LATENT BRAIN STATE-GUIDED MIXTURE-OF-EXPERTS FOR EEG-TO-FMRI RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Translating electroencephalography (EEG) into functional magnetic resonance imaging (fMRI) signals could enable region-resolved estimation of blood-oxygen-level-dependent (BOLD) dynamics and functional connectivity using only EEG at inference, without costly fMRI acquisition. However, this mapping is challenging because neurovascular coupling is temporally delayed and varies across brain regions, while existing approaches often fail to separate shared hemodynamic dynamics from region-specific decoding. To address this challenge, we propose Latent Brain State-Guided Mixture-of-Experts (**E2fMoE**), inspired by neurovascular coupling in which neural activity gives rise to delayed, region-dependent BOLD responses. Accordingly, E2fMoE performs shared-state inference before regional decoding. Specifically, a state-space mixer transforms EEG tokens into slowly evolving latent representations, while region-of-interest (ROI) queries extract region-specific information and route it through shared and specialized experts. Residual prediction and low-rank connectivity-aware calibration further refine reconstruction. Extensive experiments across four simultaneous EEG-fMRI datasets show that E2fMoE outperforms representative EEG representation models and EEG-to-fMRI baselines in temporal reconstruction and functional-connectivity preservation. Code is available in the supplementary material.

1 INTRODUCTION

Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) provide complementary views of brain activity. EEG captures electrophysiological fluctuations at millisecond resolution and supports relatively accessible, long-duration recording, including repeated longitudinal measurements, whereas fMRI measures spatially distributed blood-oxygen-level-dependent (BOLD) activity across cortical and subcortical systems Ogawa et al. (1990); Warbrick (2022); Ebrahimzadeh et al. (2022). Learning EEG-to-fMRI reconstruction from simultaneously acquired data could therefore enable estimation of regional BOLD dynamics and functional connectivity (FC) from EEG alone, offering a scalable solution for large-scale neuroimaging studies and clinical settings where fMRI acquisition is unavailable or impractical Li et al. (2024; 2026).

This cross-modal mapping remains challenging due to the complex temporal, spatial, and functional relationships between EEG and BOLD signals. As illustrated in Figure 1, BOLD activity is not an instantaneous transformation of EEG, but a delayed and temporally integrated consequence of neural activity mediated by neurovascular processes Logothetis et al. (2001); Georgopoulos et al. (2024); Mangini et al. (2025). EEG-informed fMRI analyses commonly address this mismatch using predefined EEG features and assumed hemodynamic response functions Warbrick (2022); Ebrahimzadeh et al. (2022), whereas recent EEG representation models learn richer temporal and spectral features directly from data Song et al. (2023); Yang et al. (2023a); Jiang et al. (2024); Wang et al. (2024). However, richer EEG representations alone do not explicitly model the latent dynamics linking preceding electrophysiological activity to subsequent BOLD responses.

Beyond temporal mismatch, the EEG-to-BOLD mapping is inherently region dependent, as neurovascular coupling and hemodynamic responses vary across brain regions, individuals, and cognitive states Handwerker et al. (2004); Lindquist et al. (2009); Georgopoulos et al. (2024); Mangini et al. (2025). While ROI-specific models can account for such heterogeneity, they discard shared structures and scale poorly, whereas fully shared decoders may impose overly uniform mappings

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

across regions. Recent EEG-to-fMRI models predict BOLD signals from EEG or reconstruct multiple regions Li et al. (2024; 2026), but they generally do not disentangle shared cross-region dynamics from ROI-specific signal generation.

Furthermore, accurate prediction of individual ROI trajectories does not necessarily preserve the multivariate organization of brain activity. Pointwise losses and temporal correlations may produce plausible regional signals while distorting covariance structures and functional connectivity (FC) patterns. Since coordinated BOLD fluctuations define large-scale functional systems Friston (1994); Biswal et al. (1995); Damoiseaux et al. (2006); Fox & Raichle (2007); Power et al. (2011); Smith et al. (2011), reconstruction should preserve both regional dynamics and interregional dependencies.

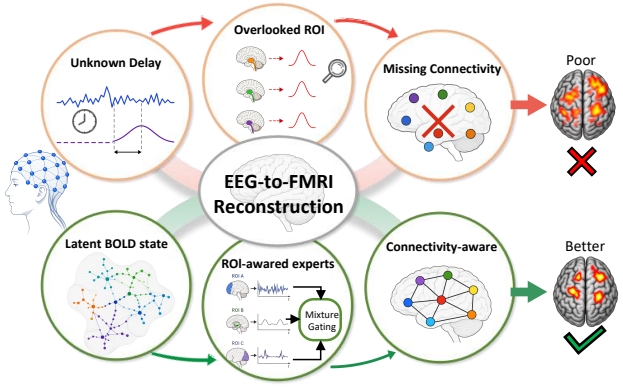


Figure 1: E2fMoE infers a shared latent BOLD state from EEG history and uses ROI-conditioned observation to reconstruct regional responses with connectivity-aware calibration of their joint dynamics.

To overcome these limitations, we investigate how neural activity is transformed into regional hemodynamic responses through latent brain dynamics. Drawing on insights from neurovascular coupling and dynamical systems theory Havlicek et al. (2015); Breakspear (2017), we formulate EEG-to-fMRI reconstruction as progressive state inference in which fast electrophysiological fluctuations are integrated into latent hemodynamic representations before being expressed across different brain regions. From a neurobiological perspective, BOLD signals do not directly mirror instantaneous neural activity, but instead emerge through temporally evolving physiological processes that integrate neural information and produce region-dependent responses.

Motivated by the above observations, we propose Latent Brain State-Guided Mixture-of-Experts (**E2fMoE**). A state-space sequence mixer transforms EEG tokens into slowly evolving latent representations, providing temporal memory for delayed BOLD responses Gu et al. (2022). Learnable region-of-interest (ROI) queries then extract target-specific information from the shared latent trajectory. A mixture of shared and specialized experts decodes these representations, capturing common cross-region transformations while adapting to regional heterogeneity. Finally, residual prediction and low-rank connectivity-aware calibration refine regional BOLD signals and their functional organization. The learned routing weights further reveal how different ROIs recruit shared and specialized decoding pathways. Our main contributions are as follows:

- We formulate EEG-to-fMRI reconstruction as latent hemodynamic state inference followed by ROI-conditioned observation, explicitly separating shared temporal dynamics from region-specific BOLD generation across heterogeneous ROIs.
- We instantiate this factorization in E2fMoE, enabling ROI-specific reconstruction from a common latent trajectory while constraining the joint predictions to retain functional connectivity across heterogeneous ROIs.
- We evaluate E2fMoE on four simultaneous EEG-fMRI datasets, achieving state-of-the-art performance in both ROI-wise and global reconstruction.

2 Related Work

Simultaneous EEG-fMRI modeling. Simultaneous EEG-fMRI combines fast electrophysiological activity with spatially resolved BOLD responses, but the two modalities differ in temporal resolution, spatial specificity, and physiological origin Ogawa et al. (1990); Logothetis et al. (2001); Georgopoulos et al. (2024); Mangini et al. (2025). Classical analyses often use hand-designed EEG

features or hemodynamic assumptions to explain fMRI responses Glover (1999); Debener et al. (2005); Buxton et al. (1998). Our work instead targets continuous EEG-to-fMRI reconstruction while preserving ROI-level FC structure Friston (1994); Biswal et al. (1995); Smith et al. (2011).

EEG representation learning. EEGNet and EEG Conformer demonstrate the effectiveness of convolutional and transformer-style EEG encoders Schirrmeister et al. (2017); Lawhern et al. (2018); Song et al. (2023); Roy et al. (2019); Craik et al. (2019), while recent self-supervised and foundation models improve cross-dataset EEG modeling through larger pretraining corpora and more general tokenization strategies Kostas et al. (2021); Banville et al. (2021); Yang et al. (2023a;b); Jiang et al. (2024); Wang et al. (2024; 2025); El Ouahidi et al. (2025); Ma et al. (2026). These models provide strong EEG feature extractors, but they are not designed to model delayed hemodynamic state transitions, ROI-specific decoding, or FC-aware reconstruction.

EEG-to-fMRI reconstruction models. NeuroBOLT studies resting-state EEG-to-fMRI synthesis, and UnEBOLT further unifies EEG-to-BOLD translation with FC reconstruction Li et al. (2024; 2026). E2fMoE separates the conditional EEG-to-BOLD transformation into shared latent hemodynamic dynamics and ROI-conditioned observation functions. E2fMoE uses state-space modeling Gu et al. (2022); Gupta et al. (2022); Smith et al. (2023); Gu & Dao (2024) and mixture-of-experts learning Jacobs et al. (1991); Shazeer et al. (2017); Lepikhin et al. (2021); Fedus et al. (2022); Riquelme et al. (2021); Du et al. (2022); Zoph et al. (2022) to instantiate these two factors for structured ROI-level reconstruction.

3 METHOD

In this section, we first introduce the **E2fMoE** architecture, as shown in Figure 2. We then detail two key components that enable factorized EEG-to-fMRI reconstruction through shared latent hemodynamic state inference and ROI-conditioned BOLD observation:

- **Latent BOLD State Space:** A compact hemodynamic state module that maps EEG tokens into slowly evolving latent BOLD states, capturing temporal dependencies that bridge high-frequency EEG activity and low-frequency fMRI responses.
- **ROI-Conditioned MoE:** An ROI-adaptive reconstruction module that uses ROI queries to route each brain region through shared and specialized experts, followed by residual prediction and low-rank FC calibration to preserve both temporal fidelity and cross-ROI connectivity.

Together, these components implement a factorized model of shared dynamics and region-specific observation while preserving temporal fidelity and cross-ROI connectivity.

3.1 THE E2FMOE ARCHITECTURE

E2fMoE reconstructs ROI-level BOLD signals through a factorized conditional mapping. Let $X \in \mathrm{R}^{C\times T}$ denote an EEG segment with C channels and T samples, and let $Y \in \mathrm{R}^{N\times L}$ denote the synchronized fMRI signal over N ROIs and L BOLD time points. The forward mapping is organized as

$$S = G_\theta(X), \quad \tilde{\mathbf{y}}_i = O_\theta(S, \mathbf{q}_i), \quad \hat{Y} = C_\theta(\tilde{Y}, A_{\mathrm{tr}}), \tag{1}$$

where G_θ infers the shared latent state sequence S, O_θ denotes the ROI-conditioned observation function, and C_θ couples the preliminary predictions through connectivity calibration. Here, $Q = \{\mathbf{q}_i\}_{i=1}^N$ is a set of learnable ROI queries and A_{tr} is estimated once from training subjects only Friston (1994); Biswal et al. (1995). As shown in Figure 2, the tokenizer first maps dense EEG samples into a shorter token sequence. We divide X into K non-overlapping windows $\{X_k\}_{k=1}^K$ and use a shared temporal encoder ϕ_θ, following common EEG representation learning practice Lawhern et al. (2018); Song et al. (2023), to extract local channel-time features $\mathbf{x}_k = W_x\phi_\theta(X_k) + \mathbf{b}_x$. The resulting sequence $X_e = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$ is propagated by the latent BOLD state-space module through long-range sequential memory. The predicted latent states are aligned with BOLD-derived states through the BOLD encoder and decoder branch. MLP pooling then summarizes the trajectory into hemodynamic tokens and a global state for downstream reconstruction. ROI queries attend to the hemodynamic tokens, and the resulting ROI features are decoded by shared and specialized

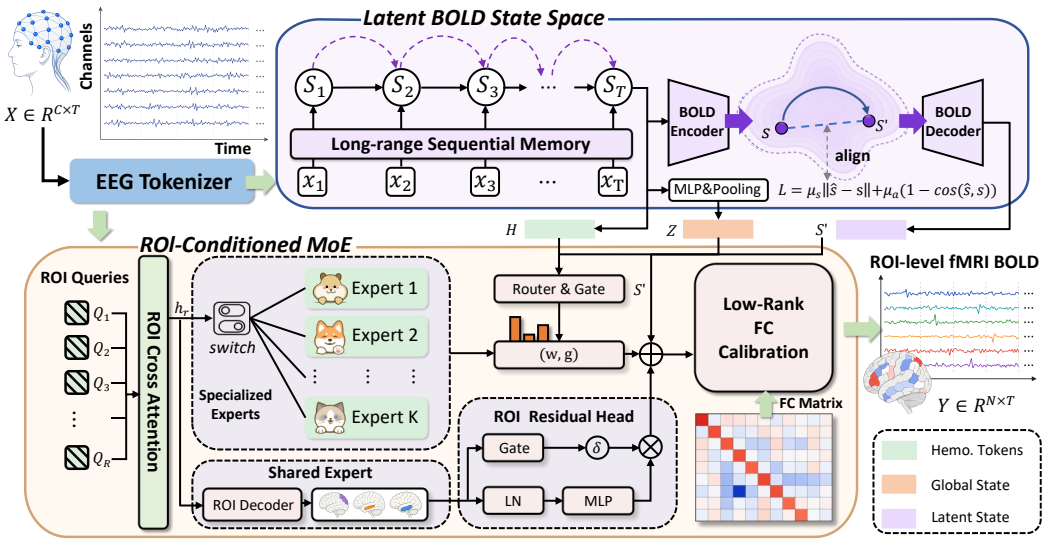


Figure 2: Overview of E2fMoE, which factorizes EEG-to-fMRI reconstruction into shared latent BOLD state inference and ROI-conditioned observation. Shared and specialized experts generate ROI signals, followed by residual prediction and FC-aware calibration.

experts. A residual BOLD head produces the preliminary prediction, which is further refined by low-rank FC calibration before final output.

