Brain-DiT: A Universal Multi-state fMRI Foundation Model with Metadata-Conditioned Pretraining

Junfeng Xia[1], Wenhao Ye[1], Xuanye Pan[1], Xinke Shen[1], Mo Wang[2,1,4*], and Quanying Liu[1,3,4*]

[1] Department of Biomedical Engineering, Southern University of Science and Technology, Shenzhen, China
[2] Department of Computer Science, University of Warwick, The UK
[3] Shenzhen Loop Area Institute, Shenzhen, China
[4] Omni-Intelligence, Shenzhen, China
12250099@mail.sustech.edu.cn;liuqy@sustech.edu.cn
*Co-corresponding authors

Abstract. Current fMRI foundation models primarily rely on a limited range of brain states and mismatched pretraining tasks, restricting their ability to learn generalized representations across diverse brain states. We present *Brain-DiT*, a universal multi-state fMRI foundation model pretrained on 349,898 sessions from 24 datasets spanning resting, task, naturalistic, disease, and sleep states. Unlike prior fMRI foundation models that rely on masked reconstruction in the raw-signal space or a latent space, *Brain-DiT* adopts metadata-conditioned diffusion pretraining with a Diffusion Transformer (DiT), enabling the model to learn multi-scale representations that capture both fine-grained functional structure and global semantics. Across extensive evaluations and ablations on 7 downstream tasks, we find consistent evidence that diffusion-based generative pretraining is a stronger proxy than reconstruction or alignment, with metadata-conditioned pretraining further improving downstream performance by disentangling intrinsic neural dynamics from population-level variability. We also observe that downstream tasks exhibit distinct preferences for representational scale: ADNI classification benefits more from global semantic representations, whereas age/sex prediction comparatively relies more on fine-grained local structure. Code and parameters are available at https://github.com/REDMAO4869/Brain-DiT.

Keywords: fMRI · Foundation Models · Diffusion Transformer · Generative Modeling

1 Introduction

The human brain is a dynamical system that continuously transits between states to flexibly support adaptive behaviors. Capturing universal representations across this state spectrum is essential for decoding brain function, yet traditional approaches often fail to generalize across different brain states. Inspired

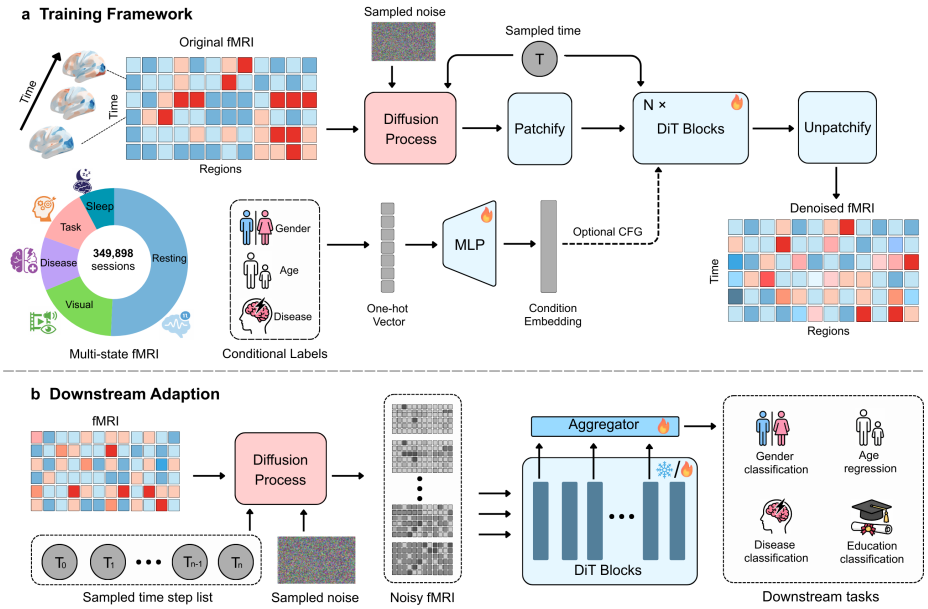

Fig. 1. Brain-DiT framework: pretraining and downstream adaptation. (a) **Pretraining.** An ROI time-series window is noised at a random timestep t and denoised by a conditional/unconditional DiT using v-prediction. (b) **Downstream adaptation.** Multiple noisy views at selected timesteps are processed by the shared pretrained DiT. Token features from selected layers are aggregated to form a representation for downstream classification or regression.

by the transformative impact of self-supervised learning in computer vision and natural language processing, neuroimaging (such as fMRI, EEG and MEG data) is undergoing a paradigm shift: moving from task-specific models to large-scale foundation models pretrained on massive unlabeled datasets.

