MSGFlowNet: Learning Effective Connectivity Network based on Sparse Generative Flow Network from fMRI and EEG Data

Zhihao Su[1], Jihao Zhai[1], Junzhong Ji[1], and Jinduo Liu[1]*

[1]Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology, Beijing Institute of Artificial Intelligence, College of Computer Science, Beijing University of Technology, Beijing, China.
jinduo@bjut.edu.cn

Abstract. Brain effective connectivity (EC) is key to understanding causal neural interactions and brain organization. However, learning EC from single-modal brain data, such as functional magnetic resonance imaging (fMRI) or electroencephalography (EEG), is limited by the inability to simultaneously capture sparse temporal and spatial information. This paper proposes a novel multimodal sparse generative flow network (MSGFlowNet), which integrates fMRI and EEG data through an attention-guided encoder and employs a multi-head self-attention sparse Transformer to extract features from the fused data. These features are then processed by two output heads of the generative flow network: one computes state transition probabilities and updates the mask, while the other determines the probability of generating a termination state. Experiments on synthetic and real-world datasets demonstrate that MSGFlowNet significantly outperforms state-of-the-art methods.

Keywords: Brain Effective Connectivity · Generative Flow Network · Sparse Transformer · functional Magnetic Resonance Imaging · Electroencephalography.

1 Introduction

Brain effective connectivity (EC) describes the causal relationships between brain regions and is crucial for understanding brain function [7, 17, 30]. Currently, EC learning methods fall into two categories: traditional assumption-based statistical methods and data-driven deep learning methods [21]. Traditional methods like structural equation modeling (SEM) [18], dynamic causal modeling (DCM) [8], and Granger causality analysis (GCA) are widely used but struggle to model complex brain networks due to their strong assumptions.

Deep learning has advanced data-driven EC learning [4]. Graph neural networks (GNNs) leverage brain network topology to model causal relationships, with models such as dynamic diffusion-variational GNN [15] and directed structure learning GNN [2] making progress in brain disease classification. Recurrent

* Corresponding Author

neural networks (RNNs) are well-suited for time-series modeling, with a Granger causality estimator based on the recurrent neural network [29] capturing time lags and nonlinear dependencies.Recurrent neural networks (RNNs) are well-suited for time-series modeling, with a Granger causality estimator based on the recurrent neural network [27]capturing time lags and nonlinear dependencies. Several approaches have been developed to address time-varying causal relationships [5] and developing more rigorous model selection frameworks [13]. Generative adversarial networks (GANs) have shown promise in data augmentation and missing data imputation. Furthermore, the framework incorporates a diffusion model-based EC learning method [12] and integrates dynamic EC with structural EC for brain network classification [3]. However, these approaches primarily infer EC from functional magnetic resonance imaging (fMRI) [25], neglecting the potential of multimodal integration.

fMRI provides high spatial resolution but low temporal resolution, whereas electroencephalography (EEG) offers high temporal resolution [31, 22] but lower spatial resolution. Combining both modalities enhances EC learning accuracy but presents challenges [26]. Anwar et al. inferred EC using GCA but did not fully integrate fMRI and EEG information [19]. Tu et al. proposed a linear state-space model that combines temporal information from both modalities but struggles with nonlinear causal relationships [28]. Liu et al. proposed a novel method to bridge modality differences but failed to address high-dimensional sparsity after data fusion [16]. Efficient integration of fMRI and EEG, while overcoming high-dimensional sparsity, remains a key challenge.

To address these issues, we propose a novel multimodal sparse generative flow network (MSGFlowNet). The model consists of three main stages: data fusion, sparse feature extraction, and state generation. MSGFlowNet redefines the EC learning problem as a generative task and effectively fuses fMRI and EEG data using an attention-based encoder. Subsequently, a sparse Transformer with multi-head self-attention is employed to extract features from the fused data, and two output heads in the generative flow network (GFN) compute the EC state and state transition probabilities. Finally, the EC is updated and generated.

Our contributions are summarized as follows:

1. This is the first study that develops a sparse generative flow network to learn brain effective connectivity from fMRI and EEG data.