3.2 Latent BOLD State Space

The first factor infers shared latent BOLD dynamics from EEG tokens while retaining the delayed influence of earlier neural activity. Motivated by hemodynamic response modeling and long-sequence state-space learning Glover (1999); Buxton et al. (1998); Gu et al. (2022); Gupta et al. (2022); Smith et al. (2023); Gu & Dao (2024), it maintains a latent state $\mathbf{s}_k$ for each EEG token and updates it recurrently as $\mathbf{s}_k = \bar{A}_k\mathbf{s}_{k-1} + \bar{B}_k\mathbf{x}_k$, where $\bar{A}_k$ and $\bar{B}_k$ are input-dependent transition parameters. The corresponding latent output is $\mathbf{u}_k = \bar{C}_k\mathbf{s}_k + \bar{D}_k\mathbf{x}_k$. We collect $S = [\mathbf{s}_1, \ldots, \mathbf{s}_K]$ as the latent BOLD state sequence and $U = [\mathbf{u}_1, \ldots, \mathbf{u}_K]$ as the hemodynamic token sequence. This state-space form gives the model an explicit memory variable, which is important because the BOLD response depends on previous neural activity rather than only the current EEG window.

During training, observed fMRI signals provide latent supervision. The BOLD encoder maps Y into a compact global state $\mathbf{z} = E_B(Y)$, and the BOLD decoder produces a BOLD-derived latent state sequence $S' = D_B(\mathbf{z})$. We regularize the predicted and decoded latent states by combining scale and angular consistency:

$$\mathcal{L}_{state} = \mu_s\|S - S'\|_2 + \mu_a\left(1 - \frac{\langle S, S'\rangle}{\|S\|_2\|S'\|_2}\right). \tag{2}$$

This alignment discourages the intermediate representation from becoming an unconstrained hidden feature and encourages it to preserve BOLD-like state structure.

3.3 ROI-Conditioned Mixture-of-Experts

The second factor implements an ROI-indexed observation model that converts shared hemodynamic tokens into region-specific BOLD features. We first obtain $H = \mathrm{MLP}(U)$ from the latent token sequence. For ROI i, the learnable query $\mathbf{q}_i$ attends to H using scaled dot-product attention Vaswani et al. (2017):

$$\alpha_i = \mathrm{softmax}\left(\frac{(W_q\mathbf{q}_i)(W_kH)^T}{\sqrt{d}}\right), \tag{3}$$

The resulting feature $\mathbf{h}_i = \alpha_iW_vH$ summarizes the latent state trajectory from the perspective of the i-th ROI.

We then decode $\mathbf{h}_i$ with a mixture of M specialized experts and one shared expert, following the mixture-of-experts principle of combining specialized local models with learned routing Jacobs et al. (1991); Shazeer et al. (2017); Lepikhin et al. (2021); Fedus et al. (2022); Riquelme et al. (2021); Du et al. (2022); Zoph et al. (2022). The router predicts normalized expert weights $\mathbf{g}_i = \mathrm{softmax}(W_g \mathbf{h}_i + \mathbf{b}_g)$. The expert output is

$$\mathbf{z}_i = \sum_{m=1}^{M} g_{i,m} E_m(\mathbf{h}_i) + E_{shared}(\mathbf{h}_i). \tag{4}$$

The shared expert captures observation patterns common to many ROIs, while the specialized experts model regional deviations. A shared decoder would impose one observation function on every ROI, whereas independent ROI heads would discard common structure. The shared-specialized form provides a structured compromise for regions that differ in response scale, delay, and temporal profile.

The ROI-conditioned decoder then converts each expert feature into a calibrated BOLD prediction. For ROI i, we compute a gated residual feature $\mathbf{r}_i = \sigma(W_r \mathbf{z}_i) \odot M_r(\mathbf{z}_i) + \mathbf{z}_i$ and predict the preliminary BOLD trace $\tilde{\mathbf{y}}_i = W_o \mathbf{r}_i + \mathbf{b}_o$. Stacking all ROI predictions gives $\tilde{Y} = [\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_N]^T$.

Point-wise reconstruction alone does not guarantee that the predicted ROI signals preserve functional connectivity Friston (1994); Biswal et al. (1995); Smith et al. (2011). We therefore add a low-rank calibration term. Given the training-only template A_{tr}, we parameterize a compact correction matrix as $\Delta A = P Q_c^T$, where $P, Q_c \in \mathrm{R}^{N \times r}$ and $r \ll N$. The calibrated prediction is

$$\hat{Y} = \tilde{Y} + \gamma (A_{\mathrm{tr}} + \Delta A) \tilde{Y}. \tag{5}$$

Validation/test fMRI is never used for template construction, calibration, model selection beyond validation loss, or inference.

We train the full model with reconstruction, FC, and latent-state losses. The reconstruction loss is

$$\mathcal{L}_{rec} = \frac{1}{NL} \sum_{i=1}^{N} \sum_{t=1}^{L} (\hat{Y}_{i,t} - Y_{i,t})^2. \tag{6}$$

The FC loss is used only in training mini-batches:

$$\mathcal{L}_{fc} = \frac{1}{N^2} \| \mathrm{Corr}(\hat{Y}) - \mathrm{Corr}(Y) \|_F^2. \tag{7}$$

The final objective is $\mathcal{L} = \mathcal{L}_{rec} + \lambda_{fc} \mathcal{L}_{fc} + \lambda_{state} \mathcal{L}_{state}$. This objective matches the evaluation protocol: $\mathcal{L}_{rec}$ optimizes temporal fidelity, $\mathcal{L}_{fc}$ optimizes connectivity preservation, and $\mathcal{L}_{state}$ constrains the intermediate hemodynamic representation during optimization.

4 Experiments

We conduct an extensive evaluation across ROI-wise and global reconstruction settings, together with interpretability, ablation, sensitivity, and representation analyses, to assess E2fMoE for EEG-to-fMRI reconstruction. The analyses cover **RQ1:** ROI-wise performance relative to representative baselines; **RQ2:** global reconstruction performance relative to representative baselines; **RQ3:** interpretability of ROI-adaptive expert routing; **RQ4:** the contribution of E2fMoE components; and **RQ5:** sensitivity to key hyperparameters.

4.1 Experimental Settings

Implementation details. All experiments were conducted on a single NVIDIA H200 GPU. To ensure direct comparability, all reproduced external baselines were evaluated under the same train-test protocol and ROI-level evaluation pipeline with configurations from their original papers or official code repositories. Additional implementation details, split protocols, and baseline adaptation protocols are provided in Appendices B and D.

Datasets. We evaluate our method on NeuroBOLT-Rest Li et al. (2024), Motor Imagery Neurofeedback Lioi et al. (2020), Confidence in Perceptual Decisions (CPD) Gherman & Philiastides (2020),

Table 1: Experimental comparison under the ROI-wise setting. Each cell reports mean ± standard deviation over five random seeds. **Best results** and second-best results are highlighted.

Metric	Model	NeuroBOLT						MI-NF		Overall
		Primary Sensory		High-level Cognitive		Subcortical		Motor System		
		Cuneus	Heschl	Mid. Front.	Prec. Ant.	Putamen	Thalamus	Precentral	Postcentral	
$R\uparrow$	EEGNet	0.517±.258	0.534±.262	0.392±.193	0.483±.230	0.363±.192	0.447±.196	0.335±.308	0.102±.006	0.456±.222
	EEG Conformer	0.435±.211	0.465±.202	0.423±.234	0.508±.163	0.344±.235	0.396±.212	0.376±.274	0.068±.021	0.429±.210
	BIOT	0.282±.185	0.348±.231	0.313±.190	0.330±.218	0.245±.213	0.302±.222	0.214±.150	0.078±.017	0.303±.210
	LaBraM	0.456±.181	0.387±.293	0.324±.297	0.317±.355	0.266±.280	0.308±.238	0.215±.229	0.020±.005	0.343±.274
	NeuroBOLT	0.568±.198	0.548±.228	0.487±.202	0.520±.172	0.424±.181	0.462±.205	**0.378±.240**	0.099±.009	0.502±.198
	CBraMod	0.391±.189	0.387±.196	0.379±.192	0.371±.179	0.289±.183	0.324±.211	0.243±.227	0.065±.041	0.357±.192
	REVE	0.466±.209	0.518±.202	0.387±.241	0.461±.206	0.335±.222	0.422±.211	0.336±.205	0.106±.031	0.432±.215
	CodeBrain	0.446±.249	0.398±.283	0.378±.289	0.326±.346	0.275±.258	0.340±.233	0.213±.146	0.039±.046	0.361±.276
	UnEBOLT	0.501±.172	0.397±.223	0.390±.221	0.426±.193	0.324±.185	0.396±.202	0.365±.264	0.043±.047	0.406±.199
	E2fMoE	**0.631±.175**	**0.611±.198**	**0.555±.163**	**0.629±.130**	**0.491±.176**	**0.534±.175**	0.369±.217	**0.106±.031**	**0.575±.170**
MSE$\downarrow$	EEGNet	0.276±.269	0.233±.151	0.259±.103	0.261±.122	0.224±.098	0.237±.108	**0.282±.112**	0.260±.002	0.248±.142
	EEG Conformer	0.305±.269	0.294±.143	0.285±.127	0.282±.127	0.265±.118	0.251±.108	0.309±.133	0.312±.078	0.280±.149
	BIOT	0.305±.267	0.266±.125	0.277±.114	0.299±.133	0.245±.101	0.252±.093	0.301±.110	0.272±.022	0.274±.139
	LaBraM	0.308±.280	0.280±.140	0.270±.101	0.326±.210	0.234±.091	0.277±.101	0.317±.125	0.259±.001	0.283±.154
	NeuroBOLT	0.261±.274	0.233±.130	**0.251±.103**	0.266±.135	0.232±.088	0.238±.103	0.297±.122	0.257±.001	0.247±.139
	CBraMod	0.307±.269	0.283±.126	0.270±.107	0.325±.177	0.245±.096	0.268±.090	0.304±.079	0.278±.018	0.283±.144
	REVE	0.307±.267	0.265±.135	0.287±.113	0.289±.119	0.244±.133	0.281±.153	0.288±.126	0.262±.008	0.279±.153
	CodeBrain	0.300±.268	0.265±.137	0.277±.113	0.296±.137	0.237±.116	0.256±.105	0.299±.123	0.303±.042	0.272±.146
	UnEBOLT	0.291±.276	0.301±.164	0.312±.148	0.320±.139	0.313±.168	0.301±.145	0.323±.133	0.281±.023	0.306±.173
	E2fMoE	**0.246±.278**	**0.226±.130**	0.253±.103	**0.254±.141**	**0.214±.080**	**0.236±.100**	0.289±.111	**0.257±.003**	**0.238±.139**
FC$\uparrow$	EEGNet	0.770±.076	0.465±.163	0.481±.099	0.529±.106	-0.193±.122	0.328±.148	-0.230±.809	0.859±.049	0.503±.136
	EEG Conformer	0.458±.086	0.267±.170	0.161±.084	0.210±.102	-0.411±.093	0.581±.086	0.349±.761	0.577±.230	0.313±.177
	BIOT	0.773±.069	0.198±.166	0.374±.089	0.519±.099	0.656±.098	**0.824±.066**	-0.013±.632	0.674±.015	0.605±.086
	LaBraM	0.772±.076	0.247±.171	0.397±.107	0.414±.101	0.672±.050	0.705±.086	-0.096±.731	0.627±.169	0.575±.121
	NeuroBOLT	0.665±.088	0.211±.172	0.323±.132	0.457±.102	0.539±.085	0.770±.074	0.266±.744	0.828±.154	0.576±.139
	CBraMod	**0.838±.071**	0.274±.154	0.302±.079	0.556±.107	0.583±.063	0.540±.091	0.262±.500	0.639±.200	0.535±.164
	REVE	0.744±.081	0.278±.175	0.310±.106	0.382±.106	0.643±.076	0.620±.093	0.135±.860	0.711±.249	0.581±.116
	CodeBrain	0.749±.070	0.268±.163	0.212±.078	0.317±.099	0.364±.086	0.427±.106	-0.023±.895	0.802±.164	0.508±.105
	UnEBOLT	0.588±.091	0.467±.144	0.291±.116	0.284±.155	-0.356±.081	0.228±.150	0.068±.709	0.766±.175	0.378±.147
	E2fMoE	0.793±.073	**0.718±.118**	**0.499±.094**	**0.806±.047**	**0.784±.046**	0.775±.072	**0.362±.802**	**0.860±.077**	**0.772±.076**