Recent efforts on fMRI foundation models have explored diverse proxy tasks, including masked autoencoding (MAE) [21,20,27], joint-embedding predictive architectures (JEPA) [4], contrastive learning [30], and diffusion-based generative modeling of fMRI dynamics [29]. However, existing fMRI foundation models exhibit limited robustness across diverse brain states, primarily due to three fundamental limitations. First, the pretraining distribution is restricted: most models are trained exclusively on resting-state data, limiting their exposure to the full spectrum of brain dynamics [20,30,4,26].

Second, the proxy tasks employed are often mismatched to learn universal representations: raw reconstruction may overfit to noise in low-SNR fMRI data [20,21], while latent-space reconstruction requires careful architectural design to avoid representational collapse [26,4]. Consequently, substantial post-training adaptation is often required to achieve acceptable downstream performance. A third, largely unexplored limitation is the underutilization of clinical

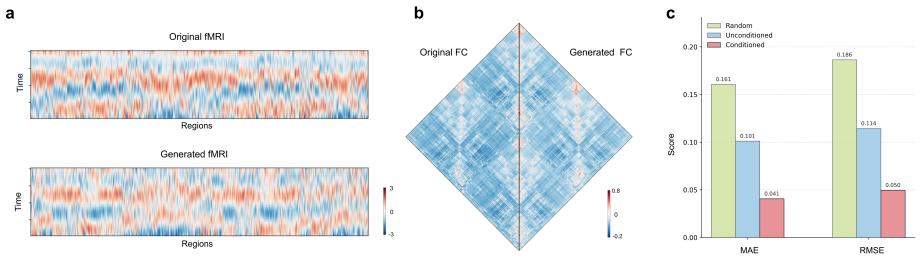

Fig. 2. Metadata-conditioned generation of Brain-DiT. (a) **Original vs. generated fMRI signals.** Generated fMRI signals exhibit similar spatiotemporal patterns to the original fMRI. (b) **Group-level FC comparison.** Empirical and synthetic group-level FC matrices for 200 metadata-conditioned virtual ASD subjects exhibit similar large-scale connectivity organization. (c) **FC error metrics.** MAE and RMSE (lower is better) across three settings: random signal, unconditional generation, and metadata-conditioned generation. All results are computed on the ABIDE dataset.

and demographic metadata during pretraining, which hinders the disentanglement of intrinsic neural dynamics from population-level variability.

Diffusion models, which learn to reverse a gradual noising process, offer a promising solution to these challenges. First, by modeling the full data distribution rather than optimizing discriminative objectives, diffusion models characterize the complete probability manifold of neural signals, enabling the learning of richer, more generalizable features [16]. Second, the iterative denoising process provides a natural hierarchy of representations, with early timesteps capturing fine-grained local structure and later timesteps encoding abstract global semantics. Most importantly, diffusion models natively support conditional generation, allowing seamless integration of biological and clinical metadata as conditioning signals—thereby reducing the "contextual blindness" of prior paradigms.

Here, we present **Brain-DiT**, a universal multi-state foundation model that leverages these advantages through a metadata-conditioned Diffusion Transformer. Our main contributions are:

- **Multi-state generative pretraining paradigm.** We introduce the first fMRI foundation model trained jointly on multiple brain states, unifying 349,898 sessions from 24 datasets spanning resting, task, naturalistic, disease, and sleep conditions under a single generative objective.
- **Metadata-conditioned pretraining.** We are the first to integrate demographic and clinical metadata as explicit conditioning signals during pretraining, enabling the model to disentangle intrinsic neural dynamics from population variability.
- **Multi-scale feature aggregation.** We treat the denoising trajectory as a hierarchical feature extractor and aggregate representations across timesteps and layers, yielding embeddings that capture both fine-grained functional structure and abstract semantic information.

Table 1. Downstream performance on in-distribution (ID) and out-of-distribution (OOD) tasks. Top: ID; bottom: OOD. Brain-DiT$_{\text{AAL424}}$ and Brain-DiT$_{\text{uncond}}$ are ablations. Best and runner-up results are highlighted by gray shading and by underlining, respectively. * indicates $p < 0.05$ vs. the best baseline (BrainLM/Brain-JEPA/BrainMass); no tests are performed among our variants.