2. We utilize fMRI and EEG data as complementary modalities, effectively leveraging the high spatial resolution of fMRI and the high temporal resolution of EEG to overcome the limitations of using a single modality.

3. We propose a novel sparse Transformer architecture that effectively extracts key features from high-dimensional sparse fused fMRI-EEG data, significantly enhancing EC learning accuracy.

2 Related work

Brain effective connectivity. The brain EC can represent as a directed graph $G = (V, E, W)$, containing $|V|$ nodes and $|E|$ directed edges, which reflect the

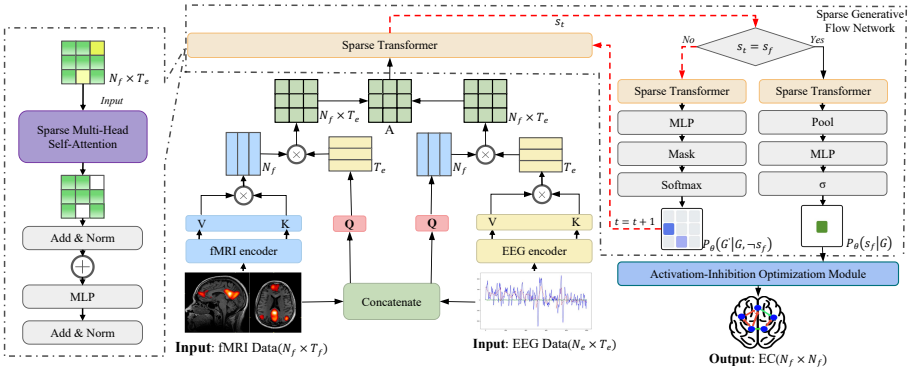

Fig. 1. Flowchart of MSGFlowNet. The framework processes multimodal data through three stages: (1) fusion of fMRI and EEG signals, (2) sparse Transformer for feature extraction, and (3) generation of EC states with GFN.

regions of interest (ROIs) in the brain and the causal relationships between these regions, respectively. The weight or adjacency matrix W is an asymmetric matrix used to represent the causal effects between brain regions of brain data.

Generative flow network. GFN is a probabilistic model over discrete sample spaces with a compositional structure [6], which is a recently introduced method to train energy-based generative models. They have now been successfully applied to a variety of settings such as biological sequences, causal discovery, discrete latent variable modeling, and computational graph scheduling.

3 Method

In this section, we present MSGFlowNet, a novel three-stage method for learning brain effective connectivity networks from complementary fMRI and EEG data.

3.1 Main Idea

To accurately learn EC from fMRI and EEG data, we propose MSGFlowNet model. MSGFlowNet redefines the EC learning problem as a generative task and effectively fuses fMRI and EEG data using an attention-based encoder. Subsequently, a sparse Transformer with multi-head self-attention is employed to extract features from the fused data, and two output heads in the generative flow network compute the EC state and state transition probabilities. Finally, the EC is updated and generated.

3.2 fMRI and EEG Feature Fusion

We propose attention-guided a multimodal encoder for integrating fMRI and EEG signals into a unified representation, thereby enabling efficient cross-modal

information integration [20]. By combining the fMRI signal x_f and EEG signal x_e, we introduce fMRI and EEG source encoding to align their spatial and temporal resolutions. These encoding modules solve the inverse problems for both modalities, generating fMRI and EEG source-level time series (s_f and s_e). For fMRI, the BOLD inverse problem is defined by the equation:

$$s_f = x_f H + \epsilon_f, \tag{1}$$

where $x_f \in \mathbb{R}^{N_f \times T_f}$ is the measured fMRI signal of N_f brain regions during T_f instants, $s_f \in \mathbb{R}^{N_f \times T_e}$ is the fMRI sources-level time series, and T_e is the duration of the EEG signal. $H \in \mathbb{R}^{T_f \times T_e}$ is the BOLD response matrix consisting of trainable neural network parameters. ϵ_f is the Gaussian noise. For EEG, the inverse problem estimates the neural sources as:

$$s_e = I x_e + \epsilon_e, \tag{2}$$

where $x_e \in \mathbb{R}^{N_e \times T_e}$ is the measured EEG signal of N_e electrodes during T_e instants. $s_e \in \mathbb{R}^{N_f \times T_e}$ is the EEG sources-level time series. $I \in \mathbb{R}^{N_e \times N_f}$ is a leadfield matrix consisting of trainable neural network parameters. ϵ_e is the Gaussian noise.

