MARBLE: Lightweight EEG-to-fMRI Translation via Mamba-Attention and ROI-Conditioned Decoding

Emil Kim[1][0009−0001−5344−7549] and Jin Kyu Gahm[2][0000−0001−8827−402X]

[1] Dept. of Information Convergence Engineering, Pusan National University, Busan 46241, Republic of Korea
[2] School of Computer Science and Engineering, Pusan National University, Busan 46241, Republic of Korea
gahmj@pusan.ac.kr

Abstract. Functional magnetic resonance imaging (fMRI) provides high-resolution whole-brain activity measurements but is constrained by high cost, limited accessibility, and restrictive acquisition conditions. Recent advances in cross-modal synthesis have enabled the inference of fMRI signals from electroencephalography (EEG), facilitating estimation of whole-brain activity from portable, low-cost recordings in naturalistic settings. Despite promising results, existing approaches rely on Transformer-based encoders that lack inductive biases for the causal, delayed nature of hemodynamic responses, and provide limited insight into how region-specific predictions are formed. In this paper, we propose MARBLE (Mamba-Attention for ROI-conditioned BOLD Estimation), an end-to-end framework for whole-brain EEG-to-fMRI translation. MARBLE employs a hybrid Mamba-Transformer encoder, where selective state-space modeling provides an inductive bias for HRF-related temporal structure while maintaining linear-time complexity. ROI-conditioned decoding heads operating on a shared latent representation generate region-specific BOLD time series. Our approach provides a favorable trade-off between reconstruction performance and computational efficiency, using 3.2× fewer model parameters and 2.3× less peak memory. Reconstructed functional connectivity matrices demonstrate preserved temporal dynamics and inter-regional structure, and cross-attention topographies reveal network-specific electrode contributions without explicit spatial supervision.

Keywords: EEG-fMRI Translation · Cross-Modal Synthesis · State-Space Models · Functional Connectivity

1 Introduction

Functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) are complementary, non-invasive neuroimaging modalities offering high spatial and temporal resolution [21, 28, 1], respectively. fMRI measures blood-oxygen-level-dependent (BOLD) signal changes associated with neuronal activity, which provide an indirect characterization of distributed neural processes. However, the intrinsic delay and temporal smoothing introduced by the hemodynamic response function (HRF) [2] limit its ability to capture rapid neural dynamics. In addition, fMRI acquisition is costly [7, 22], time-consuming [30], and constrained by its lack of portability [7, 22], thereby restricting its applicability in real-world and resource-limited clinical settings. In contrast, EEG directly records electrical brain activity at millisecond-level temporal resolution [11, 5], providing real-time insight into neural dynamics. Owing to its portability and low cost, EEG has emerged as a promising modality for estimating fMRI BOLD signal dynamics through cross-modal translation frameworks [3, 4] that reconstruct spatially resolved hemodynamic activity from temporally rich electrophysiological signals.

Early EEG–fMRI translation approaches employed encoder–decoder architectures [14, 17] within sequence-to-sequence frameworks to reconstruct regional fMRI time series. While demonstrating feasibility, these approaches typically focused on a limited set of regions of interest (ROIs) and relied on subject-specific training [18], thereby restricting cross-subject generalization and whole-brain scalability. In addition, most studies concentrated on task-based or eyes-open resting-state conditions, leaving scalable whole-brain modeling under eyes-closed resting-state conditions largely unaddressed. Resting-state fMRI captures spontaneous intrinsic neural dynamics and underpins functional connectivity analysis [23, 15]. Its minimal behavioral demands and widespread clinical use make it a compelling setting for developing robust, clinically transferable EEG-to-fMRI translation frameworks.

Recent work has made progress toward whole-brain modeling. NeuroBOLT [18] integrated temporal, spatial, and spectral EEG representations to achieve state-of-the-art performance on eyes-closed resting-state data, and UnEBOLT [16] extended this to a unified architecture that jointly predicts all brain regions via learnable ROI embeddings and adapter-based decoding. However, both methods build on Transformer-based encoders that treat EEG as a generic sequence, without incorporating inductive biases for the causal, delayed temporal structure imposed by the HRF. Furthermore, their ROI conditioning mechanisms offer limited transparency into how each region's prediction is derived from the underlying EEG features.