Model \ Dataset	HCP	ABIDE-AGE		SALD	
	ACC%↑	MSE↓	R↑	MSE↓	R↑
BrainLM	62.71 ± 4.43	0.91 ± 0.01	0.24 ± 0.02	0.68 ± 0.06	0.62 ± 0.06
Brain-JEPA	69.97 ± 2.73	0.98 ± 0.07	0.17 ± 0.02	1.13 ± 0.11	0.30 ± 0.06
BrainMass	67.65 ± 1.02	0.70 ± 0.04	0.50 ± 0.04	0.70 ± 0.08	0.63 ± 0.06
Brain-DiT$_{\text{AAL424}}$	$\underline{81.71 \pm 0.77}$	0.80 ± 0.02	0.59 ± 0.01	0.53 ± 0.02	0.65 ± 0.01
Brain-DiT$_{\text{uncond}}$	80.98 ± 1.22	0.77 ± 0.02	$\underline{0.60 \pm 0.02}$	$\underline{0.39 \pm 0.02}$	$\underline{0.75 \pm 0.01}$
Brain-DiT	$82.71 \pm 0.79^{*}$	$\underline{0.75 \pm 0.01}$	$0.61 \pm 0.01^{*}$	$0.33 \pm 0.01^{*}$	$0.79 \pm 0.01^{*}$

Model \ Dataset	PPMI	ADHD	NKI-EDU	NKI-AGE	
	ACC%↑	ACC%↑	ACC%↑	MSE↓	R↑
BrainLM	54.86 ± 1.20	58.37 ± 1.04	60.46 ± 0.73	0.50 ± 0.02	0.68 ± 0.01
Brain-JEPA	47.91 ± 3.61	54.89 ± 4.98	49.21 ± 2.15	1.03 ± 0.08	0.33 ± 0.03
BrainMass	58.68 ± 4.21	57.21 ± 3.45	62.36 ± 1.24	0.60 ± 0.06	0.62 ± 0.04
Brain-DiT$_{\text{AAL424}}$	58.99 ± 1.46	$\underline{62.38 \pm 0.67}$	60.18 ± 1.81	0.44 ± 0.04	0.77 ± 0.02
Brain-DiT$_{\text{uncond}}$	$\underline{60.99 \pm 2.14}$	62.06 ± 0.36	$\underline{69.78 \pm 3.29}$	$\underline{0.40 \pm 0.02}$	$\underline{0.80 \pm 0.01}$
Brain-DiT	$61.15 \pm 1.84^{*}$	$62.53 \pm 1.56^{*}$	$69.92 \pm 0.02^{*}$	$0.38 \pm 0.01^{*}$	$0.81 \pm 0.01^{*}$

2 Methods

2.1 Problem Formulation and Overview

An fMRI sample is denoted by $\mathbf{x} \in \mathbb{R}^{T \times N}$, where T is the number of time points and N is the number of ROIs. For patch-based Transformer modeling, we reshape $\mathbf{x}$ into an ROI×time tensor $\mathbf{x}_0 \in \mathbb{R}^{1 \times N \times T}$ (channel × ROI × time). Our pretraining objective is to learn a diffusion-based generative model for $p(\mathbf{x})$ by training a DiT denoiser, optionally conditioned on subject-level metadata.

2.2 Generative Diffusion Pretraining

Diffusion and Training Process As shown in Fig. 1 (a), we adopt a DDPM forward process with a cosine noise schedule over S diffusion steps (S=1000). For timestep $t \in \{1, \ldots, S\}$, we sample Gaussian noise $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and obtain:

$$\mathbf{x}_t \;=\; \alpha_t \mathbf{x}_0 \;+\; \sigma_t \boldsymbol{\epsilon}, \tag{1}$$

where $\alpha_t = \sqrt{\bar{\alpha}_t}$ and $\sigma_t = \sqrt{1 - \bar{\alpha}_t}$ are precomputed schedule coefficients.