and SEF-Sleep, the simultaneous EEG and fMRI sleep dataset Gu et al. (2023). NeuroBOLT-Rest is used for ROI-wise EEG-to-fMRI reconstruction over seven targets, including Cuneus, Heschl, Mid. Front., Prec. Ant., Putamen, Thalamus, and the global signal. These regions cover primary sensory, high-level cognitive, and subcortical systems. The remaining datasets support global reconstruction tests across different experimental protocols. More dataset details are provided in Appendix B.

Evaluation metrics. We use the Pearson correlation coefficient (R) to measure temporal reconstruction fidelity between predicted and real fMRI signals, with higher values indicating better reconstruction. Mean squared error (MSE) measures point-wise prediction error, with lower values indicating better reconstruction. Functional connectivity correlation (FC) measures preservation of the real fMRI signals' connectivity structure, with higher values indicating better preservation. Appendix C defines both ROI-level and whole-matrix FC scores.

Baselines. We compare our method with nine representative baselines from classical and recent model families, including compact convolutional EEG models, transformer-based biosignal foundation models, and EEG-to-fMRI reconstruction models. Specifically, the baselines include EEGNet Lawhern et al. (2018), EEG Conformer Song et al. (2023), BIOT Yang et al. (2023a), LaBraM Jiang et al. (2024), NeuroBOLT Li et al. (2024), CBraMod Wang et al. (2025), REVE El Ouahidi et al. (2025), CodeBrain Ma et al. (2026), and UnEBOLT Li et al. (2026).

4.2 Performance Comparison on ROI-wise Reconstruction (RQ1)

To evaluate ROI-wise reconstruction, Table 1 compares E2fMoE with representative baselines under the intra-scan protocol, while Appendix E1 reports inter-subject generalization results. E2fMoE achieves the best overall performance on all three metrics, with an overall R of 0.575, MSE of 0.238, and FC of 0.772. On NeuroBOLT, our method obtains the best R on all six ROIs and the best MSE on five of them, showing consistent temporal reconstruction gains across heterogeneous brain targets. The FC results further show that E2fMoE preserves cross-ROI structure, reaching the best FC on Heschl, Mid. Front., Prec. Ant., Putamen, Postcentral, and Overall, while remaining second-best on Cuneus and Thalamus. On MI-NF motor ROIs, NeuroBOLT obtains the best R on Precentral, whereas our method obtains the best FC on Precentral and the best R, MSE, and FC on

Postcentral. These results indicate that latent BOLD state modeling and ROI-conditioned decoding improve region-specific reconstruction while maintaining connectivity fidelity. Paired NeuroBOLT significance analyses are provided in Appendix E2. Figure 3 provides qualitative examples across representative ROIs. The predicted BOLD trajectories broadly follow the ground-truth fluctuations, while the window-level correlation distributions show variation within each example.

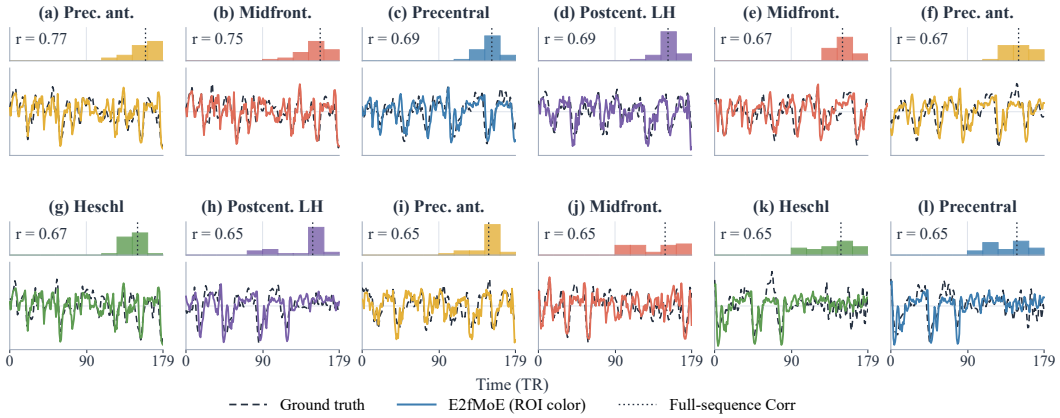


Figure 3: Qualitative EEG-conditioned BOLD prediction results on NeuroBOLT.

4.3 Performance Comparison on Global Reconstruction (RQ2)

To assess global reconstruction without ROI-level stratification, we compare E2fMoE with baselines under the global setting in Table 2. E2fMoE ranks first overall and achieves the best performance on seven of the nine metric-dataset combinations. On CPD, our method improves all three metrics, reaching the highest R of 0.303, lowest MSE of 0.311, and highest FC of 0.767. On NeuroBOLT Global, it also obtains the best R, MSE, and FC, showing that the model remains effective when evaluation is not decomposed by ROI category. On SEF-Sleep, E2fMoE achieves the best R of 0.199 and second-best FC of 0.580, while EEGNet obtains the lowest MSE and highest FC. The best average rank in Table 2 indicates that E2fMoE provides the most stable trade-off under global evaluation, where temporal accuracy and connectivity preservation must be balanced jointly.

Table 2: Experimental comparison under the global setting without ROI-level stratification. Each cell reports mean ± standard deviation over five random seeds. **Best results** and second-best results are highlighted throughout the table.

Model	CPD			NeuroBOLT Global			SEF-Sleep			Rank
	$R\uparrow$	MSE$\downarrow$	FC$\uparrow$	$R\uparrow$	MSE$\downarrow$	FC$\uparrow$	$R\uparrow$	MSE$\downarrow$	FC$\uparrow$	
EEGNet	0.232±.162	0.324±.100	0.612±.195	0.353±.049	0.231±.051	0.473±.144	0.141±.202	**0.950±.125**	**0.614±.248**	2
EEG Conformer	0.293±.062	0.330±.104	0.618±.180	0.218±.046	0.246±.048	0.424±.151	0.173±.173	0.966±.121	0.510±.281	3
BIOT	0.166±.013	0.328±.087	0.522±.167	0.326±.048	0.234±.013	0.609±.135	0.011±.039	1.010±.016	0.101±.166	10
LaBraM	0.140±.084	0.334±.094	0.363±.272	0.332±.052	0.230±.031	0.539±.143	0.052±.045	0.999±.011	0.225±.248	9
NeuroBOLT	0.204±.041	0.336±.095	0.515±.248	0.403±.054	0.217±.049	0.622±.127	0.122±.216	1.237±.328	0.565±.139	3
CBraMod	0.196±.011	0.327±.086	0.576±.221	0.341±.051	0.232±.034	0.502±.146	0.018±.028	1.035±.050	0.153±.124	6
REVE	0.177±.038	0.326±.094	0.442±.180	0.376±.054	0.226±.040	0.542±.147	0.040±.032	1.044±.093	0.257±.260	5
CodeBrain	0.150±.076	0.326±.094	0.293±.237	0.335±.052	0.231±.029	0.416±.159	0.055±.049	0.998±.012	0.279±.255	6
UnEBOLT	0.227±.055	0.334±.113	0.504±.216	0.096±.031	0.259±.038	0.125±.220	0.151±.227	1.001±.162	0.444±.198	8
E2fMoE	**0.303±.027**	**0.311±.107**	**0.767±.085**	**0.420±.038**	**0.206±.013**	**0.662±.106**	**0.199±.089**	1.008±.195	0.580±.134	**1**

4.4 ROI-adaptive Expert Routing Analysis (RQ3)

Figure 4 shows how E2fMoE routes each ROI across four experts. Panels (a)–(d) summarize expert usage, ROI preferences, anatomical mixtures, and routing confidence. Table A3 shows distinct dominant experts for Cuneus (E1), Heschl (E2), and Putamen (E4), whereas the global signal remains more evenly routed. This variation supports region-dependent expert selection while retaining soft mixtures without hard assignment. Appendix I shows the ROI-to-expert flow, and Appendix H provides the ROI embedding visualization.

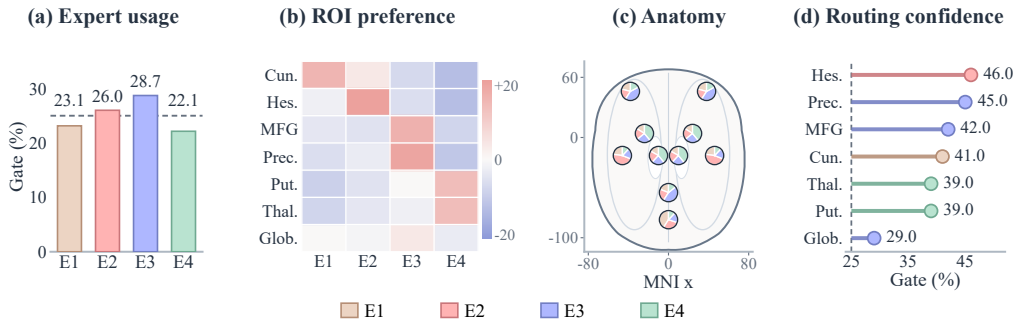


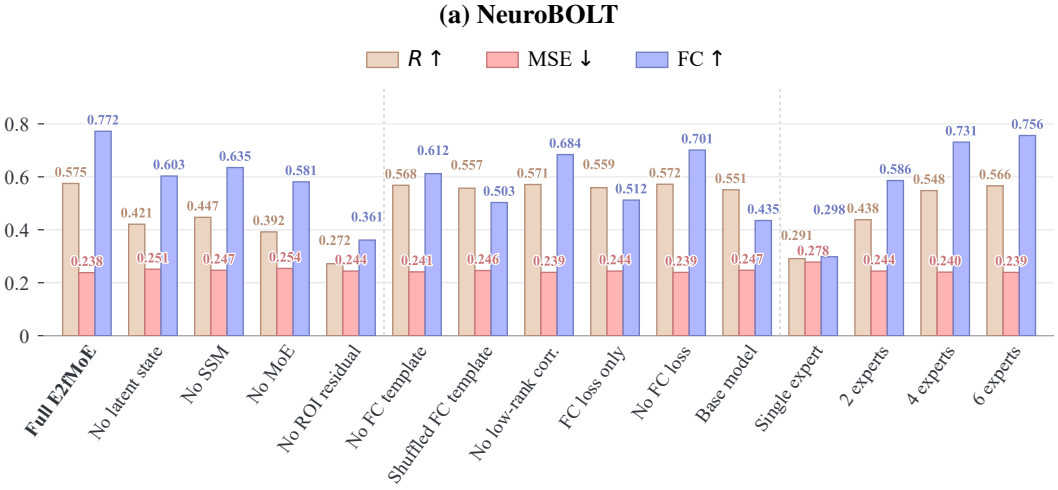
Figure 4: Expert routing on NeuroBOLT. Bars and lollipops show gate percentages; the heatmap shows percentage-point deviations from uniform routing, and anatomical markers show four-expert mixtures at schematic MNI coordinates. Dashed lines mark the 25% uniform gate.

