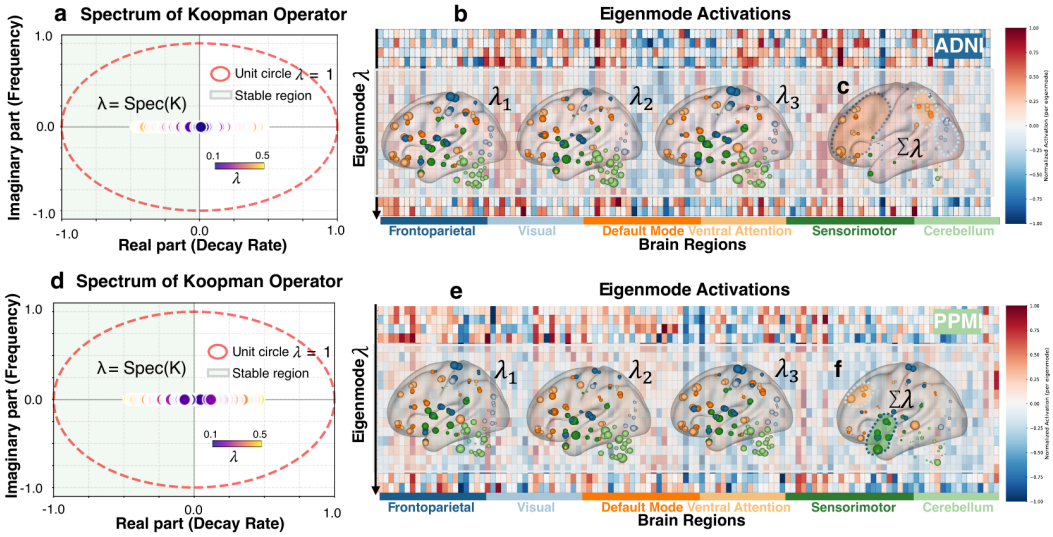

Figure 11: Eigenmode dynamics analysis of BRAINMECH on ADNI (top) and PPMI (bottom) datasets.

G Notation Summary

For ease of reference, we summarize the main notation used throughout the paper in Table 8.

Table 8: Summary of notation used in BRAINMECH.

Notation	Description	Shape / Domain
Observed fMRI data and sampling		
x_τ	Observed regional BOLD signal at discrete fMRI frame τ.	$\mathbb{R}^N$
$x_{1:T}$	Full observed BOLD sequence over T frames.	$\mathbb{R}^{T\times N}$
N	Number of brain regions/ROIs under a chosen atlas.	scalar
T	Sequence length of an fMRI segment.	scalar
S	Sliding-window stride used to extract fMRI segments.	scalar
τ	Discrete fMRI sampling index.	$\tau = 1, \ldots, T$
t_τ	Continuous sampling time corresponding to frame τ.	$t_\tau = \tau\Delta t$