Given the noised input $\mathbf{x}_t \in \mathbb{R}^{1\times N\times T}$ and timestep t, we learn a DiT directly in raw signal space. DiT first patchifies $\mathbf{x}_t$ using a 2D convolution with patch size p and stride p, producing a token sequence of length

$$L \;=\; (N/p)\cdot(T/p), \tag{2}$$

with embedding dimension D. Tokens are processed by K DiT blocks, and a linear head followed by an invertible unpatchify operation maps tokens back to a dense prediction in $\mathbb{R}^{1\times N\times T}$. We then train the model with v-prediction, where

$$\mathbf{v} \;=\; \alpha_t\boldsymbol{\epsilon} \;-\; \sigma_t\mathbf{x}_0. \tag{3}$$

The DiT is trained to predict $\hat{\mathbf{v}}_\theta(\mathbf{x}_t, t, \mathbf{c})$ by minimizing:

$$\mathcal{L}_{\text{diff}} \;=\; \mathbb{E}_{\mathbf{x}_0,t,\boldsymbol{\epsilon}}\left[\|\hat{\mathbf{v}}_\theta(\mathbf{x}_t,t,\mathbf{c}) - \mathbf{v}\|_2^2\right]. \tag{4}$$

Metadata-Conditioned Pretraining Brain-DiT uses optional explicit metadata conditioning during pretraining. Given subject metadata $\mathbf{c}$ (e.g., age, sex, diagnosis), we form a tabular vector with continuous values plus observation masks and one-hot diagnosis indicators, then map it to

$$\mathbf{y}_c = f_\phi(\mathbf{c}) \in \mathbb{R}^{d_c}. \tag{5}$$

With timestep embedding $\mathbf{y}_t \in \mathbb{R}^{d_c}$, the condition is

$$\mathbf{y} = \mathbf{y}_t + \mathbf{y}_c, \tag{6}$$

which is injected into every DiT block via AdaLN-Zero to condition the denoising process. During training, CFG-style regularization replaces $\mathbf{y}_c$ with an unconditional embedding with probability p_{drop}; samples with missing metadata are either dropped or treated as unconditional according to the missing-label policy.

2.3 Multi-scale Feature Aggregation

Feature Extraction Given a window $\mathbf{x}$, we construct $\mathbf{x}_0 \in \mathbb{R}^{1\times N\times T}$ and choose a timestep list $\mathcal{T} = \{t_1, \ldots, t_M\}$ (e.g., $\{0, 50, 100, 150\}$) and a set of captured Transformer blocks $\mathcal{L}$ (e.g., the last few blocks). For each $t \in \mathcal{T}$, we generate $\mathbf{x}_t$ using the same forward process:

$$\mathbf{x}_t \;=\; \alpha_t\mathbf{x}_0 \;+\; \sigma_t\boldsymbol{\epsilon}. \tag{7}$$

In Fig. 1 (b), we pass $\mathbf{x}_t$ through the pretrained DiT and extract features from selected blocks $\ell \in \mathcal{L}$. For each (t, ℓ), we mean-pool $\mathbf{H}_t^{(\ell)} \in \mathbb{R}^{B\times L\times D}$ to obtain

$$\mathbf{e}_{t,\ell} = \text{MeanPool}\Big(\mathbf{H}_t^{(\ell)}\Big) \in \mathbb{R}^{B\times D}. \tag{8}$$

We then concatenate features across timesteps and layers:

$$\mathbf{E} = \text{Concat}_{t\in\mathcal{T},\,\ell\in\mathcal{L}}\mathbf{e}_{t,\ell} \in \mathbb{R}^{B\times Q\times D}, \quad Q = |\mathcal{T}||\mathcal{L}|. \tag{9}$$

Table 2. Downstream performance with a frozen backbone. Best results are shown by gray shading; underlined values denote the runner-up. * indicates $p < 0.05$ vs. the best baseline.

Model \ Dataset	ADHD		SALD		NKI-EDU		NKI-AGE	
	ACC%↑		MSE↓		ACC%↑		MSE↓	
BrainLM	53.57 ± 2.77		0.80 ± 0.01		47.95 ± 1.12		0.76 ± 0.02	
Brain-JEPA	52.24 ± 2.66		0.76 ± 0.01		38.69 ± 0.23		0.95 ± 0.02	
BrainMass	54.89 ± 3.38		0.78 ± 0.03		54.89 ± 3.38		0.72 ± 0.01	
$\text{SlimBrain}_{\text{jepa}}$	$\underline{55.74 \pm 6.51}$		0.72 ± 0.01		53.83 ± 6.62		0.89 ± 0.02	
$\text{SlimBrain}_{\text{mae}}$	55.72 ± 1.01		$\underline{0.71 \pm 0.07}$		$\underline{60.59 \pm 2.77}$		$\underline{0.68 \pm 0.01}$	
Brain-DiT	$62.04 \pm 1.22^*$		$0.67 \pm 0.01^*$		$62.80 \pm 0.63^*$		$0.58 \pm 0.03^*$	