4.5 Ablation Study (RQ4)

Table 3 compares architectural components, FC-source controls, expert-count variants, and latent encoders. Removing an architectural component degrades at least one key metric, and no alternative expert count matches the full model. Panels (c)–(d) isolate the effect of BOLD-state alignment. SSM performs best in both settings, while alignment improves R and MSE for every encoder.

Table 3: Ablations on NeuroBOLT and MI-NF. Panel (a) shows raw R, MSE, and FC values; arrows mark the favorable direction.

(a) NeuroBOLT



(b) MI-NF

Variant	$R\uparrow$	MSE↓	FC↑
Full E2fMoE	**0.238**	**0.273**	**0.611**
w/o latent state	0.171	0.287	0.542
w/o state-space mixer	0.186	0.282	0.557
w/o MoE	0.162	0.289	0.531
w/o ROI residual	0.204	0.284	0.524
w/o FC calibration head	0.228	0.276	0.496
Single expert	0.213	0.283	0.497
2 experts	0.205	0.279	0.548
4 experts	0.229	0.275	0.596
6 experts	0.235	0.274	0.604
FC head only	0.128	0.312	0.571

(c) Without BOLD-state alignment

Latent encoder	$R\uparrow$	MSE↓	FC↑
GRU	0.493	0.277	0.682
Transformer	0.536	0.245	0.701
SSM (Ours)	**0.548**	**0.242**	**0.718**

(d) With BOLD-state alignment

Latent encoder	$R\uparrow$	MSE↓	FC↑
GRU	0.501	0.245	0.635
Transformer	0.562	0.239	0.748
SSM (Ours)	**0.575**	**0.238**	**0.772**

In Table 3(a), removing or shuffling the FC template reduces FC from 0.772 to 0.612 or 0.503, respectively. Removing the low-rank correction or FC loss yields 0.684 or 0.701, while the FC-loss-only and base variants reach 0.512 and 0.435. The corresponding R and MSE values change less sharply for the FC-source controls, which isolates their strongest effect to connectivity preservation. On MI-NF, removing the FC calibration head reduces FC from 0.611 to 0.496, while its R changes from 0.238 to 0.228. Increasing the expert count from one to six raises FC from 0.497 to 0.604, still below the full model's 0.611. BOLD-state alignment improves R and MSE for all three encoders, while GRU's FC decreases from 0.682 to 0.635. These patterns support the role of FC calibration beyond temporal prediction alone. Figure 4 provides a complementary view: experts share the aggregate workload, while individual ROIs exhibit distinct gate preferences. This is consistent with region-adaptive allocation contributing alongside expert capacity. Exact bar values are in Table A9; ROI-level FC results for NeuroBOLT and MI-NF appear in Appendix E4.

4.6 Hyperparameter Sensitivity Analysis (RQ5)

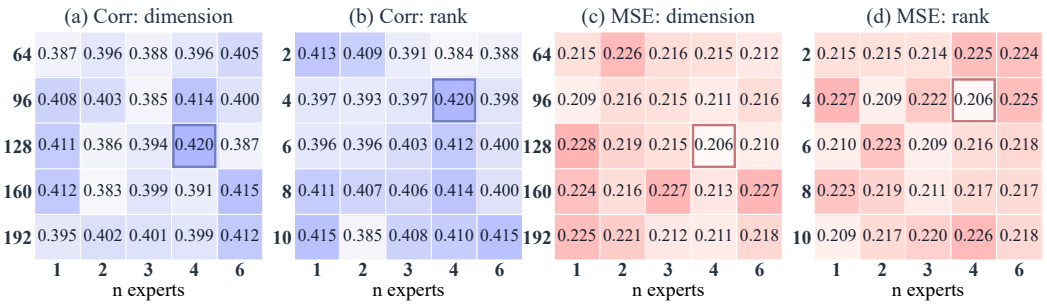


Figure 5: NeuroBOLT hyperparameter sensitivity: Corr and MSE across expert count, latent dimension, and routing rank.

We conduct an empirical sensitivity analysis on the NeuroBOLT global setting by varying the expert count, latent dimension, and routing rank in representative two-dimensional grids. In Figure 5, the highlighted cells correspond to the reference configuration used in the main experiments. This configuration achieves the best Corr and MSE, while alternative settings remain in a comparable range. The absence of a strict monotonic pattern suggests that performance is determined by the interaction among expert count, latent dimension, and routing rank rather than by increasing one hyperparameter alone. This supports the robustness of the proposed latent state, routing, and expert decoding design under key variations.

5 Conclusion

E2fMoE separates shared latent hemodynamic state inference from ROI-conditioned BOLD prediction and couples regional outputs through FC calibration. It improves temporal reconstruction and FC preservation across ROI-wise and global settings over EEG encoders and EEG-to-fMRI baselines. The FC-source ablations show that the template and low-rank correction contribute to connectivity preservation, while latent-state and expert-count variants identify the roles of temporal integration and regional specialization. Similar temporal scores can therefore conceal substantial differences in FC preservation, reinforcing the need to evaluate both regional trajectories and inter-ROI relationships when assessing EEG-to-fMRI reconstruction. Routing analysis further shows distinct ROI preferences within the shared expert pool. Together, these results support factorizing the EEG-to-fMRI mapping into shared dynamics and ROI-dependent responses.

Future work. We will test cross-dataset transfer and the stability of latent states and ROI routing under different acquisition protocols. We will extend reconstruction to longer recordings and finer parcellations, assessing temporal fidelity and functional connectivity at these scales.

AI USE STATEMENT

We used generative AI tools for language polishing and translation of the manuscript, assistance with modifying code, and LaTeX template migration and reference formatting.

REFERENCES

Hubert Banville, Omar Chehab, Aapo Hyvärinen, Denis-Alexander Engemann, and Alexandre Gramfort. Uncovering the structure of clinical EEG signals with self-supervised learning. *Journal of Neural Engineering*, 18(4):046020, 2021. doi: 10.1088/1741-2552/abca18.

Bharat Biswal, F. Zerrin Yetkin, Victor M. Haughton, and James S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34(4):537–541, 1995. doi: 10.1002/mrm.1910340409.

Michael Breakspear. Dynamic models of large-scale brain activity. *Nature Neuroscience*, 20(3): 340–352, 2017. doi: 10.1038/nn.4497.

Richard B. Buxton, Eric C. Wong, and Lawrence R. Frank. Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magnetic Resonance in Medicine*, 39(6): 855–864, 1998. doi: 10.1002/mrm.1910390602.

Alexander Craik, Yongtian He, and Jose L. Contreras-Vidal. Deep learning for electroencephalogram (EEG) classification tasks: A review. *Journal of Neural Engineering*, 16(3):031001, 2019. doi: 10.1088/1741-2552/ab0ab5.

Jessica S. Damoiseaux, Serge A. R. B. Rombouts, Frederik Barkhof, Philip Scheltens, Cornelis J. Stam, Stephen M. Smith, and Christian F. Beckmann. Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences*, 103(37):13848–13853, 2006. doi: 10.1073/pnas.0601417103.

Stefan Debener, Markus Ullsperger, Markus Siegel, Katja Fiehler, D. Yves von Cramon, and Andreas K. Engel. Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience*, 25(50):11730–11737, 2005. doi: 10.1523/JNEUROSCI.3286-05.2005.

Nan Du, Yanping Huang, Andrew M. Dai, Simon Tong, Dmitry Lepikhin, Yuanzhong Xu, Maxim Krikun, Yanqi Zhou, Adams Wei Yu, Orhan Firat, Barret Zoph, Liam Fedus, Maarten Bosma, Zongwei Zhou, Tao Wang, Yu Emma Wang, Kellie Webster, Marie Pellat, Kevin Robinson, Kathleen Meier-Hellstern, Toju Duke, Lucas Dixon, Kun Zhang, Quoc V. Le, Yonghui Wu, Zhifeng Chen, and Claire Cui. GLaM: Efficient scaling of language models with mixture-of-experts. In *International Conference on Machine Learning*, pp. 5547–5569, 2022.

Elias Ebrahimzadeh, Saber Saharkhiz, Lila Rajabion, Homayoun Baghaei Oskouei, Masoud Seraji, Farahnaz Fayaz, Sarah Saliminia, Seyyed Mostafa Sadjadi, and Hamid Soltanian-Zadeh. Simultaneous electroencephalography-functional magnetic resonance imaging for assessment of human brain function. *Frontiers in Systems Neuroscience*, 16:934266, 2022. doi: 10.3389/fnsys.2022.934266.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. REVE: A foundation model for EEG - adapting to any setup with large-scale pretraining on 25,000 subjects. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/20a917f77773ac0fa8bea2bdd6606b66-Abstract-Conference.html.

William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022.

Michael D. Fox and Marcus E. Raichle. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, 8(9):700–711, 2007. doi: 10.1038/nrn2201.

Karl J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1–2):56–78, 1994. doi: 10.1002/hbm.460020107.

Apostolos P. Georgopoulos, Peka Christova, Arthur C. Leuthold, and Lisa M. James. Association between neurovascular coupling and neural signals in the resting state as revealed by a combined fmri and magnetoencephalography study in humans. *Journal of Neurophysiology*, 132(4):1231–1234, 2024. doi: 10.1152/jn.00360.2024.

Sabina Gherman and Marios G. Philiastides. Simultaneous EEG-fMRI - confidence in perceptual decisions, 2020. URL https://openneuro.org/datasets/ds002739/versions/1.0.0.

Gary H. Glover. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, 9 (4):416–429, 1999. doi: 10.1006/nimg.1998.0419.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=tEYskw1VY2.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=uYLFoz1vlAC.

Yameng Gu, Lucas E. Sainburg, Feng Han, and Xiao Liu. Simultaneous EEG and functional MRI data during rest and sleep from humans. *Data in Brief*, 48:109059, 2023. doi: 10.1016/j.dib.2023.109059. URL https://doi.org/10.1016/j.dib.2023.109059.

Ankit Gupta, Albert Gu, and Jonathan Berant. Diagonal state spaces are as effective as structured state spaces. In *Advances in Neural Information Processing Systems*, volume 35, pp. 22982–22994, 2022.

Daniel A. Handwerker, John M. Ollinger, and Mark D'Esposito. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, 21(4):1639–1651, 2004. doi: 10.1016/j.neuroimage.2003.11.029.

Martin Havlicek, Alard Roebroeck, Karl Friston, Anna Gardumi, Dimo Ivanov, and Kamil Uludag. Physiologically informed dynamic causal modeling of fMRI data. *NeuroImage*, 122:355–372, 2015. doi: 10.1016/j.neuroimage.2015.07.078.

Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3(1):79–87, 1991. doi: 10.1162/neco.1991.3.1.79.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous EEG data in BCI. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=QzTpTRVtrP.

Demetres Kostas, Stephane Aroca-Ouellette, and Frank Rudzicz. BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. *Frontiers in Human Neuroscience*, 15:653659, 2021. doi: 10.3389/fnhum.2021.653659.

Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, and Brent J. Lance. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018. doi: 10.1088/1741-2552/aace8c.

Dmitry Lepikhin, HyoukJoong Lee, Yuanzhong Xu, Dehao Chen, Orhan Firat, Yanping Huang, Maxim Krikun, Noam Shazeer, and Zhifeng Chen. GShard: Scaling giant models with conditional computation and automatic sharding. In *International Conference on Learning Representations*, 2021.

Yamin Li, Ange Lou, Ziyuan Xu, Shengchao Zhang, Shiyu Wang, Dario J. Englot, Soheil Kolouri, Daniel Moyer, Roza G. Bayrak, and Catie Chang. NeuroBOLT: Resting-state EEG-to-fMRI synthesis with multi-dimensional feature mapping. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-0736. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/29c294ddc628c94cd2c636383ef106c1-Abstract-Conference.html.