In the attention-guided layer, the query(Q_m), key(K_m), and value(V_m) vectors are generated from the input data through fully connected layers. Specifically, Q_m is derived from the input s_m, while K_m and V_m are extracted from the fMRI source-level time series s_f and EEG source-level time series s_e. The attention map A is generated by computing the dot product between Q_m and K_m. Then, A is multiplied by V_m to obtain the fused feature time series, enabling efficient capture and fusion of intersecting information across modalities.

The fusion data is represented as a two-dimensional matrix A:

$$A^{N_f \times T_e} = \begin{bmatrix} a_{11} & \cdots & a_{1N_f} \\ \vdots & \ddots & \vdots \\ a_{T_e,1} & \cdots & a_{T_e,N_f} \end{bmatrix}. \tag{3}$$

where N_f is the number of brain regions, T_e is the number of time points, and a_{ij} represents the activity level of the i-th brain region in the j-th time point.

3.3 Sparse Transformer for Feature Extraction

The fMRI data ($N_f \times T_f$) and EEG data ($N_e \times T_e$) have high spatial and temporal resolutions, respectively, but their fusion results in high-dimensional sparsity. To address this issue, we propose a sparse feature extraction model.

A causal graph is a mathematical model based on a directed acyclic graph (DAG), represented as $G = \langle N, E \rangle$, where N is the set of brain regions and E is the set of edges. These edges represent the connectivity relationships between two brain regions. Causal graphs focus not merely on probabilistic dependencies

between variables, but on causal mechanisms. Their joint probability distribution is decomposed via the causal structure into:

$$P(X_1, X_2, X_3, \cdots X_n) = \prod_{i=1}^{n} P(X_i | Pa(X_i)), \tag{4}$$

where $P_a(X_i)$ is the set of parent nodes for X_i. The task is to learn the hidden EC $G = \langle N, E \rangle$ behind the high-dimensional sparse data.

Due to the sparsity of the fused data, direct input often leads to less accurate learned EC. Therefore, we design a sparse Transformer module, which uses the sparse multi-head self-attention mechanism to extract important features from the sparse data and integrates them into a global representation. This mechanism is more advantageous than the conventional single-head self-attention in learning long-term dependencies. The sparse Transformer module mainly relies on the sparse multi-head self-attention mechanism, which is defined as follows:

$$Atten_i(q_i, K, V) = \sum_j \frac{k(q_i, k_j)}{\sum_l k(q_i, k_l)} v_j = E_{p(k_j|q_i)}[v_j]. \tag{5}$$

The self-attention combines the value matrix V by computing the conditional probability $p(k_j|q_i)$. Sparsity is introduced by limiting the attention probability distribution, ensuring that only a portion of the keys interact with the current query matrix. The sparsity measure $M(q_i, K)$ is given by:

$$M(q_i, K) = \ln \sum_{j=1}^{L} exp(q_i k_j^T / \sqrt{d_k}) - \frac{1}{L} \sum_{j=1}^{L} (\frac{q_i k_j^T}{\sqrt{d_k}}). \tag{6}$$

Due to the high complexity of the dot product operation and potential instability caused by the log-sum-exp (LSE) function, we use the maximum average method to reduce complexity, yielding:

$$\widetilde{M} = max_j \left\{ \frac{q_i k_j^T}{\sqrt{d_k}} \right\} - \frac{1}{L} \sum_{j=1}^{L} (\frac{q_i k_j^T}{\sqrt{d_k}}). \tag{7}$$

Finally, the sparse multi-head self-attention is optimized by traversing each key and adopting a Top-u query, as shown in:

$$Atten^{sparse}(\overline{Q}, K, V) = softmax \left(\frac{\overline{Q} K^T}{\sqrt{d_k}} V \right). \tag{8}$$

This approach significantly improves the efficiency of the Transformer in handling long sequences, reducing both computational and storage overheads while maintaining the expressive power of the attention mechanism. It is especially suitable for tasks that involve large-scale data and high-dimensional features.