Aggregation Given $\mathbf{E} \in \mathbb{R}^{B \times Q \times D}$ with $Q = |\mathcal{T}||\mathcal{L}|$, we fuse multi-timestep, multi-layer features using a query-based aggregator module. A learnable query $\mathbf{q} \in \mathbb{R}^{1 \times 1 \times D}$ is broadcast across the batch and used for cross-attention:

$$\mathbf{z},\ \mathbf{w} = \mathrm{Attn}(\mathbf{q}, \mathbf{E}, \mathbf{E}), \quad \mathbf{z} \in \mathbb{R}^{B \times D},\ \mathbf{w} \in \mathbb{R}^{B \times Q}. \tag{10}$$

The fused representation $\mathbf{z}$ is fed to the downstream prediction head, while $\mathbf{w}$ provides the relative importance of each (t, ℓ) pair.

3 Experiments and Results

3.1 Experimental Setup

Datasets and Preprocessing. We utilize 24 multi-state datasets covering resting, task, disease, naturalistic, and sleep states, partitioned into ID and OOD sets. The ID pretraining corpus comprises 21 datasets: 9 resting-state (HCP [25], CHCP [7], ABCD [2], PIOP1 [23], PIOP2 [23], ISYB [6], BHRC [19], SALD [28], SLIM [12]), 2 disease (ABIDE [3], ADNI [11]), and 10 specialized states including task-based (HCP Task [25], CHCP Task [7], N&N [17]), naturalistic (StudyForrest [9], Emo Film [15], NSD [1], Things [10], CineBrain [5], HCP Movie [25]), and sleep fMRI [8]. To validate generalization, the NKI [24], ADHD-200 [14], and PPMI [13] datasets are strictly reserved for OOD evaluation. All fMRI volumes were resampled to 2 mm isotropic resolution and cropped/padded to $96 \times 96 \times 96$, with time series interpolated to a uniform TR of 0.72 s (cubic B-spline). After global z-score normalization, voxel signals were parcellated into ROI time series using the Schaefer-1000 [22] and AAL424 atlases [18], with the resulting ROI time series further segmented into fixed-length 40-frame temporal windows.

Baselines and Evaluation. We compare Brain-DiT with ROI and voxel baselines (BrainLM, Brain-JEPA, BrainMass, $\text{SlimBrain}_{\text{mae}}$, $\text{SlimBrain}_{\text{jepa}}$), covering MAE, JEPA, and contrastive learning. Evaluation metrics include Accu-

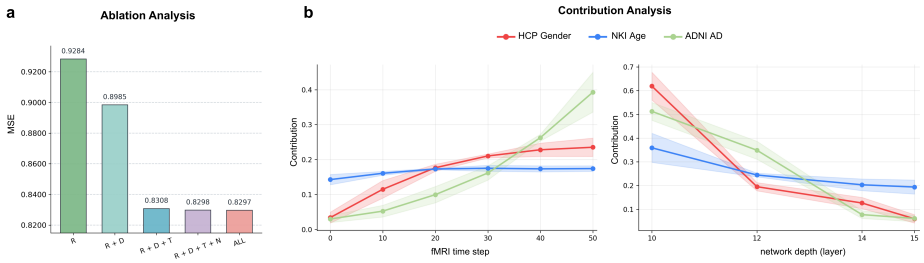

Fig. 3. State diversity ablation and contribution analysis. (a) **State diversity ablation.** Mixtures of multi-state fMRI data are used during pretraining with a fixed pretraining budget (∼30k sessions): R (resting), +D (disease), +T (task), +N (naturalistic), and ALL (the full multi-state corpus). (b) **Contribution analysis.** Aggregator weights show contributions across timesteps (left) and network depth (right), reflecting global (later/noisier) vs. local (earlier/cleaner) features.

racy for classification, and Mean Squared Error (MSE) and Pearson's r for regression. Generative fidelity is quantified by comparing synthesized and empirical functional connectivity (FC) patterns via MSE and RMSE.