To overcome these limitations, we propose MARBLE (Mamba-Attention for ROI-conditioned BOLD Estimation), a lightweight framework for whole-brain EEG-to-fMRI translation that enables joint and efficient modeling across brain regions. MARBLE introduces a hybrid Mamba-Transformer encoder where selective state-space modeling provides an architectural inductive bias for HRF-related temporal structure, while attention mechanisms further refine the resulting latent representation. From this shared representation, dedicated ROI-wise decoders with cross-attention generate region-specific BOLD predictions. This design supports interpretable multi-region inference while substantially reducing computational and memory overhead. Our contributions are summarized as follows:

1. A hybrid Mamba-Transformer encoder that provides an inductive bias for HRF-related temporal structure via state-space modeling, achieving 3.2× fewer parameters and 2.3× lower memory than prior methods.
2. Interpretable ROI-wise cross-attention decoders that generate region-specific predictions from a shared latent representation.
3. A favorable balance between reconstruction performance and computational efficiency while preserving functional connectivity patterns essential for downstream clinical analysis.

2 Methods

2.1 Problem Definition

The mapping between neural activity and BOLD responses is neither instantaneous nor one-to-one [18, 12]. Thus, we adopt a sequence-to-one modeling strategy that aligns temporally rich EEG inputs with delayed BOLD responses. Specifically, for each fMRI time point t, we extract a preceding EEG window spanning the hemodynamic response function (HRF) support. Formally, the EEG input is represented as $X_t \in \mathbb{R}^{C\times T}$, where C denotes the number of channels and T denotes the number of temporal samples in the window.

Directly modeling full 4D fMRI volumes is computationally prohibitive [26, 6]. Therefore, we operate at the level of image-derived phenotypes obtained through brain parcellation. Voxel-wise signals are averaged within predefined regions of interest (ROIs), yielding stable regional time courses while preserving large-scale functional organization. The fMRI target at time t is thus represented as $Y_t \in \mathbb{R}^P$, where P denotes the number of ROIs. Given a parameterized model $f_\theta(\cdot)$, the task is formulated as $\hat{Y}_t = f_\theta(X_t)$, with $\hat{Y}_t$ denoting the predicted ROI-level BOLD signal at time t.

2.2 Model Architecture

We propose MARBLE, a lightweight end-to-end framework for predicting multi-region BOLD time series from EEG. The architecture comprises (i) multi-scale spectral embedding (MSSE), (ii) a shared EEG encoder and (iii) a ROI-wise fMRI decoder (Fig. 1).

MSSE. To capture oscillatory dynamics relevant to BOLD fluctuations, we employ an MSSE inspired by NeuroBOLT [18]. Given EEG input $X \in \mathbb{R}^{C\times T}$, we extract multi-scale spectral representations using the short-time Fourier transform (STFT). For each scale level $l \in \{0, \dots, L\}$, the Fourier window size is defined as $n_{\text{fft}}^{(l)} = l_{\text{base}} \cdot 2^l$, where l_{base} is a base window size. In our implementation, we set $l_{\text{base}} = 100$ and $L = 2$. The hop length is set to half the window size, resulting in 50% overlap, a common choice in EEG time-frequency analysis [29]. STFT [9] is applied independently per channel, and the magnitude of the complex spectrum is retained. Spectral features at each scale are linearly projected to

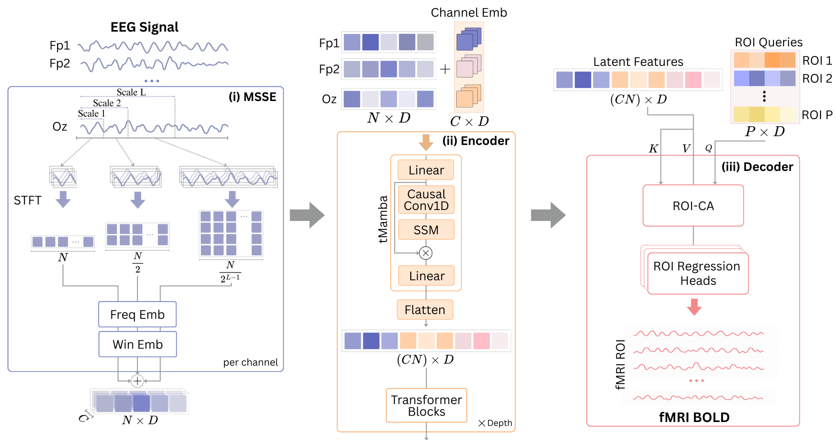

Fig. 1. Overview of MARBLE. A 16-second EEG segment is processed through (i) multi-scale spectral embedding (MSSE), (ii) a shared encoder combining Mamba and Transformer blocks, and (iii) a ROI-wise decoder that generates region-specific BOLD predictions via cross-attention.