Yamin Li, Ange Lou, Chang Li, Shiyu Wang, Haatef Pourmotabbed, Ziyuan Xu, Shengchao Zhang, Dario J. Englot, Soheil Kolouri, Daniel Moyer, Roza G. Bayrak, and Catie Chang. UnEBOLT: A unified model for EEG-to-BOLD translation and functional connectivity reconstruction. In Yuankai Huo, Mingchen Gao, Chang-Fu Kuo, Yueming Jin, and Ruining Deng (eds.), *Proceedings of The 9th International Conference on Medical Imaging with Deep Learning*, volume 315 of *Proceedings of Machine Learning Research*, pp. 2338–2351. PMLR, 08–10 Jul 2026. URL https://proceedings.mlr.press/v315/li26g.html.

Martin A. Lindquist, Ji Meng Loh, Lauren Y. Atlas, and Tor D. Wager. Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. *NeuroImage*, 45(1 Suppl):S187–S198, 2009. doi: 10.1016/j.neuroimage.2008.10.065.

Giulia Lioi, Claire Cury, Lorraine Perronnet, Marsel Mano, Elise Bannier, Anatole Lécuyer, and Christian Barillot. Simultaneous EEG-fMRI during a neurofeedback task, a brain imaging dataset for multimodal data integration. *Scientific Data*, 7:173, 2020. doi: 10.1038/s41597-020-0498-3. URL https://doi.org/10.1038/s41597-020-0498-3.

Giulia Lioi, Claire Cury, Lorraine Perronnet, Marsel Mano, Elise Bannier, Anatole Lécuyer, and Christian Barillot. A multi-modal human neuroimaging dataset for data integration: simultaneous EEG and fMRI acquisition during a motor imagery neurofeedback task: XP1, 2021. URL https://openneuro.org/datasets/ds002336/versions/2.0.2.

Giulia Lioi, Claire Cury, Lorraine Perronnet, Marsel Mano, Elise Bannier, Anatole Lécuyer, and Christian Barillot. A multi-modal human neuroimaging dataset for data integration: simultaneous EEG and fMRI acquisition during a motor imagery neurofeedback task: XP2, 2022. URL https://openneuro.org/datasets/ds002338/versions/2.0.2.

Nikos K. Logothetis, Jon Pauls, Mark Augath, Torsten Trinath, and Axel Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843):150–157, 2001. doi: 10.1038/35084005.

Jingying Ma, Feng Wu, Qika Lin, Yucheng Xing, Chenyu Liu, Ziyu Jia, and Mengling Feng. CodeBrain: Bridging decoupled tokenizer and multi-scale architecture for EEG foundation model. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=msJgEkjwh5.

Fabio Mangini, Marta Moraschi, Daniele Mascali, Maria Guidi, Michela Fratini, Silvia Mangia, Mauro DiNuzzo, Fabrizio Frezza, and Federico Giove. Towards whole brain mapping of the hemodynamic response function. *Journal of Cerebral Blood Flow and Metabolism*, 2025. doi: 10.1177/0271678X251325413.

Seiji Ogawa, Tso-Ming Lee, Alan R. Kay, and David W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990. doi: 10.1073/pnas.87.24.9868.

Jonathan D. Power, Alexander L. Cohen, Steven M. Nelson, Gagan S. Wig, Kelly Anne Barnes, Jessica A. Church, Alecia C. Vogel, Timothy O. Laumann, Fran M. Miezin, Bradley L. Schlaggar, and Steven E. Petersen. Functional network organization of the human brain. *Neuron*, 72(4):665–678, 2011. doi: 10.1016/j.neuron.2011.09.006.

Carlos Riquelme, Joan Puigcerver, Basil Mustafa, Maxim Neumann, Rodolphe Jenatton, André Susano Pinto, Daniel Keysers, and Neil Houlsby. Scaling vision with sparse mixture of experts. In *Advances in Neural Information Processing Systems*, volume 34, pp. 8583–8595, 2021.

Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H. Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: A systematic review. *Journal of Neural Engineering*, 16(5):051001, 2019. doi: 10.1088/1741-2552/ab260c.

Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, 38(11):5391–5420, 2017. doi: 10.1002/hbm.23730.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=B1ckMDqlg.

Jimmy T. H. Smith, Andrew Warrington, and Scott W. Linderman. Simplified state space layers for sequence modeling. In *International Conference on Learning Representations*, 2023.

Stephen M. Smith, Karla L. Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F. Beckmann, Thomas E. Nichols, Joseph D. Ramsey, and Mark W. Woolrich. Network modelling methods for FMRI. *NeuroImage*, 54(2):875–891, 2011. doi: 10.1016/j.neuroimage.2010.08.063.

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. EEG Conformer: Convolutional transformer for EEG decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2023. doi: 10.1109/TNSRE.2022.3230250.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc 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. doi: 10.1006/nimg.2001.0978.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5998–6008, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Guangyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, and Haifeng Li. EEGPT: Pretrained transformer for universal and reliable representation of EEG signals. In *Advances in Neural Information Processing Systems*, volume 37, pp. 39249–39280, 2024. doi: 10.52202/079017-1239. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/4540d267eeec4e5dbd9dae9448f0b739-Abstract-Conference.html.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. CBraMod: A criss-cross brain foundation model for EEG decoding. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=NPNUHgHF2w.

Tracy Warbrick. Simultaneous EEG-fMRI: What have we learned and what does the future hold? *Sensors*, 22(6):2262, 2022. doi: 10.3390/s22062262.

Chaoqi Yang, M. Brandon Westover, and Jimeng Sun. BIOT: Biosignal transformer for cross-data learning in the wild. In *Advances in Neural Information Processing Systems*, volume 36, pp. 78240–78260, 2023a. URL https://openreview.net/forum?id=c2LZyTyddi.

Chaoqi Yang, Danica Xiao, M. Brandon Westover, and Jimeng Sun. Self-supervised electroencephalogram representation learning for automatic sleep staging: Model development and evaluation study. *JMIR AI*, 2:e46769, 2023b. doi: 10.2196/46769.

Barret Zoph, Irwan Bello, Sameer Kumar, Nan Du, Yanping Huang, Jeff Dean, Noam Shazeer, and William Fedus. ST-MoE: Designing stable and transferable sparse expert models. *arXiv preprint arXiv:2202.08906*, 2022.

TECHNICAL APPENDIX

Appendix A. Theoretical Analysis

This section provides theoretical analysis for the main components of E2fMoE. Algorithm A1 first summarizes the overall training and inference pipeline, showing how EEG tokens are converted into latent BOLD states, decoded through ROI-conditioned experts, and refined by FC-aware calibration. The following propositions then formalize four design properties: latent state memory, ROI-adaptive expressiveness, parameter-efficient connectivity calibration, and consistency between the training objective and the reported metrics.

Proposition 1. The latent BOLD state forms a history-dependent hemodynamic filter. For the state update $\mathbf{s}_k = \bar{A}_k \mathbf{s}_{k-1} + \bar{B}_k \mathbf{x}_k$, define $\Phi_{k,\tau} = \prod_{j=\tau+1}^{k} \bar{A}_j$ for $\tau < k$ and $\Phi_{k,k} = I$. The latent state at step k admits the expansion

$$\mathbf{s}_k = \left(\prod_{j=1}^{k} \bar{A}_j \right) \mathbf{s}_0 + \sum_{\tau=1}^{k} \Phi_{k,\tau} \bar{B}_\tau \mathbf{x}_\tau, \tag{8}$$

where an empty product is the identity matrix. Thus $\mathbf{s}_k$ depends on all EEG tokens $\{\mathbf{x}_\tau\}_{\tau=1}^{k}$ through transition-dependent temporal filters.

Proof. For $k = 1$, the statement gives $\mathbf{s}_1 = \bar{A}_1 \mathbf{s}_0 + \bar{B}_1 \mathbf{x}_1$, which is exactly the recurrent update. Assume the expression holds for $k-1$. Substituting it into the update at step k gives

$$\begin{aligned} \mathbf{s}_k &= \bar{A}_k \mathbf{s}_{k-1} + \bar{B}_k \mathbf{x}_k \\ &= \bar{A}_k \left[\left(\prod_{j=1}^{k-1} \bar{A}_j \right) \mathbf{s}_0 + \sum_{\tau=1}^{k-1} \Phi_{k-1,\tau} \bar{B}_\tau \mathbf{x}_\tau \right] + \bar{B}_k \mathbf{x}_k \\ &= \left(\prod_{j=1}^{k} \bar{A}_j \right) \mathbf{s}_0 + \sum_{\tau=1}^{k} \Phi_{k,\tau} \bar{B}_\tau \mathbf{x}_\tau. \end{aligned} \tag{9}$$

This proves the expansion. The coefficients multiplying each historical token depend on the product of later transition matrices, so the module can represent delayed and temporally smoothed EEG-to-BOLD effects rather than only pointwise EEG features.

Proposition 2. ROI-conditioned MoE strictly contains shared and single-expert decoders. Let the E2fMoE decoder for ROI i be

$$\mathbf{z}_i = \sum_{m=1}^{M} g_{i,m} E_m(\mathbf{h}_i) + E_{shared}(\mathbf{h}_i), \tag{10}$$

where $\mathbf{g}_i = \mathrm{softmax}(W_g \mathbf{h}_i + \mathbf{b}_g)$. This hypothesis class contains a shared decoder and a single-expert decoder as special cases, and therefore is at least as expressive as these restricted alternatives.

Proof. Consider a shared decoder $D(\mathbf{h}_i)$. Set $E_{shared} = D$ and set each specialized expert to the zero map, $E_m(\cdot) = 0$. Then $\mathbf{z}_i = D(\mathbf{h}_i)$ independently of the routing weights. Hence every shared decoder is realizable by E2fMoE. For a single-expert decoder D_1, set $E_1 = D_1$, set $E_m = 0$ for $m > 1$, set $E_{shared} = 0$, and choose router logits so that $g_{i,1} \to 1$ for all ROIs. Since softmax can approach a one-hot distribution by increasing the logit margin, the E2fMoE output approaches $D_1(\mathbf{h}_i)$ arbitrarily closely as $\lim_{\eta \to \infty} \mathrm{softmax}([\eta, 0, \ldots, 0])_1 = 1$. Thus the proposed decoder includes the restricted shared and single-expert designs while additionally allowing ROI-dependent mixtures over multiple specialized experts. This formalizes the expressiveness advantage tested in the ablation study.

Proposition 3. Low-rank FC calibration is a constrained graph correction. Let the calibrated prediction be $\hat{Y} = \tilde{Y} + \gamma(A_{\mathrm{tr}} + PQ_c^T)\tilde{Y}$, where $P, Q_c \in \mathbb{R}^{N \times r}$ and $r \ll N$. The correction $\Delta A = PQ_c^T$ has rank at most r, uses $2Nr$ parameters, and modifies the prediction through ROI graph mixing.

Algorithm A1: E2fMoE Overall Architecture

Require: Training dataset $\mathcal{D} = \{(X, Y)\}$, ROI queries Q, frozen FC template A_{tr}, number of experts M, total iterations I

Initialize tokenizer ϕ_θ, latent mixer F_θ, BOLD encoder-decoder (E_B, D_B), ROI router R_θ, experts $\{E_s, E_m\}_{m=1}^{M}$, residual head H_θ, and FC calibration parameters P, Q_c
for $iter = 1$ to I **do**
  Sample a mini-batch $\{(X_b, Y_b)\} \sim \mathcal{D}$
  Tokenize EEG and predict latent BOLD states: $S_b = F_\theta(\phi_\theta(X_b))$
  Obtain BOLD-derived latent states $S'_b = D_B(E_B(Y_b))$
  Decode each ROI with query attention and routed experts
  Stack ROI outputs into $\tilde{Y}_b$ and calibrate $\hat{Y}_b = \tilde{Y}_b + \gamma(A_{\text{tr}} + PQ_c^T)\tilde{Y}_b$
  Compute $\mathcal{L} = \mathcal{L}_{rec} + \lambda_{fc}\mathcal{L}_{fc} + \lambda_{state}\mathcal{L}_{state}$
  Update model parameters with $\nabla_\theta \mathcal{L}$
end for
Inference: Given test EEG X, apply decoding and FC calibration with fixed A_{tr}
return Reconstructed ROI-level BOLD signals $\hat{Y}$

Proof. The calibration can be rewritten as $\hat{Y} = (I + \gamma A_{\text{tr}} + \gamma \Delta A)\tilde{Y}$. Thus each calibrated ROI signal is a linear combination of preliminary ROI predictions, with mixing weights determined by the frozen training-subject template A_{tr} and the learned correction ΔA. For the correction term, the product-rank inequality gives

$$\text{rank}(\Delta A) \leq \min(\text{rank}(P), \text{rank}(Q_c)) \leq r. \tag{11}$$

The parameter count is $Nr + Nr = 2Nr$, compared with N^2 parameters for a dense unconstrained correction. Therefore the calibration is a constrained graph correction that can adjust coordinated ROI-level errors while limiting the number of free connectivity parameters.