3.4 GFN-based EC State Generation

After extracting sparse features, MSGFlowNet initiates the dynamic construction process of EC through GFN. The goal is to model the posterior of the explanatory graph $Q(G|D)$ given the data and jointly learn the distribution over parameters θ that parameterize the latent drift function $f(x)$. The generative process of the GDN proceeds as follows::

$$p(G, \theta, D) = p(G)p(\theta|G)p(D|G, \theta). \tag{9}$$

The terminal state s_f indicates the termination of the sequential construction. Some states $s \in Y$ are connected to s_f, called complete states. Each complete state has a reward $R(s) \geq 0$, and for incomplete states $s \in S \setminus Y$, $R(s) = 0$.

The goal is to find a flow that satisfies the flow-matching condition for all states s':

$$\sum_{s \in Pa(s^{'})} F_\theta(s \to s^{'}) - \sum_{s^{''} \in Ch(s^{'})} F_\theta(s^{'} \to s^{''}) = R(s^{'}). \tag{10}$$

When this condition is satisfied, MSGFlowNet samples from the complete states $s \in Y$ with probability proportional to $R(s)$. To fit the parameters θ of the forward transition probabilities, we minimize the detailed-balance loss:

$$Loss(\theta) = \mathbb{E}_\pi \left[\left[\log \frac{R(G^{'})P_B(G|G^{'})P_\theta(s_f|G)}{R(G)P_\theta(G^{'}|G)P_\theta(s_f|G^{'})} \right]^2 \right]. \tag{11}$$

The forward transition probabilities are parameterized by neural network. We model these transitions with two parts: one for the termination probability $P_\theta(s_f|G)$, and another for the non-termination transition $P_\theta(G'|G, \neg s_f)$:

$$P_\theta(G'|G) = (1 - P_\theta(s_f|G))P_\theta(G'|G, \neg s_f). \tag{12}$$

The transition $G \to G'$ involves adding edges to the graph, and a mask Mask_0 is used to filter out invalid actions. The reward for graph G is given by:

$$R(G) = P(G)P(D|G). \tag{13}$$

Finally, after sampling transitions randomly and updating the parameters, we obtain the probability distribution over the graph G as the output of MSGFlowNet after several iterations.

4 Experiments

4.1 Simulated Dataset Analysis

We generated three simulated datasets to evaluate MSFlowNet and other EC learning methods. For simulated fMRI data, we followed prior work [16, 32] and used the forward dynamic causal model (DCM) to generate time series data with 5 brain regions (200 time points). For simulated EEG data, we used a

Table 1. Performance comparison of different methods on simulation datasets

Methods	sim1				sim2				sim3			
	Prec↑	Rec↑	F1↑	SHD↓	Prec↑	Rec↑	F1↑	SHD↓	Prec↑	Rec↑	F1↑	SHD↓
pwLiNGAM [11]	0.50	0.81	0.59	2.00	0.30	0.29	0.289	5.62	0.36	0.56	0.42	5.22
Two-Step [24]	0.55	0.53	0.53	3.60	0.51	0.51	0.51	**1.80**	0.34	0.41	0.37	4.90
GFN [14]	**0.73**	0.74	0.70	1.00	0.78	0.54	0.63	2.90	0.77	0.51	0.61	2.55
DiffAN [23]	0.59	0.59	0.59	1.20	0.28	0.24	0.27	5.20	0.32	0.39	0.35	5.50
MetaRLEC [32]	0.33	0.25	0.28	3.00	0.32	0.28	0.31	5.25	0.26	0.29	0.27	5.83
MCAN [16]	0.33	0.54	0.39	3.00	**1.00**	0.25	0.40	5.00	**1.00**	0.35	0.51	3.00
SGFlowNet	0.76	0.74	0.73	**0.90**	0.78	0.54	0.63	2.90	0.73	**0.64**	0.63	3.50
MSGFlowNet	0.66	**1.00**	**0.80**	1.00	0.98	0.54	**0.70**	2.01	0.82	0.52	**0.64**	**2.26**

state-space model [28] to generate synchronized time series data with 3 channels (20,000 time points). Each dataset contains data from 10 subjects. To assess performance under sparse conditions, we set the network edges to 3, 5, and 7, and uniformly assigned edge weights of 0.4 to test the methods' ability to detect weak connectivity. We employ four metrics: F1 score (F1), precision (Pre), recall (Rec), and structural Hamming distance (SHD) [1, 10] to evaluate the performance of EC learning.