dimension D, then passed through a second linear projection along time to unify all scales to a common temporal resolution $N = \lfloor \frac{T}{l_{\text{base}}} \rfloor$, producing scale-specific embeddings $e^{(l)}$. These are aggregated via element-wise summation: $e_{\text{sp}} = \sum_{l=0}^{L} e^{(l)}$, where $e_{\text{sp}} \in \mathbb{R}^{C \times N \times D}$.

Channel identity is preserved via learnable channel embeddings $e_{\text{ch}} \in \mathbb{R}^{C \times D}$, which are added to the spectral embedding at each temporal position. Finally, sinusoidal positional encoding [27] is applied along the temporal dimension, yielding the final encoder input $e \in \mathbb{R}^{C \times N \times D}$.

Encoder. To model temporal EEG dynamics, we employ a temporal Mamba (tMamba) block. Mamba, a selective state-space sequence model (SSM) [10], represents temporal evolution through a learnable state-space formulation that propagates information via a recurrent hidden state. In continuous time, the SSM is defined as

$$\dot{h}(t) = Ah(t) + Bx(t), \quad y(t) = Ch(t), \tag{1}$$

where $\dot{h}(t)$ denotes the latent state and A, B, C are learnable parameters. After discretization, the model yields a structured convolutional form, enabling efficient unidirectional sequence processing. This unidirectional processing provides an inductive bias for HRF-related temporal structure, consistent with delayed, causal dependence of the BOLD response on preceding neural activity, while maintaining linear-time complexity.

Following the tMamba layers, we apply a series of Transformer [27] blocks to model inter-channel interactions. The encoder representation is reshaped into a flattened channel–time sequence $Z \in \mathbb{R}^{(CN) \times D}$, allowing self-attention to capture global dependencies across channels and time.

Decoder. We introduce a ROI-wise cross-attention (ROI-CA) mechanism to generate region-specific BOLD predictions from the shared EEG latent representation. Let $Z \in \mathbb{R}^{(CN)\times D}$ denote the encoded EEG features. We initialize a set of learnable ROI query embeddings $Q \in \mathbb{R}^{P\times D}$, where P denotes the number of ROIs. Each query vector q_p serves as a region-conditioned token that selectively attends to relevant EEG features. For each ROI, cross-attention is performed using the corresponding ROI query and the encoded EEG representation as key–value pairs:

$$z_p = \mathrm{CrossAttn}(q_p, Z, Z), \tag{2}$$

yielding ROI-specific feature representations $z_p \in \mathbb{R}^D$.

To produce final BOLD predictions, we employ independent linear regression heads for each ROI. This design encourages region-specific specialization while maintaining a shared global representation in the encoder, allowing region-dependent signal mappings without sacrificing shared cross-regional structure.

3 Experiments

3.1 Dataset

We conducted experiments on the VU dataset from NeuroBOLT [18], adopting the same preprocessing procedures. The dataset consists of 29 simultaneous EEG–fMRI recordings from 22 healthy volunteers during eyes-closed resting-state conditions, with 7 participants contributing two separate sessions. Each recording session has a duration of approximately 20 minutes. Scalp EEG was acquired using a 32-channel MR-compatible system following the international 10–20 placement scheme. After removing ECG, EOG, and EMG channels, 26 EEG channels were retained for analysis. fMRI regional time series were extracted using the Dictionary of Functional Modes (DiFuMo) [6] atlas (64 ROIs). For each fMRI time point, a 16-second segment of preceding EEG data was extracted and resampled to 200 Hz. We used a subject-wise partitioning strategy with a split ratio of 3:1:1, ensuring that scans from the same subject appear only within a single split.