Proposition 4. The objective is optimization-consistent with temporal and FC evaluation. Let the training loss be $\mathcal{L} = \mathcal{L}_{rec} + \lambda_{fc}\mathcal{L}_{fc} + \lambda_{state}\mathcal{L}_{state}$, where $\mathcal{L}_{rec}$ is the pointwise MSE and $\mathcal{L}_{fc} = \frac{1}{N^2}\|\text{Corr}(\hat{Y}) - \text{Corr}(Y)\|_F^2$. If $\lambda_{fc} > 0$, any update that decreases $\mathcal{L}_{rec}$ or $\mathcal{L}_{fc}$ while not increasing the other terms improves the total objective. Thus the objective explicitly couples temporal reconstruction and FC preservation, matching the two families of metrics reported in the experiments.

Proof. The total differential of the objective with respect to model parameters θ is $d\mathcal{L} = d\mathcal{L}_{rec} + \lambda_{fc} d\mathcal{L}_{fc} + \lambda_{state} d\mathcal{L}_{state}$. Since $\lambda_{fc} > 0$, reductions in FC mismatch contribute directly to a reduction of the total objective whenever the remaining terms do not offset the improvement. Similarly, $\mathcal{L}_{rec}$ is exactly the MSE-style reconstruction error optimized during training. Therefore the optimization target is aligned with both temporal fidelity and connectivity preservation. The latent-state term does not replace these metrics; it regularizes the intermediate hemodynamic representation that feeds the ROI decoder.

Appendix B. Dataset Descriptions

Table A1 summarizes the datasets used in the experiments, including their public scale and evaluation role. The global results in Table 2 of the main paper are cross-subject experiments, where training, validation, and test splits are subject-disjoint.

All experiments were conducted on a Linux workstation equipped with a single NVIDIA H200 GPU. The software environment was based on Python and PyTorch with CUDA acceleration. Standard scientific computing and neuroimaging packages, including NumPy, SciPy, MNE, and scikit-learn, were used for data processing, model training, and evaluation.

Split protocol and FC template construction. All validation and test samples are kept outside model fitting. For cross-subject experiments, training, validation, and test partitions are subject-disjoint. The FC template A_{tr} is estimated once from training subjects only and is frozen before validation and test evaluation. Validation/test fMRI is used only for metric computation, and is never used for template construction, calibration, model selection beyond validation loss, or inference.

Table A1: Summary of datasets used for EEG-to-fMRI reconstruction experiments.

Dataset	Source	Scale	Split and usage
NB-Rest	Resting EEG-fMRI Li et al. (2024)	29 scans, 22 subjects	ROI-wise and cross-subject global reconstruction
MI-NF	OpenNeuro XP1/XP2 Lioi et al. (2020; 2021; 2022)	10 XP1 and 23 XP2 participants	Motor-system ROI evaluation
CPD	OpenNeuro ds002739 Gherman & Philiastides (2020)	24 BIDS subjects	Cross-subject global reconstruction
SEF-Sleep	Sleep EEG-fMRI Gu et al. (2023)	33 healthy participants	Cross-subject global reconstruction under sleep shifts

NeuroBOLT-Rest. NeuroBOLT-Rest follows the NeuroBOLT EEG-to-fMRI synthesis benchmark, which studies resting-state translation from EEG to BOLD signals. In this paper, it provides the main ROI-wise setting, covering primary sensory, high-level cognitive, and subcortical ROIs, and also contributes the NeuroBOLT Global setting for non-stratified reconstruction. This dataset is central for evaluating whether E2fMoE can jointly improve temporal BOLD reconstruction and FC preservation in resting-state signals.

Motor Imagery Neurofeedback. The Motor Imagery Neurofeedback data come from a simultaneous EEG-fMRI neurofeedback resource released with XP1 and XP2 OpenNeuro datasets. The recordings are designed around motor imagery neurofeedback, making them suitable for testing motor-system targets such as Precentral and Postcentral ROIs. We use this dataset to evaluate whether ROI-conditioned decoding remains effective when the target regions are dominated by motor imagery and feedback-related activity.

Confidence in Perceptual Decisions. The CPD dataset is the OpenNeuro ds002739 simultaneous EEG-fMRI dataset for confidence in perceptual decisions. Unlike resting-state or motor-imagery data, CPD reflects a decision-making protocol in which neural responses are tied to perceptual evidence and confidence. We use CPD in the global reconstruction setting to test whether E2fMoE generalizes beyond resting-state reconstruction to task-evoked cognitive dynamics.

SEF-Sleep. SEF-Sleep denotes the simultaneous EEG and fMRI dataset collected during rest and sleep. Sleep recordings introduce substantial changes in physiological state, vigilance, and signal distribution compared with awake resting or task datasets. We therefore use SEF-Sleep as a global reconstruction benchmark for assessing robustness under state-dependent EEG-fMRI shifts.

Appendix C. Evaluation Metric Descriptions

This section defines the three evaluation metrics used in the main experiments. Let $Y^{(s)} \in \mathbb{R}^{N\times L}$ and $\hat{Y}^{(s)} \in \mathbb{R}^{N\times L}$ denote the ground-truth and reconstructed BOLD signals for sample s, where N is the number of ROIs and L is the number of fMRI time points. We report temporal correlation, point-wise error, and functional-connectivity preservation because EEG-to-fMRI reconstruction should recover both regional BOLD traces and their cross-ROI organization.

Pearson correlation coefficient. The temporal reconstruction score R measures the Pearson correlation between the predicted and ground-truth time series. For ROI i in sample s, it is computed as

$$\rho(\hat{\mathbf{y}}_i^{(s)}, \mathbf{y}_i^{(s)}) = \frac{\sum_{\ell=1}^{L}(\hat{Y}_{i,\ell}^{(s)} - \bar{\hat{Y}}_i^{(s)})(Y_{i,\ell}^{(s)} - \bar{Y}_i^{(s)})}{\sigma(\hat{\mathbf{y}}_i^{(s)})\sigma(\mathbf{y}_i^{(s)})}, \tag{12}$$

where $\sigma(\mathbf{y}_i^{(s)}) = \sqrt{\sum_{\ell=1}^{L}(Y_{i,\ell}^{(s)} - \bar{Y}_i^{(s)})^2}$. The reported score is $R_i^{(s)} = \rho(\hat{\mathbf{y}}_i^{(s)}, \mathbf{y}_i^{(s)})$, where $\bar{\hat{Y}}_i^{(s)}$ and $\bar{Y}_i^{(s)}$ are temporal means. The reported R is averaged over evaluated samples and ROIs, and a higher value indicates better temporal reconstruction.

Mean squared error. Mean squared error measures point-wise reconstruction deviation between the predicted and ground-truth BOLD signals:

$$\mathrm{MSE} = \frac{1}{SNL} \sum_{s=1}^{S} \sum_{i=1}^{N} \sum_{\ell=1}^{L} \left(\hat{Y}_{i,\ell}^{(s)} - Y_{i,\ell}^{(s)} \right)^2, \tag{13}$$

where S is the number of evaluated samples. A lower MSE indicates smaller point-wise prediction error.

Functional connectivity correlation. Functional connectivity preservation is evaluated by comparing the ROI-wise correlation matrices of the reconstructed and ground-truth BOLD signals. For sample s, define $C^{(s)} = \mathrm{Corr}(Y^{(s)})$ and $\hat{C}^{(s)} = \mathrm{Corr}(\hat{Y}^{(s)})$, where each matrix entry is the Pearson correlation between two ROI time series. We then vectorize the upper-triangular entries excluding the diagonal and compute

$$\mathrm{FC}^{(s)} = \mathrm{Pearson}\left(\mathrm{vec}_u(\hat{C}^{(s)}), \mathrm{vec}_u(C^{(s)}) \right). \tag{14}$$

This whole-matrix score is reported in the Overall FC column. For an individual ROI i, we instead compare its connections to the other evaluated ROIs. Let $c_i^{(s)} = (C_{ij}^{(s)})_{j \neq i}$ and $\hat{c}_i^{(s)} = (\hat{C}_{ij}^{(s)})_{j \neq i}$, with the diagonal entry excluded. The ROI-level score is

$$\mathrm{FC}_i^{(s)} = \mathrm{Pearson}\left(\hat{c}_i^{(s)}, c_i^{(s)} \right). \tag{15}$$

Thus, the named ROI columns in Tables 1 and A2 measure agreement of incident connectivity edges, whereas Overall FC uses all unique ROI pairs as in Eq. (14). Scores are averaged over evaluation samples, and table entries summarize the resulting scores across runs. Higher values indicate better preservation of the corresponding connectivity pattern.

Appendix D. Baseline Methods

This section describes the baseline methods used in the main experiments. The comparison covers three representative families: compact supervised EEG encoders, recent EEG foundation or representation models, and task-specific EEG-to-fMRI reconstruction models. For a fair comparison, all baselines are adapted to predict the same target BOLD representation and are evaluated using the metrics defined in Appendix C.

Baseline adaptation protocol. All baselines use the same EEG input windows, target BOLD preprocessing, train/validation/test splits, and evaluation metrics as E2fMoE. EEG representation models are used as encoders and connected to a BOLD prediction head for the same reconstruction target. Task-specific EEG-to-fMRI baselines are reproduced with their official configurations when available and are evaluated under the same ROI-wise or global protocol. No baseline or proposed model accesses validation/test fMRI during inference.

EEGNet. EEGNet is a compact convolutional EEG decoder that captures temporal and channel-wise patterns with few parameters Lawhern et al. (2018). We use it as a classical supervised EEG baseline before BOLD regression.

EEG Conformer. EEG Conformer combines convolutional EEG feature extraction with self-attention Song et al. (2023). It tests whether a general attention-based EEG encoder can support BOLD reconstruction without explicit latent BOLD state modeling.

BIOT. BIOT is a biosignal transformer for transferable biomedical time-series representation learning Yang et al. (2023a). It serves as a cross-biosignal baseline for testing whether generic biosignal representations transfer to EEG-to-fMRI reconstruction.

LaBraM. LaBraM is a large EEG representation model trained to learn general-purpose brain signal features Jiang et al. (2024). We include it as a strong EEG foundation baseline without ROI-adaptive BOLD decoding.

CBraMod. CBraMod learns general EEG representations across subjects and recording conditions Wang et al. (2025). We compare with it to evaluate whether cross-brain EEG modeling alone preserves ROI-wise BOLD dynamics and FC structure.

Table A2: Inter-subject ROI-wise reconstruction comparison. ROI cells report mean ± standard deviation over five random seeds. Overall R and MSE average the six displayed ROI means and standard deviations; overall FC is reported separately. **Best results** and second-best results are highlighted.