Implementation Details. Brain-DiT uses a DiT backbone (d_{model}=1024, 16 layers, 8 heads, patch size 4) and is trained on Schaefer-1000 ROI windows (T=40) with a cosine DDPM schedule (1,000 steps) and a v-prediction objective. We train for 150 epochs on 6 NVIDIA A800 (80GB) GPUs with AdamW (LR 1e−4, WD 1e−4; batch size 36/GPU, mixed precision) and inject age/sex/disease via AdaLN (d_{cond}=256) with optional 10% condition dropout.

3.2 Generative Fidelity

Before evaluating downstream tasks, we first verify that Brain-DiT learns meaningful representations through generative experiments. As shown in Fig. 2 (a), we generated fMRI signals conditioned on specific subject metadata from the ABIDE dataset. The generated BOLD responses exhibit a high temporal correlation with the empirical ground truth, indicating that the model successfully generates individual-level neural dynamics guided by metadata. To further assess population-level fidelity, we generated fMRI signals for 200 virtual ASD subjects and computed group-level FC. The synthetic FC topography closely matches the empirical FC of the real cohort, and quantitative metrics (MSE/RMSE) show that conditional generation outperforms both unconditional and random-conditioning baselines. These results suggest that Brain-DiT learns subject-generalizable representations and supports controllable conditional generation.

3.3 Downstream Task Performance

As shown in Table 1, Brain-DiT achieves the best or competitive performance across ID and OOD benchmarks, with improvements that are significant against

the best baseline. For fair comparison, we report two variants: Brain-DiT$_{\text{AAL424}}$, aligned with the AAL424 atlas used by BrainLM, and Brain-DiT$_{\text{uncond}}$, an unconditional ablation.

Even the ablated variants remain strong. For example, Brain-DiT$_{\text{AAL424}}$ achieves 81.71% accuracy on HCP, outperforming the best baseline (Brain-JEPA, 69.97%). Moreover, the full model outperforms Brain-DiT$_{\text{uncond}}$ across benchmarks, highlighting the benefit of metadata conditioning for generalization.

3.4 Representation Quality

We assess representation quality by freezing the pretrained backbone and training a lightweight linear head as shown in Table 2. Despite using ROI-level inputs, Brain-DiT surpasses **voxel-level** MAE/JEPA baselines significantly; for instance, it achieves 62.80% on NKI-EDU versus 60.59% for SlimBrain$_{\text{mae}}$. This advantage persists under distribution shift, with Brain-DiT reaching 62.04% accuracy on ADHD-200 and 0.67 MSE on SALD.

3.5 Interpretability Analysis

State diversity drives downstream gains. To isolate the effect of state diversity, we fix the pretraining set size at 30k sessions and vary only the state mixture. As shown in Fig. 3 (a), downstream performance improves consistently as state coverage increases, suggesting that the gains mainly come from broader brain-state coverage rather than a larger sample size.

Global–local contributions across timesteps and depth.

We analyze step- and depth-wise contributions using normalized aggregator weights $w_{t,\ell}$, where larger t indicates higher noise. Later timesteps emphasize coarse, network-level structure, whereas earlier steps preserve finer-grained variations. As shown in Fig. 3 (b), ADNI shows the strongest preference for later timesteps, consistent with system-level network disruptions in Alzheimer's disease; age prediction is more evenly distributed, consistent with age effects spanning both global and local connectivity, with gender in between. Across depth, contributions concentrate in earlier-to-intermediate layers and decay in deeper blocks, with a sharper drop for disease classification, consistent with prior observations.

4 Conclusion

We present Brain-DiT, a universal multi-state fMRI foundation model with metadata-conditioned pretraining to capture complex brain dynamics and multi-scale aggregation to improve downstream generalization. Brain-DiT outperforms representative baselines on ID and OOD benchmarks and remains strong under frozen-backbone transfer. Conditional generation reproduces realistic subject-level fMRI dynamics and preserves cohort-level FC patterns under metadata conditioning. Contribution analysis suggests that multi-state pretraining yields

multi-scale representations, with diffusion timesteps and layers providing complementary signals for downstream tasks. Future work will extend Brain-DiT with richer metadata and broader state coverage, while improving scaling efficiency for longer sequences and higher-resolution inputs.