3.2 Experimental Settings

All experiments were implemented in PyTorch 2.4.1 using Python 3.10.12 and conducted on a single NVIDIA H100 GPU with CUDA 12.4. Models were trained for 20 epochs with a batch size of 64 using the AdamW optimizer [20] (lr 3×10^{-4}, weight decay 0.05) with cosine scheduling [19] (min lr 1×10^{-6}). MARBLE uses an embedding dimension of 64 and four encoder layers. The tMamba module employs a state size of 16 and convolution width of 4, with dropout set to 0.2 throughout the network. The ROI-conditioned cross-attention decoder uses 64 learnable ROI queries with the same embedding dimension. The training objective combines mean squared error (MSE) with temporal and spatial correlation terms:

$$\mathcal{L} = \lambda_{mse}\mathcal{L}_{mse} + \lambda_t\mathcal{L}_t + \lambda_s\mathcal{L}_s,$$

Table 1. Quantitative comparison on 6 unseen test scans from the VU dataset. Results are reported as mean ± standard deviation. Bold indicates best performance. Gray-shaded rows indicate ablation studies.

Model	Corr (r) ↑	MSE ↓	FC-Corr ↑	FC-F1 ↑	Params ↓	Mem ↓
BIOT	0.396 ± 0.041	0.229 ± 0.020	0.439 ± 0.144	0.404 ± 0.025	**2.0M**	**4.4 GB**
LaBraM	0.179 ± 0.075	0.250 ± 0.009	0.359 ± 0.114	0.429 ± 0.006	5.8M	11.5 GB
NeuroBOLT	0.406 ± 0.038	0.223 ± 0.017	0.469 ± 0.115	0.418 ± 0.012	7.9M	13.7 GB
UnEBOLT	0.412 ± 0.049	0.218 ± 0.019	$\mathbf{0.469 \pm 0.093}$	0.427 ± 0.016	8.2M	13.8 GB
MARBLE	$\mathbf{0.434 \pm 0.047}$	$\mathbf{0.210 \pm 0.020}$	0.456 ± 0.129	$\mathbf{0.454 \pm 0.024}$	2.6M	6.0 GB
W/O ROI CA	0.413 ± 0.040	0.218 ± 0.019	0.395 ± 0.084	0.431 ± 0.018	-	-
W/O tMamba	0.424 ± 0.042	0.211 ± 0.018	0.440 ± 0.125	0.442 ± 0.019	-	-

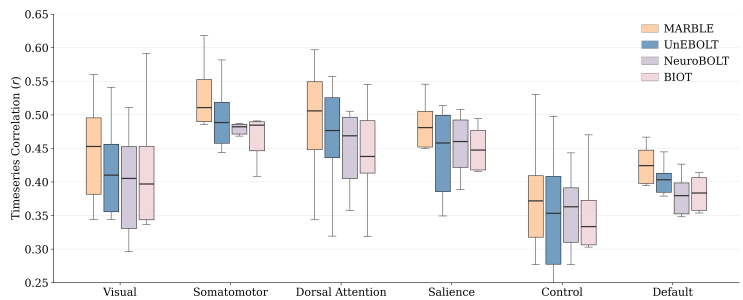

Fig. 2. ROI-level correlation distribution across functional networks on unseen test scans. Boxes show median (center line), IQR (box), and full range (whiskers).

where $\mathcal{L}_t = \frac{1}{P}\sum_{p=1}^{P}(1 - r(\hat{Y}_p, Y_p))$ encourages temporal fidelity per ROI, and $\mathcal{L}_s = \frac{1}{B}\sum_{b=1}^{B}(1 - r(\hat{Y}_b, Y_b))$ promotes spatial fidelity across ROIs per time point. Here, $r(\cdot)$ is Pearson correlation coefficient computed after mean-centering along the corresponding axis, and B is batch size. We set $\lambda_{mse} = 0.8$, $\lambda_t = 0.1$, and $\lambda_s = 0.1$ in all experiments.

We evaluated signal-level (Corr, MSE) and connectivity-level (FC-Corr, FC-F1) metrics. FC matrices were computed as Pearson correlations across ROI time series (upper-triangular only). FC-F1 is computed by thresholding at the 75th percentile and comparing predicted edges at the same threshold. To facilitate reproducibility, we provide the implementation and pretrained model weights at `https://github.com/emeelkaa/marble`.