Metric	Model	NeuroBOLT						Overall
		Primary Sensory		High-level Cognitive		Subcortical		
		Cuneus	Heschl	Mid. Front.	Prec. Ant.	Putamen	Thalamus	
$R\uparrow$	EEGNet	0.230±.038	0.352±.032	0.289±.045	0.283±.040	0.239±.037	0.251±.039	0.274±.039
	EEG Conformer	0.151±.041	0.174±.043	0.149±.042	0.192±.042	0.132±.038	0.020±.039	0.136±.041
	BIOT	0.245±.033	0.332±.032	0.284±.047	0.276±.045	0.248±.035	0.233±.034	0.270±.038
	LaBraM	0.228±.035	0.327±.041	0.296±.050	0.320±.047	0.262±.045	0.258±.046	0.282±.044
	NeuroBOLT	0.274±.042	0.409±.036	0.317±.035	0.359±.053	0.286±.029	0.295±.036	0.323±.039
	CBraMod	0.234±.036	0.333±.034	0.286±.043	0.294±.045	0.240±.031	0.240±.037	0.271±.038
	REVE	0.266±.037	0.382±.035	0.320±.052	0.330±.051	0.276±.040	0.263±.050	0.306±.044
	CodeBrain	0.253±.040	0.328±.041	0.291±.053	0.286±.050	0.252±.038	0.245±.041	0.276±.044
	UnEBOLT	0.059±.034	0.048±.039	0.076±.038	0.048±.031	0.038±.039	0.009±.035	0.046±.036
	E2fMoE	**0.300**±.036	**0.424**±.029	**0.341**±.057	**0.377**±.033	**0.310**±.037	**0.320**±.046	**0.345**±.040
MSE$\downarrow$	EEGNet	0.245±.014	0.229±.024	0.236±.021	0.247±.045	0.230±.036	0.251±.030	0.240±.028
	EEG Conformer	0.257±.018	0.253±.038	0.249±.051	0.258±.040	0.240±.028	0.267±.041	0.254±.036
	BIOT	0.250±.030	0.228±.048	0.234±.041	0.247±.031	0.229±.012	0.252±.049	0.240±.035
	LaBraM	0.248±.024	0.228±.039	0.233±.017	0.240±.046	0.228±.048	0.248±.047	0.238±.037
	NeuroBOLT	0.240±.039	0.214±.017	0.233±.011	0.232±.055	0.224±.011	0.248±.012	0.232±.024
	CBraMod	0.245±.017	0.229±.025	0.234±.042	0.245±.012	0.230±.051	0.251±.021	0.239±.028
	REVE	0.241±.038	0.221±.030	0.230±.027	0.238±.051	0.227±.033	0.249±.012	0.234±.032
	CodeBrain	0.243±.043	0.232±.031	0.235±.016	0.247±.012	0.230±.051	0.258±.012	0.241±.028
	UnEBOLT	0.262±.015	0.260±.019	0.255±.025	0.270±.047	0.248±.011	0.272±.023	0.261±.023
	E2fMoE	**0.228**±.012	**0.204**±.012	**0.217**±.050	**0.222**±.013	**0.211**±.045	**0.230**±.040	**0.219**±.029
FC$\uparrow$	EEGNet	0.612±.082	0.381±.142	0.394±.091	0.431±.103	0.286±.096	0.421±.112	0.438±.119
	EEG Conformer	0.381±.091	0.221±.151	0.184±.086	0.236±.097	0.102±.089	0.362±.101	0.283±.137
	BIOT	0.641±.074	0.364±.147	0.401±.092	0.448±.096	0.491±.083	0.536±.089	0.447±.101
	LaBraM	0.623±.079	0.351±.152	0.417±.101	0.472±.094	0.516±.071	0.521±.095	0.446±.108
	NeuroBOLT	0.684±.076	0.452±.133	0.438±.108	0.521±.091	0.548±.078	0.594±.081	0.532±.104
	CBraMod	0.652±.073	0.369±.139	0.407±.088	0.461±.099	0.433±.079	0.476±.094	0.446±.112
	REVE	0.672±.078	0.421±.148	0.447±.104	0.493±.101	0.502±.082	0.517±.097	0.490±.116
	CodeBrain	0.648±.076	0.358±.151	0.398±.081	0.439±.095	0.402±.084	0.451±.099	0.449±.109
	UnEBOLT	0.294±.087	0.156±.136	0.173±.093	0.192±.108	0.071±.085	0.138±.111	0.240±.121
	E2fMoE	**0.731**±.069	**0.516**±.121	**0.482**±.091	**0.574**±.083	**0.603**±.067	**0.641**±.078	**0.583**±.096

REVE. REVE learns reusable EEG embeddings for downstream neural signal analysis El Ouahidi et al. (2025). It provides another modern EEG representation baseline for separating encoding strength from ROI-conditioned BOLD decoding.

CodeBrain. CodeBrain learns code-based brain representations for scalable EEG modeling Ma et al. (2026). We use it to compare against a recent neural coding baseline that does not explicitly align its latent space with BOLD and FC objectives.

NeuroBOLT. NeuroBOLT is a task-specific EEG-to-fMRI synthesis model for reconstructing BOLD signals from EEG Li et al. (2024). It serves as a task-specific reference for evaluating E2fMoE against EEG-to-BOLD reconstruction.

UnEBOLT. UnEBOLT jointly studies EEG-to-BOLD prediction and FC reconstruction Li et al. (2026). It is the closest EEG-to-fMRI baseline for testing whether E2fMoE adds gains through latent state modeling, expert routing, and FC calibration.

Appendix E. Additional Experimental Results

E1. Inter-subject ROI-wise Reconstruction Results. Table A2 reports inter-subject ROI-wise reconstruction results. It groups the six NeuroBOLT targets into primary sensory, high-level cognitive, and subcortical regions.

E2fMoE achieves the best R, MSE, and FC on all six reported NeuroBOLT ROIs and overall. The gains are especially clear for Heschl, Prec. Ant., and Putamen, where temporal accuracy and connectivity preservation improve together.

E2. Paired Statistical Significance on NeuroBOLT. We assess whether the NeuroBOLT gains remain consistent over 203 paired scan–ROI units. Corr deltas are defined as E2fMoE minus baseline, while MSE deltas reverse the subtraction so positive values always favor E2fMoE.

Figure A1 reports paired bootstrap 95% confidence intervals. Against NeuroBOLT, E2fMoE improves Corr by 0.0705 with a 95% CI of [0.0563, 0.0851] and MSE by 0.0223 with a 95% CI of [0.0139, 0.0302]. Table A5 shows that all FDR-corrected paired tests remain significant, indicating that the gains are not driven by a small number of favorable units.

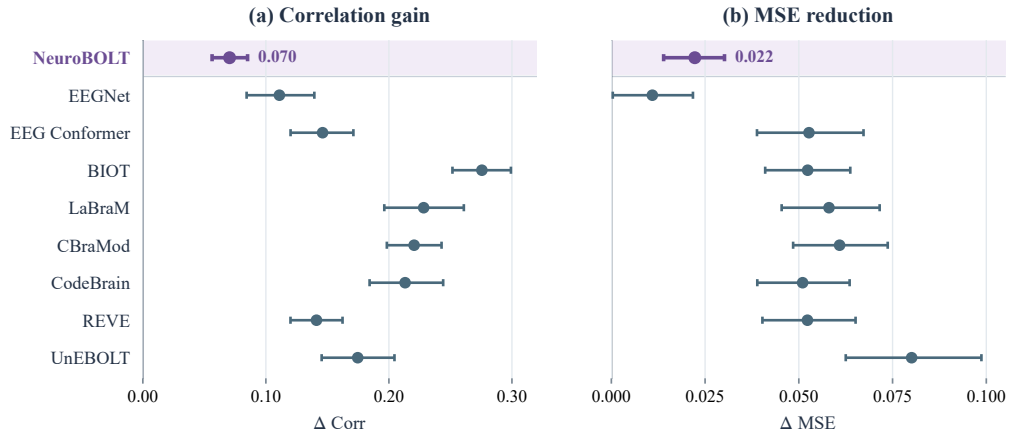


Figure A1: Paired bootstrap 95% confidence intervals for E2fMoE gains over nine baselines on NeuroBOLT (203 scan–ROI units). Correlation and MSE deltas are defined so positive values favor E2fMoE.

Table A5: Paired NeuroBOLT significance summary over 203 scan–ROI units. All listed 95% confidence intervals exclude zero, and all FDR-corrected paired tests satisfy $q < 0.05$.

Baseline	Corr ↑			MSE ↓		
	Δ 95% CI	Win	d_z	Δ 95% CI	Win	d_z
NeuroBOLT	.070 [.056, .085]	74.4%	.664	.022 [.014, .030]	65.0%	.379
EEGNet	.111 [.084, .139]	67.0%	.548	.011 [.000, .022]	49.8%	.140
EEG Conformer	.146 [.120, .171]	80.8%	.786	.053 [.039, .067]	68.5%	.507
BIOT	.276 [.252, .299]	96.6%	1.572	.052 [.041, .064]	70.0%	.624
LaBraM	.228 [.196, .261]	86.2%	.968	.058 [.045, .072]	73.9%	.606
CBraMod	.220 [.198, .243]	91.1%	1.353	.061 [.048, .074]	74.4%	.664
CodeBrain	.213 [.184, .244]	89.7%	.978	.051 [.039, .064]	72.9%	.571
REVE	.141 [.120, .162]	84.2%	.910	.052 [.040, .065]	74.9%	.573
UnEBOLT	.175 [.145, .205]	77.8%	.818	.080 [.063, .099]	72.4%	.603

E3. Expert Specialization and Contribution. To complement the routing visualization in RQ3, Table A3 reports the ROI-level soft routing distributions. Sensory, association, and subcortical targets exhibit different dominant experts, while the global target retains a more distributed mixture. Tables A4 and A6 evaluate each expert through removal tests. The largest target-specific drops occur when the corresponding dominant expert is removed, and the MI-NF Motor column provides a representative result outside the NeuroBOLT targets.

Table A3: ROI-level expert routing distributions on NeuroBOLT. Bold values identify the dominant expert for each target.

ROI	Category	E1	E2	E3	E4	Dom.
Cuneus	Visual	**0.41**	0.29	0.18	0.12	E1
Heschl	Auditory	0.23	**0.46**	0.19	0.12	E2
Mid. Frontal	Association	0.21	0.20	**0.42**	0.17	E3
Prec. Ant.	Association	0.19	0.22	**0.45**	0.14	E3
Putamen	Subcortical	0.16	0.20	0.25	**0.39**	E4
Thalamus	Subcortical	0.17	0.21	0.23	**0.39**	E4
Global	Whole brain	0.25	0.24	**0.29**	0.22	E3

Table A4: Contribution of individual experts on NeuroBOLT ROI-wise reconstruction.

Variant	Overall $R\uparrow$	MSE $\downarrow$	FC $\uparrow$
Full E2fMoE	**0.575**	**0.238**	**0.772**
w/o Expert 1	0.561	0.244	0.751
w/o Expert 2	0.558	0.246	0.748
w/o Expert 3	0.563	0.243	0.756
w/o Expert 4	0.559	0.245	0.749

Table A6: Representative target-specific degradation after expert removal. More negative values indicate larger losses.

Removed	Cuneus ΔR	Heschl ΔR	MI-NF Motor ΔR	Putamen ΔR
w/o E1	**-0.037**	-0.011	-0.014	-0.010
w/o E2	-0.014	**-0.041**	-0.013	-0.012
w/o E3	-0.015	-0.016	**-0.039**	-0.018
w/o E4	-0.011	-0.012	-0.017	**-0.036**

E4. FC Source Ablation. Table A7 gives the ROI-level FC scores for the seven FC-source variants in Table 3(a). The largest reductions occur when the FC template is removed or shuffled. The per-ROI values also show that the low-rank correction and FC loss contribute beyond the base model.

Table A7: ROI-level FC source ablation on NeuroBOLT and MI-NF. Values are mean $\pm$ standard deviation.