Acknowledgments. This work was supported by National Science and Technology Major Project (2021ZD0200500),National Natural Science Foundation of China (3254100307, 62472206), National Key R&D Program of China (2025YFC3410000), Shenzhen Science and Technology Innovation Committee (RCYX20231211090405003, JCYJ20220818100213029), Guangdong Basic and Applied Basic Research Foundation (2026B1515020099), Guangdong S&T Program (2026B0101110003), Shanghai Municipal Special Program for Basic Research on General AI Foundation Models (2025SHZDZX 026D05), GuangDong Basic and Applied Basic Research Foundation (2025A1515011645 to ZC.L.), GuangDong Basic and Applied Basic Research Foundation (2026A1515010121 to XK.S.), Shenzhen Doctoral Startup Project (RCBS20231211090748082 to XK.S.), Shenzhen Loop Area Institute (FPF10120250012), and the open research fund of the Guangdong Provincial Key Laboratory of Mathematical and Neural Dynamical Systems, the Center for Computational Science and Engineering at Southern University of Science and Technology, Shenzhen Key Laboratory of Smart Healthcare Engineering.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Allen, E.J., St-Yves, G., Wu, Y., Breedlove, J.L., Prince, J.S., Dowdle, L.T., Nau, M., Caron, B., Pestilli, F., Charest, I., et al.: A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. Nature neuroscience **25**(1), 116–126 (2022)
2. Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., et al.: The adolescent brain cognitive development (abcd) study: imaging acquisition across 21 sites. Developmental cognitive neuroscience **32**, 43–54 (2018)
3. Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry **19**(6), 659–667 (2014)
4. Dong, Z., Li, R., Wu, Y., Nguyen, T.T., Chong, J., Ji, F., Tong, N., Chen, C., Zhou, J.H.: Brain-jepa: Brain dynamics foundation model with gradient positioning and spatiotemporal masking. Advances in Neural Information Processing Systems **37**, 86048–86073 (2024)
5. Gao, J., Liu, Y., Yang, B., Feng, J., Fu, Y.: Cinebrain: A large-scale multi-modal brain dataset during naturalistic audiovisual narrative processing. arXiv preprint arXiv:2503.06940 (2025)
6. Gao, P., Dong, H.M., Liu, S.M., Fan, X.R., Jiang, C., Wang, Y.S., Margulies, D., Li, H.F., Zuo, X.N.: A chinese multi-modal neuroimaging data release for increasing diversity of human brain mapping. Scientific Data **9**(1), 286 (2022)

7. Ge, J., Yang, G., Han, M., Zhou, S., Men, W., Qin, L., Lyu, B., Li, H., Wang, H., Rao, H., et al.: Increasing diversity in connectomics with the chinese human connectome project. Nature Neuroscience **26**(1), 163–172 (2023)
8. Gu, Y., Han, F., Sainburg, L.E., Schade, M.M., Liu, X.: "simultaneous eeg and fmri signals during sleep from humans" (2026). https://doi.org/doi:10.18112/openneuro.ds003768.v1.0.13
9. Hanke, M., Adelhöfer, N., Kottke, D., Iacovella, V., Sengupta, A., Kaule, F.R., Nigbur, R., Waite, A.Q., Baumgartner, F., Stadler, J.: A studyforrest extension, simultaneous fmri and eye gaze recordings during prolonged natural stimulation. Scientific data **3**(1), 160092 (2016)
10. Hebart, M.N., Contier, O., Teichmann, L., Rockter, A.H., Zheng, C.Y., Kidder, A., Corriveau, A., Vaziri-Pashkam, M., Baker, C.I.: Things-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. Elife **12**, e82580 (2023)
11. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al.: The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine **27**(4), 685–691 (2008)
12. Liu, W., Wei, D., Chen, Q., Yang, W., Meng, J., Wu, G., Bi, T., Zhang, Q., Zuo, X.N., Qiu, J.: Longitudinal test-retest neuroimaging data from healthy young adults in southwest china. Scientific data **4**(1), 170017 (2017)
13. Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., et al.: The parkinson progression marker initiative (ppmi). Progress in neurobiology **95**(4), 629–635 (2011)
14. Milham, M.P., Fair, D., Mennes, M., Mostofsky, S.H.: The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. Frontiers in Systems Neuroscience **Volume 6 - 2012** (2012)
15. Morgenroth, E., Moia, S., Vilaclara, L., Fournier, R., Muszynski, M., Ploumitsakou, M., Almató-Bellavista, M., Vuilleumier, P., Van De Ville, D.: Emo-film: a multimodal dataset for affective neuroscience using naturalistic stimuli. Scientific Data **12**(1), 684 (2025)
16. Mukhopadhyay, S., Gwilliam, M., Yamaguchi, Y., Agarwal, V., Padmanabhan, N., Swaminathan, A., Zhou, T., Ohya, J., Shrivastava, A.: Do text-free diffusion models learn discriminative visual representations? In: European Conference on Computer Vision. pp. 253–272. Springer (2024)
17. Nakai, T., Nishimoto, S.: Quantitative models reveal the organization of diverse cognitive functions in the brain. Nature communications **11**(1), 1142 (2020)
18. Nemati, S., Akiki, T.J., Roscoe, J., Ju, Y., Averill, C.L., Fouda, S., Dutta, A., McKie, S., Krystal, J.H., Deakin, J.W., et al.: A unique brain connectome fingerprint predates and predicts response to antidepressants. IScience **23**(1) (2020)
19. de Oliveira, I.P., Fernandéz, A.C., Salum, G.A., Gadelha, A., Pan, P.M., Miguel, E.C., Mograbi, D.C., Bado, P.: Longitudinal patterns of disordered eating behaviors in children and adolescents from the brazilian high-risk cohort study for mental conditions. Brazilian Journal of Psychiatry **47**, e20243867 (2025)
20. Ortega Caro, J., de Oliveira Fonseca, A.H., Rizvi, S., Rosati, M., Averill, C., Cross, J., Mittal, P., Zappala, E., Dhodapkar, R., Abdallah, C., et al.: Brainlm: A foundation model for brain activity recordings. In: International Conference on Learning Representations. vol. 2024, pp. 565–576 (2024)