3.3 Experimental Results

We compared MARBLE against four baselines: (i) BIOT [29], a general-purpose biosignal foundation model, (ii) LaBraM [13], an EEG-specific foundation model pre-trained on approximately 2,500 hours of data, (iii) NeuroBOLT [18], a dedicated EEG-to-fMRI translation framework built upon LaBraM, and (iv) UnEBOLT [16], a unified whole-brain extension of NeuroBOLT based on adapter-based decoding. For LaBraM, NeuroBOLT, and UnEBOLT, we initialized with

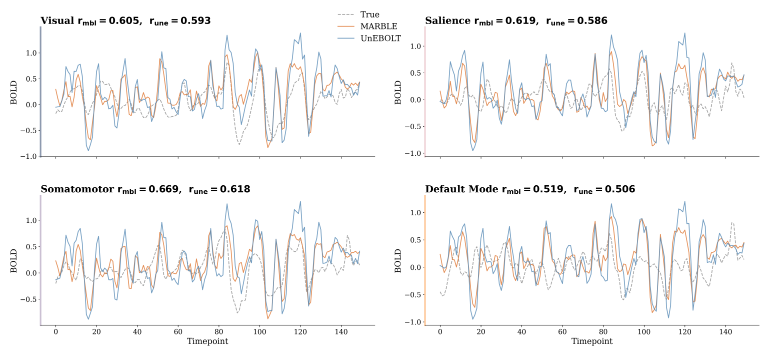

Fig. 3. Representative ROI-level BOLD time-series reconstructions across functional networks. For each selected network, one representative ROI is visualized. Solid orange, solid blue, and dashed gray lines denote MARBLE, UnEBOLT, and ground-truth BOLD signals, respectively.

the publicly available LaBraM-base checkpoint [13]. All baselines were reproduced using publicly available code and trained under identical conditions; for single-region models, we modified only the final regression head to enable multi-region prediction.

As shown in Table 1, MARBLE achieved the best signal-level reconstruction, improving ROI-level Pearson correlation from 0.412 (UnEBOLT) to 0.434 and reducing MSE from 0.218 to 0.210. It also attained the highest FC-F1 score (0.454 ± 0.024), indicating improved preservation of functional topology. These gains were achieved with substantially fewer parameters (2.6M vs. 8.2M) and lower memory usage (6.0 GB vs. 13.8 GB). Ablation studies further support our architectural design. Removing ROI-CA caused the largest drop (FC-Corr 0.061), while removing tMamba primarily affected signal-level metrics (Corr 0.010).

Figure 2 shows the distribution of ROI-level Pearson correlations grouped by the Yeo functional networks [24]. Consistent performance across sensory, attentional, and association networks demonstrates that MARBLE generalized effectively across diverse brain systems while maintaining stable reconstruction fidelity. To further illustrate qualitative reconstruction accuracy, Figure 3 presents representative BOLD time-series predictions from selected networks.

To assess preservation of large-scale functional organization, we compared predicted and ground-truth functional connectivity (FC) matrices (Fig. 4). MARBLE's predictions closely resembled ground-truth connectivity structure, capturing both intra-network blocks and inter-network interactions. When thresholding to retain the top 25% strongest connections, the reconstructed networks demonstrated substantial overlap with the ground-truth high-strength edges, confirming accurate recovery of dominant functional interactions.

To further examine ROI-CA, we visualized attention maps averaged across unseen test subjects (Fig. 5). Distinct topographic biases emerged across networks. Visual and Somatomotor showed posterior contributions, with Somatomotor

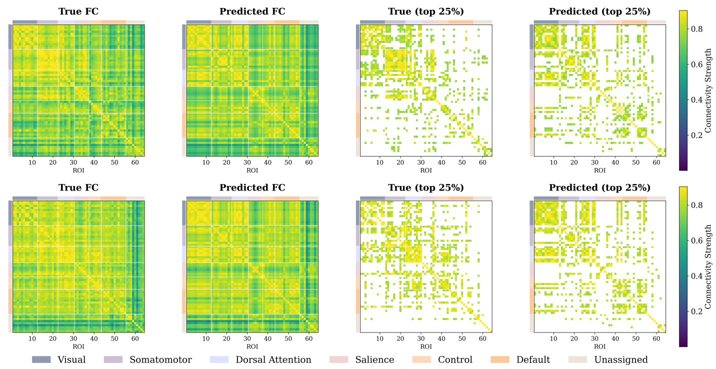

Fig. 4. FC reconstruction on unseen test scans. Each row corresponds to a representative subject, showing ground-truth and predicted FC matrices alongside thresholded top 25% strongest connections. Network assignments are indicated along matrix borders.