(a) NeuroBOLT sensory, cognitive, and subcortical ROIs

Variant	Cuneus	Heschl	Mid. Front.	Prec. Ant.	Putamen	Thalamus
Full E2fMoE	$.793 \pm .073$	$.718 \pm .118$	$.499 \pm .094$	$.806 \pm .047$	$.784 \pm .046$	$.775 \pm .072$
No FC template	$.681 \pm .089$	$.562 \pm .137$	$.397 \pm .108$	$.646 \pm .081$	$.594 \pm .083$	$.632 \pm .090$
Shuffled FC template	$.556 \pm .106$	$.451 \pm .151$	$.318 \pm .119$	$.537 \pm .096$	$.463 \pm .102$	$.501 \pm .108$
No low-rank correction	$.724 \pm .081$	$.631 \pm .128$	$.439 \pm .102$	$.711 \pm .070$	$.682 \pm .067$	$.697 \pm .080$
FC loss only	$.579 \pm .101$	$.463 \pm .146$	$.337 \pm .114$	$.552 \pm .093$	$.485 \pm .096$	$.532 \pm .103$
No FC loss	$.742 \pm .079$	$.649 \pm .126$	$.451 \pm .100$	$.732 \pm .067$	$.711 \pm .063$	$.716 \pm .078$
Base model	$.508 \pm .112$	$.376 \pm .158$	$.281 \pm .125$	$.461 \pm .110$	$.392 \pm .113$	$.438 \pm .117$

(b) MI-NF motor ROIs and NeuroBOLT overall FC

Variant	Precentral	Postcentral	Overall
Full E2fMoE	$.362 \pm .802$	$.860 \pm .077$	$.772 \pm .076$
No FC template	$.224 \pm .817$	$.741 \pm .101$	$.612 \pm .091$
Shuffled FC template	$.105 \pm .834$	$.622 \pm .121$	$.503 \pm .104$
No low-rank correction	$.286 \pm .811$	$.791 \pm .089$	$.684 \pm .084$
FC loss only	$.139 \pm .827$	$.638 \pm .116$	$.512 \pm .097$
No FC loss	$.302 \pm .808$	$.806 \pm .086$	$.701 \pm .082$
Base model	$.061 \pm .846$	$.564 \pm .133$	$.435 \pm .105$

E5. Computational Efficiency. We benchmark all models on one NVIDIA H200 GPU using the same 16-s EEG input (26×3200) and seven-ROI output. Inference and training use BF16 computation with FP32 parameters and AdamW states. Batch-1 latency is averaged over 20 forward

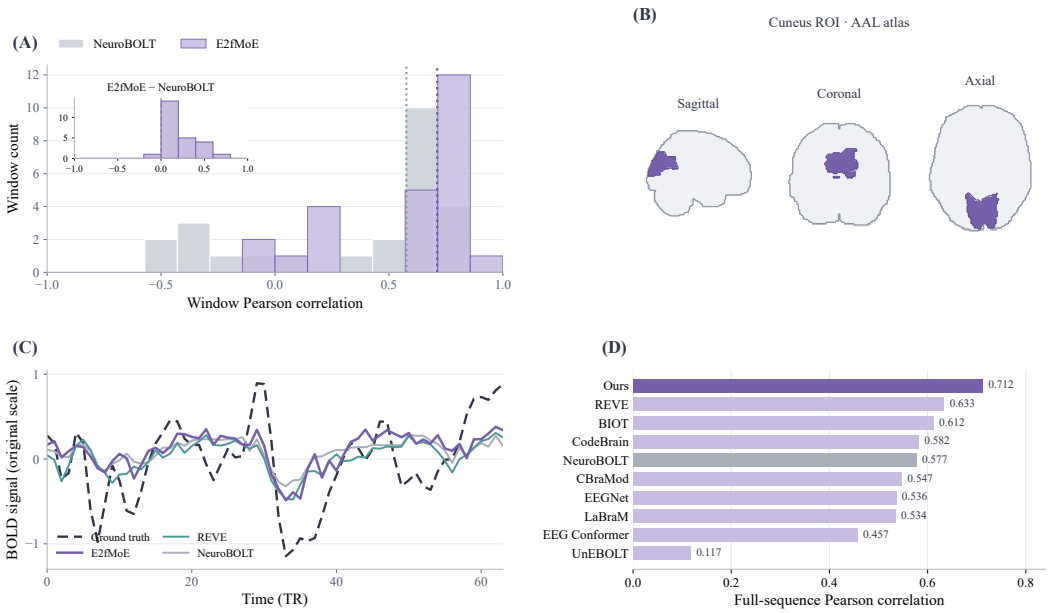


Figure A2: Same-input Cuneus reconstruction. (A) Pearson correlations over 16-TR windows (stride 2) for E2fMoE and NeuroBOLT; the inset shows paired differences and dotted lines mark full-sequence correlations. (B) Bilateral Cuneus in three projections of the AAL SPM12 atlas by Tzourio-Mazoyer et al. (2002); purple marks anatomy only. (C) Original-scale 64-TR BOLD truth and predictions. (D) Full-sequence correlations for all ten methods.

passes, and batch-8 training cost includes the forward, backward, and optimizer-update steps; five warm-up iterations are excluded. Peak memory denotes CUDA peak allocated memory. Corr and MSE are averaged over 203 paired scan–ROI units under the NeuroBOLT diagnostic protocol.

Table A8 shows that E2fMoE achieves the best reconstruction results while remaining substantially lighter than NeuroBOLT. It uses 0.423M parameters, reducing the parameter count by 94.7%, and provides 28.4× lower batch-1 latency, 22.7× faster training steps, and 88.7% lower training peak memory. Compact EEG encoders remain competitive in individual efficiency metrics, but their reconstruction accuracy is lower, indicating that E2fMoE offers a favorable accuracy–efficiency trade-off.

Table A8: Efficiency comparison across representative EEG-to-fMRI reconstruction baselines. Best and second-best results are highlighted with the same colors used in the main tables.

Model	Corr ↑	MSE ↓	Params (M) ↓	FP32 (MiB) ↓	Latency (ms) ↓	Infer. Mem. (MiB) ↓	Train Step (ms) ↓	Train Mem. (MiB) ↓
NeuroBOLT	0.516	0.245	7.908	30.930	22.336	377.8	80.950	1850.1
EEGNet	0.475	0.234	**0.107**	**0.409**	**0.516**	217.1	**3.484**	**193.2**
EEG Conformer	0.440	0.276	0.462	1.763	1.350	350.8	5.304	295.4
BIOT	0.311	0.276	0.439	1.673	1.040	245.3	5.301	320.0
LaBraM	0.358	0.281	0.432	1.649	1.133	253.7	4.947	321.8
CBraMod	0.366	0.284	0.571	2.179	1.473	240.7	7.136	370.6
CodeBrain	0.373	0.274	0.516	1.969	1.724	405.4	8.336	622.0
REVE	0.445	0.276	0.444	1.718	1.177	266.7	5.175	334.6
UnEBOLT	0.411	0.303	0.478	1.832	2.273	210.1	8.717	342.8
Ours	**0.586**	**0.223**	0.423	1.625	0.786	**200.3**	3.568	208.7

E6. Qualitative Reconstruction Visualization. Figure A2 compares E2fMoE with representative external baselines on the same EEG input. The correlation ranking covers all ten methods, while the temporal panel shows the target signal alongside E2fMoE, REVE, and NeuroBOLT on a shared original-scale axis.

E7. Additional EEG-Conditioned BOLD Prediction Examples. Figure A3 displays six further cases in the same format as Figure 3 in the main paper.

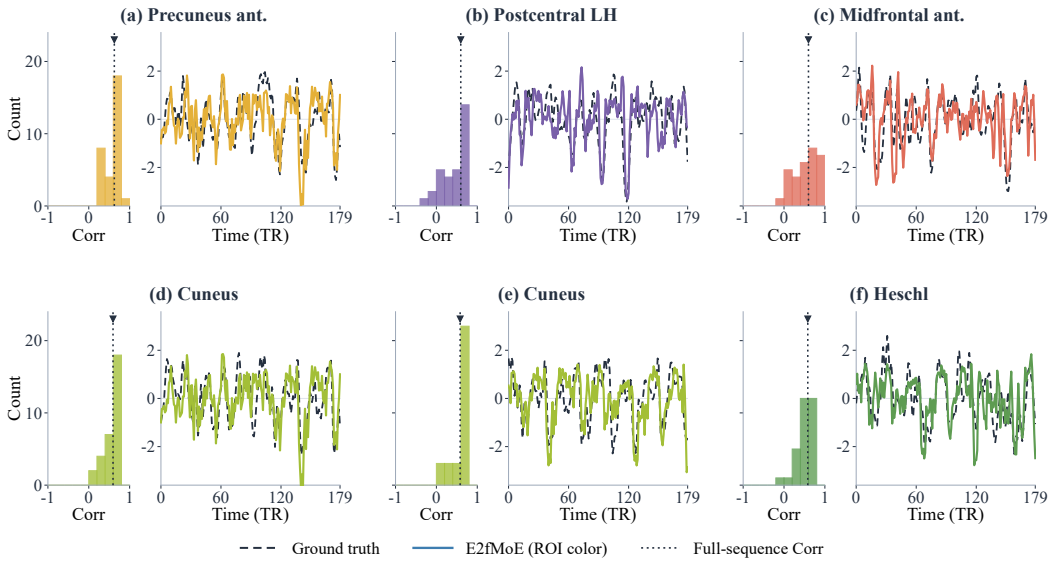


Figure A3: Additional EEG-conditioned BOLD prediction examples. Histograms show Pearson correlations over 30-TR windows with a 5-TR stride; dotted lines indicate full-sequence correlations. Curves show the ground-truth and predicted BOLD trajectories.

Appendix F. Neuroscience Interpretation

The ROI-wise results in Table 1 span sensory, association, subcortical, and motor systems, suggesting that E2fMoE does not improve only one favorable ROI type. Its soft expert routing shares common hemodynamic computations while retaining region-dependent decoding, consistent with shared neurovascular mechanisms and regional response variability Handwerker et al. (2004); Georgopoulos et al. (2024); Mangini et al. (2025). FC calibration further encourages the reconstructed signals to preserve cross-ROI coordination rather than only point-wise waveform similarity Friston (1994).

Appendix G. Limitations

The supplementary visualizations focus on representative ROI-level predictions and routing behavior rather than exhaustively showing every scan, subject, and ROI. This choice keeps the appendix readable while the complete quantitative evidence is still provided by the ROI-wise, global, ablation, sensitivity, and efficiency tables. Future work can extend the visualization protocol with more interactive subject-level inspection tools.

Table A9: Exact NeuroBOLT values underlying the grouped bars in Table 3(a).

Architecture and experts	$R\uparrow$	MSE↓	FC↑
Full E2fMoE	0.575	0.238	0.772
No latent state	0.421	0.251	0.603
No state-space mixer	0.447	0.247	0.635
No MoE	0.392	0.254	0.581
No ROI residual	0.272	0.244	0.361
Single expert	0.291	0.278	0.298
2 experts	0.438	0.244	0.586
4 experts	0.548	0.240	0.731
6 experts	0.566	0.239	0.756

FC controls	$R\uparrow$	MSE↓	FC↑
No FC template	0.568	0.241	0.612
Shuffled FC template	0.557	0.246	0.503
No low-rank correction	0.571	0.239	0.684
FC loss only	0.559	0.244	0.512
No FC loss	0.572	0.239	0.701
Base model	0.551	0.247	0.435

Appendix H. ROI-Discriminative Representation Analysis

To examine whether representations encode ROI-specific structure, Figure A4 compares ROI-level t-SNE embeddings from E2fMoE and NeuroBOLT on matched test windows. E2fMoE forms more compact and separated ROI-related groups, whereas the NeuroBOLT representations are more mixed across several regions. This contrast indicates that latent BOLD state modeling does not collapse different brain targets into a single undifferentiated representation. Instead, ROI-conditioned expert decoding preserves region-dependent organization while still allowing the model to share hemodynamic information across targets. Consistent with the soft routing patterns in RQ3, these visualizations suggest that E2fMoE improves reconstruction by learning an intermediate representation more closely aligned with ROI-specific BOLD dynamics.

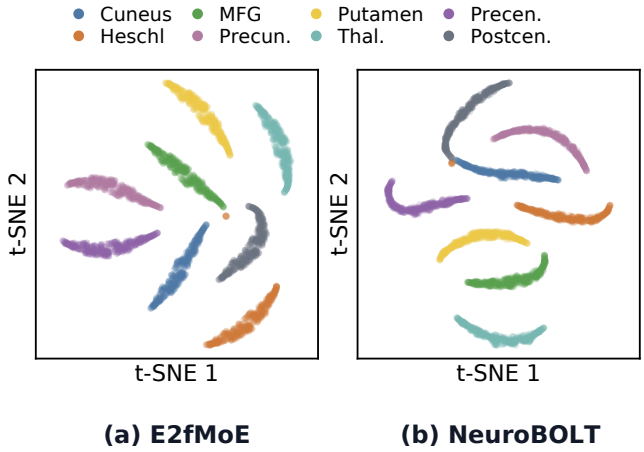


Figure A4: t-SNE visualization of ROI-level latent representations learned by E2fMoE and NeuroBOLT on matched test windows. Colors indicate ROI identities.

Appendix I. ROI-to-Expert Soft Routing

Figure A5 shows the ROI-to-expert flow diagram. The width of each connection represents its mean soft routing weight.

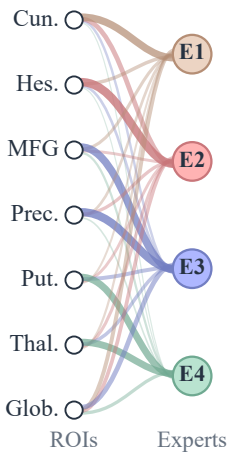


Figure A5: ROI-to-expert soft routing for representative brain targets. Connection widths encode mean gate weights over the four experts.