21. Qu, Y., Xia, J., Jian, X., Li, W., Peng, K., Liang, Z., Wu, H., Liu, Q.: Uncovering cognitive taskonomy through transfer learning in masked autoencoder-based fmri reconstruction. In: International Workshop on Human Brain and Artificial Intelligence. pp. 35–50. Springer (2024)
22. Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.: Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. Cerebral cortex **28**(9), 3095–3114 (2018)
23. Snoek, L., van der Miesen, M.M., Beemsterboer, T., Van Der Leij, A., Eigenhuis, A., Steven Scholte, H.: The amsterdam open mri collection, a set of multimodal mri datasets for individual difference analyses. Scientific data **8**(1), 85 (2021)
24. Telesford, Q.K., Gonzalez-Moreira, E., Xu, T., Tian, Y., Colcombe, S.J., Cloud, J., Russ, B.E., Falchier, A., Nentwich, M., Madsen, J., et al.: An open-access dataset of naturalistic viewing using simultaneous eeg-fmri. Scientific Data **10**(1), 554 (2023)
25. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.M.H., et al.: The WU-Minn Human Connectome Project: An Overview. Neuroimage **80**, 62–79 (2013)
26. Wang, M., Xia, J., Ye, W., Liu, E., Peng, K., Feng, J., Liu, Q., Wen, H.: SLIM-Brain: A Data-and Training-Efficient Foundation Model for fMRI Data Analysis. arXiv preprint arXiv:2512.21881 (2025)
27. Wang, M., Ye, W., Xia, J., Zhang, J., Pan, X., Xu, M., Deng, H., Wen, H., Liu, Q.: Omni-fmri: A universal atlas-free fmri foundation model. arXiv preprint arXiv:2601.23090 (2026)
28. Wei, D., Zhuang, K., Ai, L., Chen, Q., Yang, W., Liu, W., Wang, K., Sun, J., Qiu, J.: Structural and functional brain scans from the cross-sectional Southwest University adult lifespan dataset. Scientific Data **5**(1), 180134 (Jul 2018)
29. Xia, J., Ye, W., Zhang, J., Pan, X., Wang, M., Liu, Q.: Brainworld: A structural-prior-conditioned generative model for whole-brain 4d fmri dynamics. arXiv preprint arXiv:2606.17742 (2026)
30. Yang, Y., Ye, C., Su, G., Zhang, Z., Chang, Z., Chen, H., Chan, P., Yu, Y., Ma, T.: Brainmass: Advancing brain network analysis for diagnosis with large-scale self-supervised learning. IEEE transactions on medical imaging **43**(11), 4004–4016 (2024)