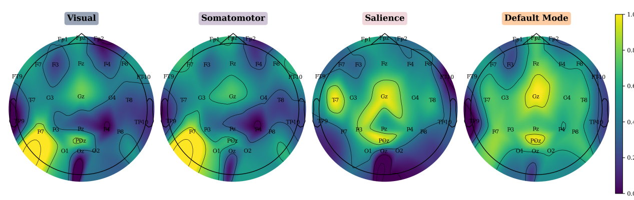

Fig. 5. Cross-attention topography averages across unseen test subjects for Visual, Somatomotor, Salience, and Default Mode networks. Maps are displayed with the nose at the top and the left hemisphere on the left. Color intensity indicates normalized attention weight per electrode.

additionally exhibiting stronger central weighting; Salience and DMN shared midline–posterior components, while Salience demonstrated comparatively greater lateral–temporal emphasis. The posterior pattern for the Visual network aligns with the established posterior dominance of visual-related scalp activity in resting EEG [8], whereas the topographies of Somatomotor, DMN, and Salience warrant further validation via source-level analysis. Shared posterior contributions likely reflect resting-state posterior oscillations and volume conduction inherent to scalp EEG [25]. In contrast to adapter-based approaches where region-specific decoding remains opaque, cross-attention weights provide direct insight into which electrodes drive each regional prediction.

4 Conclusion

We presented MARBLE, a lightweight and interpretable framework for whole-brain EEG-to-fMRI translation. MARBLE combines a hybrid Mamba-Transformer encoder with ROI-conditioned cross-attention decoders to provide a favorable balance between reconstruction performance and computational efficiency, achieving substantially fewer parameters and lower memory than prior approaches. Our analyses confirm preserved functional connectivity across diverse brain networks, while cross-attention topographies reveal network-specific electrode contributions aligned with known EEG–BOLD coupling patterns, providing interpretability absent in adapter-based approaches. This study is limited by the use of a single-site cohort of 22 subjects and eyes-closed resting-state recordings, which may constrain generalizability to broader populations and experimental conditions. Future work will extend MARBLE to larger multi-site cohorts, task-based fMRI, and clinical populations where simultaneous EEG-fMRI acquisition is infeasible.

Acknowledgments. This work was supported by an Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korean Government (MSIT; IITP-2026-RS-2023-00254177, IITP-2026-RS-2024-00360227, IITP-2026-RS-2025-02219116)

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

References

1. Bondi, E., Ding, Y., Zhang, Y., Maggioni, E., He, B.: Investigating the neurovascular coupling across multiple motor execution and imagery conditions: a whole-brain EEG-informed fMRI analysis. NeuroImage **317**, 121311 (2025)
2. Buxton, R.B., Uludağ, K., Dubowitz, D.J., Liu, T.T.: Modeling the hemodynamic response to brain activation. NeuroImage **23**, S220–S233 (2004)
3. Calhas, D., Henriques, R.: EEG to fMRI synthesis: Is deep learning a candidate? arXiv preprint arXiv:2009.14133 (2020)
4. Calhas, D., Henriques, R.: EEG to fMRI synthesis benefits from attentional graphs of electrode relationships. arXiv preprint arXiv:2203.03481 (2022)
5. Chang, C., Chen, J.E.: Multimodal EEG-fMRI: advancing insight into large-scale human brain dynamics. Current Opinion in Biomedical Engineering **18**, 100279 (2021)
6. Dadi, K., Varoquaux, G., Machlouzarides-Shalit, A., Gorgolewski, K.J., Wassermann, D., Thirion, B., Mensch, A.: Fine-grain atlases of functional modes for fMRI analysis. NeuroImage **221**, 117126 (2020)
7. Demene, C., Baranger, J., Bernal, M., Delanoe, C., Auvin, S., Biran, V., Alison, M., Mairesse, J., Harribaud, E., Pernot, M., et al.: Functional ultrasound imaging of brain activity in human newborns. Science Translational Medicine **9**(411), eaah6756 (2017)
8. Goldman, R.I., Stern, J.M., Engel Jr, J., Cohen, M.S.: Simultaneous EEG and fMRI of the alpha rhythm. Neuroreport **13**(18), 2487–2492 (2002)
9. Griffin, D., Lim, J.: Signal estimation from modified short-time Fourier transform. IEEE Transactions on Acoustics, Speech, and Signal Processing **32**(2), 236–243 (1984)
10. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. In: First Conference on Language Modeling (2024)
11. He, B., Sohrabpour, A., Brown, E., Liu, Z.: Electrophysiological source imaging: a noninvasive window to brain dynamics. Annual Review of Biomedical Engineering **20**(1), 171–196 (2018)