As shown in Fig.2, MSGFlowNet outperforms other EC learning methods, including SGFlowNet (the MSGFlowNet using single-modal fMRI), across various datasets. In particular, MSGFlowNet performs well on the sim1 and sim2 datasets. As presented in Table1, this underscores MSGFlowNet's ability to effectively capture the intricate relationships in sparse and complex data, demonstrating its robustness in learning EC under challenging conditions.

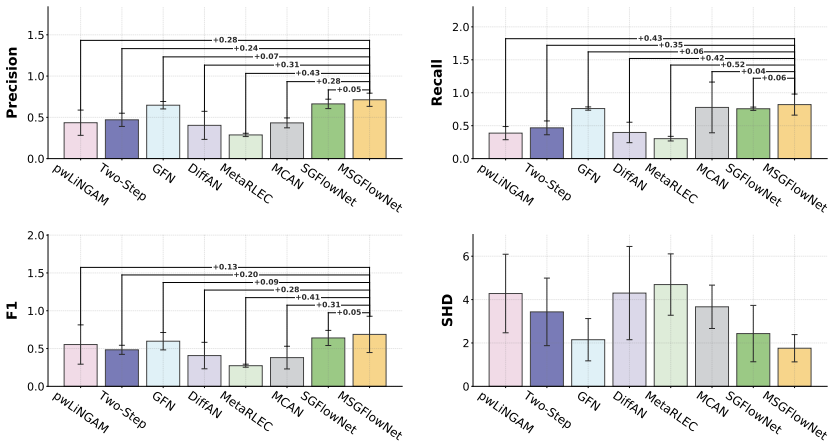

Fig. 2. Average performance comparison of various methods on 3 simulated datasets. The height of each bar represents the model's performance, and the performance differences relative to MSGFlowNet are indicated by connecting lines.

Additionally, MSGFlowNet outperforms MCAN (also employs fMRI and EEG data) across all core evaluation metrics. These results validate MSGFlowNet's

effective integration of fMRI and EEG data, successfully addressing high-dimensional sparsity and outperforming other multimodal methods.

4.2 Real Dataset Analysis (downstream classification task)

To further validate the effectiveness of MSDlowNet, we utilize simultaneous from the OpenNeuro dataset [9] for downstream classification tasks, and compare its performance against state-of-the-art methods.

The raw dataset is preprocessed using statistical parametric mapping (for fMRI) and EEGLAB (for EEG) toolboxes, yielding 65 resting-state and 183 sleep-state simultaneous fMRI-EEG recordings. We employ 15 functionally critical brain regions as the ROIs based on the research of Enzo Tagliazucchi et al. [27] and retain 30 EEG channels after excluding electrooculography (EOG) and electrocardiography (ECG) artifacts.

We evaluated EC learning performance by classifying rest-state vs. sleep-state brain activity using ground truth labels. All methods were tested with the same polynomial kernel SVM classifier, and assessed using four metrics: accuracy, F1-score, precision, and recall.

Fig. 3 shows MSGFlowNet obtains better classification results by effectively combining fMRI's spatial resolution and EEG's temporal resolution to capture EC characteristics. Our method outperforms traditional approaches [27] using fewer regions and less data.