Continued on next page

Notation	Description	Shape / Domain
Δt	Sampling interval or normalized frame-time scale. In experiments, $\Delta t = 1$.	scalar
A	Functional connectivity matrix computed from regional time series.	$\mathbb{R}^{N\times N}$
v	Phenotypic/context covariates used to condition dynamics, instantiated with age in this work.	task/context vector
Latent states and stochastic dynamics		
Z_t	Continuous-time latent stochastic process.	$\mathbb{R}^d$
z_τ	Discrete latent state at sampling time t_τ, i.e., $z_\tau := Z_{t_\tau}$.	$\mathbb{R}^d$
d	Latent state dimension.	scalar
u_t	Continuous-time control/context process modulating the latent dynamics.	$\mathbb{R}^{d_c}$
u_τ	Discrete control signal used on interval $[t_\tau, t_{\tau+1})$.	$\mathbb{R}^{d_c}$
d_c	Control/context dimension.	scalar
K	Continuous-time latent drift/generator matrix in the Koopman-structured SDE.	$\mathbb{R}^{d\times d}$
C	Control matrix mapping u_t into latent dynamics.	$\mathbb{R}^{d\times d_c}$
β	Bias term in the latent drift.	$\mathbb{R}^d$
W_t	Standard Brownian motion driving stochastic latent dynamics.	$\mathbb{R}^d$ or compatible dimension
$\Sigma_\theta(Z_t, u_t)$	Learnable diffusion coefficient modeling stochasticity and subject-level variability.	compatible with Z_t
$Q_\theta(z_\tau, u_\tau)$	Diffusion covariance, defined as $\Sigma_\theta(z_\tau, u_\tau)\Sigma_\theta(z_\tau, u_\tau)^\top$.	$\mathbb{R}^{d\times d}$
ξ_τ	Standard Gaussian noise used in Euler–Maruyama discretization.	$\mathcal{N}(0, I)$
$\Phi_{\Delta t}$	Exact sampled-time deterministic transition induced by K.	$\Phi_{\Delta t} = \exp(\Delta t K)$
Koopman and stability notation		
g	Observable function on the latent state space.	$g : \mathbb{R}^d \to \mathbb{R}$
$\mathcal{K}^t$	Stochastic Koopman semigroup acting on observables by conditional expectation.	operator
$(\mathcal{K}^t g)(z)$	Expected observable value at time t given initial state $Z_0 = z$.	$\mathbb{E}[g(Z_t) \mid Z_0 = z]$
$\mathcal{G}$	Infinitesimal generator of the stochastic Koopman semigroup.	differential operator
$f(z, u)$	Latent drift function.	$Kz + Cu + \beta$
$\nabla g, \nabla^2 g$	Gradient and Hessian of observable g.	vector / matrix
$\lambda_i(K)$	i-th eigenvalue of the continuous-time drift matrix K.	complex scalar
$\mathrm{Re}(\lambda_i(K))$	Real part of an eigenvalue of K, used for continuous-time stability interpretation.	scalar
$\alpha(K)$	Spectral abscissa of K.	$\max_i \mathrm{Re}(\lambda_i(K))$
$\mu_2(K)$	Logarithmic norm/dissipativity surrogate from the symmetric part of K.	$\lambda_{\max}\left(\frac{K+K^\top}{2}\right)$
$\rho(\cdot)$	Spectral radius operator. For continuous-time systems, it applies naturally to the sampled transition $\Phi_{\Delta t}$.	scalar
I	Identity matrix.	dimension-dependent
DPLR and low-rank parameterization		
$\hat{K}$	DPLR-parameterized approximation of the latent drift matrix.	$\mathbb{R}^{d\times d}$
k	Diagonal parameter vector of the DPLR drift matrix.	$\mathbb{R}^d$
$\mathrm{diag}(k)$	Diagonal drift component.	$\mathbb{R}^{d\times d}$
U, V	Low-rank factors in the DPLR correction $UV^\top$.	$\mathbb{R}^{d\times r}$

Continued on next page

Notation	Description	Shape / Domain
r	Rank of the low-rank Koopman correction.	scalar, $r \ll d$
R	Residual matrix approximated by the low-rank term.	$\hat{K} - \mathrm{diag}(k)$
$\hat{C}$	Low-rank parameterization of the control matrix.	$\mathbb{R}^{d \times d_c}$
C_1, C_2	Low-rank factors of $\hat{C} = C_1 C_2^\top$.	$C_1 \in \mathbb{R}^{d \times r_c}$, $C_2 \in \mathbb{R}^{d_c \times r_c}$
r_c	Rank of the low-rank control factorization.	$r_c = \min(r, d_c)$
Network components		
$f_{\mathrm{spa}}^{\theta_2}$	Spatial graph encoder that maps ROI activity and FC topology into node/graph-aware features.	neural network
$f_{\mathrm{tem}}^{\theta_3}$	Temporal encoder that models long-range dependencies across spatial embeddings.	neural network
$f_{\mathrm{fuse}}^{\theta_4}$	Fusion module combining episodic feedback and context embeddings into the control signal.	neural network
$f_{\mathrm{epi}}^{\theta_5}$	Episodic feedback encoder summarizing recent latent trajectory history.	neural network
$f_{\mathrm{ctx}}^{\theta_6}$	Context encoder mapping phenotypic covariates or task vectors into context embeddings.	neural network
$\mu_{\theta_1}(z_\tau)$	Decoder mean for the observation model.	$\mathbb{R}^N$
$\sigma^2_{\theta_1}(z_\tau)$	Decoder variance for the observation model.	positive scalar or vector
$p(x_\tau \mid z_\tau)$	Stochastic observation likelihood.	Gaussian density
$p(z_{\tau+1} \mid z_\tau, u_\tau)$	One-step latent transition density induced by Euler–Maruyama discretization.	Gaussian density
Training losses and evaluation		
$\hat{z}_{\tau+1}$	Koopman-predicted next latent state.	$\mathbb{R}^d$
$\hat{x}_{\tau+1}$	Forecasted next observation decoded from $\hat{z}_{\tau+1}$.	$\mathbb{R}^N$
$\mathcal{L}_{\mathrm{pre}}$	Forecasting negative log-likelihood loss in observation space.	scalar
$\mathcal{L}_{\mathrm{dyn}}$	Dynamics-consistency loss matching Koopman-predicted and encoder-derived latent states.	scalar
$\mathcal{L}_{\mathrm{stab}}$	Stability regularization term promoting bounded latent dynamics.	scalar
$\mathcal{L}$	Total training objective.	scalar
λ_1, λ_2	Loss weights for dynamics consistency and stability regularization.	scalars
ρ	Target spectral-radius threshold when using sampled-time spectral projection.	scalar
ϵ	Small positive stability margin or numerical constant, depending on context.	scalar
Mediation analysis		
X	Mean BOLD signal of a given ROI in mediation analysis.	scalar
M	Scalar mediator derived from pooled latent features.	scalar
Y	Downstream label, e.g., disease status.	scalar / categorical
C	Confounders in mediation analysis, such as age and sex.	vector
α_X	Regression coefficient for path $X \to M$.	scalar
β_M	Regression coefficient for path $M \to Y$ controlling for X and confounders.	scalar
β_X	Direct-effect coefficient for path $X \to Y$ controlling for M and confounders.	scalar
$\alpha_X \beta_M$	Indirect mediation effect.	scalar