12. Heeger, D.J., Ress, D.: What does fMRI tell us about neuronal activity? Nature Reviews Neuroscience **3**(2), 142–151 (2002)
13. Jiang, W.B., Zhao, L.M., Lu, B.L.: Large brain model for learning generic representations with tremendous EEG data in BCI. arXiv preprint arXiv:2405.18765 (2024)
14. Kovalev, A., Mikheev, I., Ossadtchi, A.: fMRI from EEG is only deep learning away: the use of interpretable DL to unravel EEG-fMRI relationships. arXiv preprint arXiv:2211.02024 (2022)
15. Lee, M.H., Smyser, C.D., Shimony, J.S.: Resting-state fMRI: a review of methods and clinical applications. American Journal of Neuroradiology **34**(10), 1866–1872 (2013)
16. Li, Y., Lou, A., Li, C., Wang, S., Pourmotabbed, H., Xu, Z., Zhang, S., Englot, D.J., Kolouri, S., Moyer, D., et al.: UnEBOLT: A unified model for EEG-to-BOLD translation and functional connectivity reconstruction. In: Medical Imaging with Deep Learning (2026)
17. Li, Y., Lou, A., Xu, Z., Wang, S., Chang, C.: Leveraging sinusoidal representation networks to predict fMRI signals from EEG. In: Medical Imaging 2024: Image Processing. vol. 12926, pp. 795–800. SPIE (2024)
18. Li, Y., Lou, A., Xu, Z., Zhang, S., Wang, S., Englot, D., Kolouri, S., Moyer, D., Bayrak, R., Chang, C.: NeuroBOLT: Resting-state EEG-to-fMRI synthesis with multi-dimensional feature mapping. Advances in Neural Information Processing Systems **37**, 23378–23405 (2024)
19. Loshchilov, I., Hutter, F.: SGDR: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)
20. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
21. Ritter, P., Villringer, A.: Simultaneous EEG–fMRI. Neuroscience & Biobehavioral Reviews **30**(6), 823–838 (2006)
22. Scarapicchia, V., Brown, C., Mayo, C., Gawryluk, J.R.: Functional magnetic resonance imaging and functional near-infrared spectroscopy: insights from combined recording studies. Frontiers in Human Neuroscience **11**, 419 (2017)
23. Smith, S.M., Vidaurre, D., Beckmann, C.F., Glasser, M.F., Jenkinson, M., Miller, K.L., Nichols, T.E., Robinson, E.C., Salimi-Khorshidi, G., Woolrich, M.W., et al.: Functional connectomics from resting-state fMRI. Trends in Cognitive Sciences **17**(12), 666–682 (2013)
24. Thomas Yeo, B., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al.: The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology **106**(3), 1125–1165 (2011)
25. Van Dijk, H., Schoffelen, J.M., Oostenveld, R., Jensen, O.: Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. Journal of Neuroscience **28**(8), 1816–1823 (2008)
26. Varoquaux, G., Craddock, R.C.: Learning and comparing functional connectomes across subjects. NeuroImage **80**, 405–415 (2013)
27. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems **30** (2017)
28. Warbrick, T.: Simultaneous EEG-fMRI: what have we learned and what does the future hold? Sensors **22**(6), 2262 (2022)

29. Yang, C., Westover, M., Sun, J.: BIOT: Biosignal transformer for cross-data learning in the wild. Advances in Neural Information Processing Systems **36**, 78240–78260 (2023)
30. Yao, W., Lyu, Z., Mahmud, M., Zhong, N., Lei, B., Wang, S.: CATD: Unified representation learning for EEG-to-fMRI cross-modal generation. IEEE Transactions on Medical Imaging **44**(7), 2757–2767 (2025)