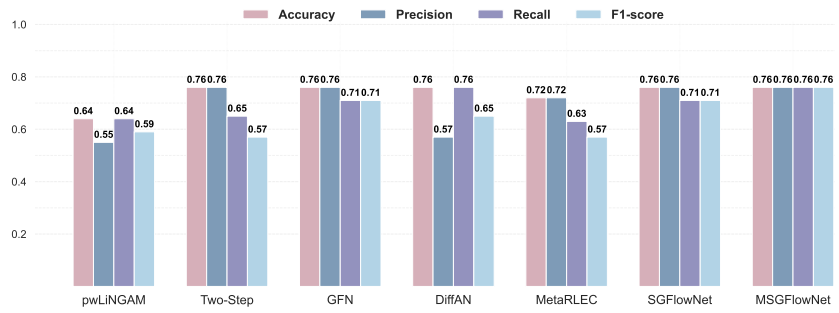

Fig. 3. Performance comparison of different methods on classification tasks

In addition to classification metrics, we also examine the neurobiological interpretability of the learned EC networks.EC analyses revealed increased directed links from the right supramarginal gyrus to the left precuneus during rest compared to sleep. These regions are associated with attention, self-awareness, and sensory integration.The enhanced connectivity during rest reflects heightened cognitive readiness and environmental monitoring, which diminish during sleep as consciousness declines.

5 Conclusion and Limitation

This study introduces MSGFlowNet, for learning EC from fMRI and EEG data. The model combines an attention-guided encoder with a multi-head attention-guided sparse Transformer to efficiently fuse multimodal information, addressing the high-dimensional sparsity issue after data fusion. The experimental results indicate that MSGFlowNet performs well on multiple brain analysis tasks. A limitation of the current work is the computational efficiency for large-scale brain network analysis. Future work will address this by integrating advanced neuroimaging modalities to enable deeper insights into brain information processing mechanisms.

Acknowledgements. This work was sponsored by Beijing Nova Program (20240484635), supported by National Natural Science Foundation of China (62106009, 62276010, 62171297).

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

References

1. Cai, R., Qiao, J., Zhang, Z., Hao, Z.: Self: structural equational likelihood framework for causal discovery. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
2. Cao, J., Yang, L., Sarrigiannis, P.G., Blackburn, D., Zhao, Y.: Dementia classification using a graph neural network on imaging of effective brain connectivity. Computers in Biology and Medicine **168**, 107701 (2024)
3. Chen, D., Liu, M., Wang, S., Li, Z., Bai, L., Wang, Q., Shen, D., Zhang, L.: Guiding fusion of dynamic functional and effective connectivity in spatio-temporal graph neural network for brain disorder classification. Knowledge-Based Systems **309**, 112856 (2025)
4. Chen, D., Zhang, L.: Fe-stgnn: Spatio-temporal graph neural network with functional and effective connectivity fusion for mci diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 67–76. Springer (2023)
5. Chuang, K.C., Ramakrishnapillai, S., Bazzano, L., Carmichael, O.: Nonlinear conditional time-varying granger causality of task fmri via deep stacking networks and adaptive convolutional kernels. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 271–281. Springer (2022)
6. Deleu, T., Góis, A., Emezue, C., Rankawat, M., Lacoste-Julien, S., Bauer, S., Bengio, Y.: Bayesian structure learning with generative flow networks. In: Uncertainty in Artificial Intelligence. pp. 518–528. PMLR (2022)
7. Friston, K.J.: Functional and effective connectivity: a review. Brain connectivity **1**(1), 13–36 (2011)
8. Friston, K., Harrison, L., Penny, W.: Dynamic causal modelling. Neuroimage **19**(4), 1273–1302 (2003)