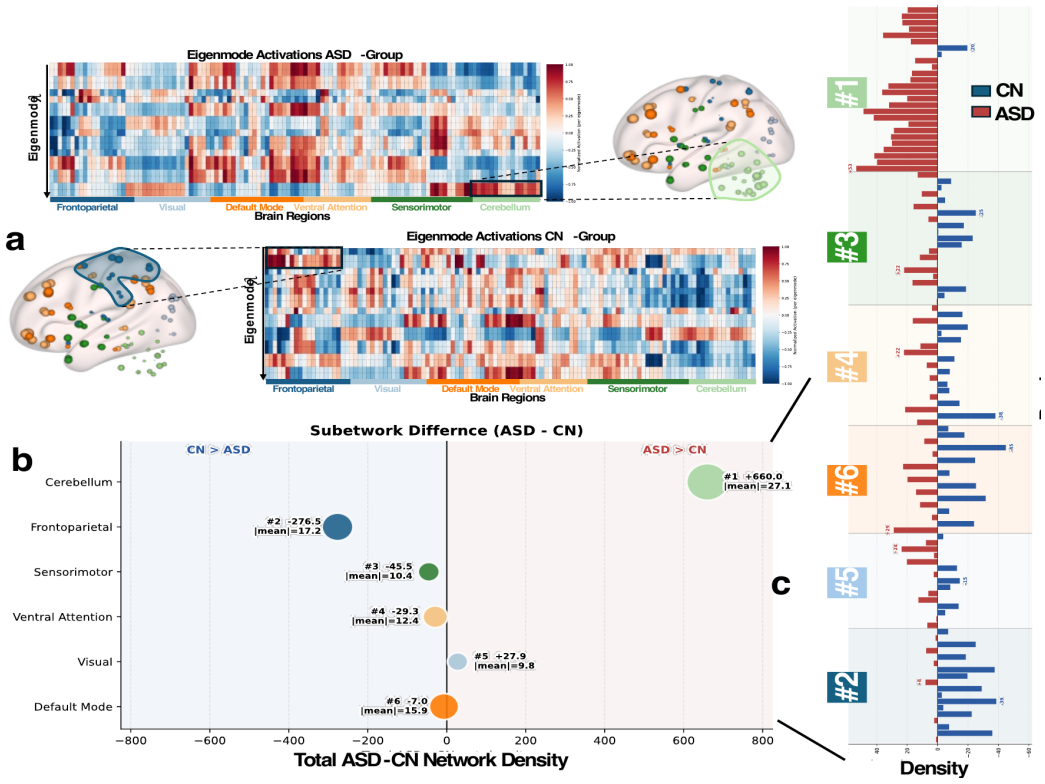

Figure 12: Group-level contrastive analysis of learned eigenmode dynamics between ASD and normal control (CN) cohorts. **(a)** Normalized eigenmode activation heatmaps and corresponding 3D spatial projections for the ASD (top) and CN (a) groups. The learned representations autonomously highlight distinct network engagements: the ASD cohort exhibits concentrated activation within the Cerebellum (green highlighting), whereas the CN cohort demonstrates dominant activation in the Frontoparietal network (blue highlighting). **(b)** Quantitative subnetwork-level differences in total net connectivity density (ASD - CN), revealing pronounced cerebellar hyperconnectivity (+660.0, ASD > CN) alongside significant frontoparietal hypoconnectivity (-276.5, CN > ASD) in the autism cohort. **(c)** Fine-grained, region-level breakdown of the connectivity differences within each subnetwork. This detailed view confirms that the macro-level subnetwork alterations observed in (b) are consistently driven by localized, region-specific density variations.