9. Gu, Y., Han, F., Sainburg, L.E., Schade, M.M., Buxton, O.M., Duyn, J.H., Liu, X.: An orderly sequence of autonomic and neural events at transient arousal changes. NeuroImage **264**, 119720 (2022)
10. Huang, B., Zhang, K., Lin, Y., Schölkopf, B., Glymour, C.: Generalized score functions for causal discovery. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 1551–1560 (2018)
11. Hyvarinen, A.: Pairwise measures of causal direction in linear non-gaussian acyclic models. In: Proceedings of 2nd Asian Conference on Machine Learning. pp. 1–16. JMLR Workshop and Conference Proceedings (2010)
12. Ji, J., Fan, J., Liu, J.: Msad-lec: Estimating large-scale brain effective connectivity network based on multi-subgraph attention diffusion. Knowledge-Based Systems **309**, 112858 (2025)
13. Li, F., Wang, X., Lin, Q., Hu, Z.: Unified model selection approach based on minimum description length principle in granger causality analysis. IEEE Access **8**, 68400–68416 (2020)
14. Li, W., Li, Y., Zhu, S., Shao, Y., Hao, J., Pang, Y.: Gflowcausal: Generative flow networks for causal discovery (2023), https://arxiv.org/abs/2210.08185
15. Liang, G., Tiwari, P., Nowaczyk, S., Byttner, S., Alonso-Fernandez, F.: Dynamic causal explanation based diffusion-variational graph neural network for spatiotemporal forecasting. IEEE Transactions on Neural Networks and Learning Systems (2024)
16. Liu, J., Han, L., Ji, J.: Mcan: Multimodal causal adversarial networks for dynamic effective connectivity learning from fmri and eeg data. IEEE Transactions on Medical Imaging **43**(8), 2913–2923 (2024)
17. Liu, J., Ji, J., Xun, G., Zhang, A.: Inferring effective connectivity networks from fmri time series with a temporal entropy-score. IEEE transactions on neural networks and learning systems **33**(10), 5993–6006 (2021)
18. Mclntosh, A.: Structural equation modeling and its application to network analysis in functional brain imaging. Human brain mapping **2**(1-2), 2–22 (1994)
19. Muthuraman, M.: Effective connectivity of cortical sensorimotor networks during finger movement tasks: a simultaneous fnirs, fmri, eeg study. Brain topography **29**, 645–660 (2016)
20. Noorzadeh, S., Maurel, P., Oberlin, T., Gribonval, R., Barillot, C.: Multi-modal eeg and fmri source estimation using sparse constraints. In: Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20. pp. 442–450. Springer (2017)
21. Rossini, P.M., Di Iorio, R., Bentivoglio, M., Bertini, G., Ferreri, F., Gerloff, C., Ilmoniemi, R.J., Miraglia, F., Nitsche, M.A., Pestilli, F., et al.: Methods for analysis of brain connectivity: An ifcn-sponsored review. Clinical Neurophysiology **130**(10), 1833–1858 (2019)
22. Sakkalis, V.: Review of advanced techniques for the estimation of brain connectivity measured with eeg/meg. Computers in biology and medicine **41**(12), 1110–1117 (2011)
23. Sanchez, P., Liu, X., O'Neil, A.Q., Tsaftaris, S.A.: Diffusion models for causal discovery via topological ordering. arXiv preprint arXiv:2210.06201 (2022)
24. Sanchez-Romero, R., Ramsey, J.D., Zhang, K., Glymour, M.R., Huang, B., Glymour, C.: Estimating feedforward and feedback effective connections from fmri time series: Assessments of statistical methods. Network Neuroscience **3**(2), 274–306 (2019)

25. Smith, S.M.: The future of fmri connectivity. Neuroimage **62**(2), 1257–1266 (2012)
26. Song, X., Zhou, F., Frangi, A.F., Cao, J., Xiao, X., Lei, Y., Wang, T., Lei, B.: Multicenter and multichannel pooling gcn for early ad diagnosis based on dual-modality fused brain network. IEEE Transactions on Medical Imaging **42**(2), 354–367 (2022)
27. Tagliazucchi, E., Laufs, H.: Decoding wakefulness levels from typical fmri resting-state data reveals reliable drifts between wakefulness and sleep. Neuron **82**(3), 695–708 (2014)
28. Tu, T., Paisley, J., Haufe, S., Sajda, P.: A state-space model for inferring effective connectivity of latent neural dynamics from simultaneous eeg/fmri. Advances in Neural Information Processing Systems **32** (2019)
29. Wang, Y., Lin, K., Qi, Y., Lian, Q., Feng, S., Wu, Z., Pan, G.: Estimating brain connectivity with varying-length time lags using a recurrent neural network. IEEE Transactions on Biomedical Engineering **65**(9), 1953–1963 (2018)
30. Xiong, W., Liu, J., Ji, J., Ma, F.: Brain effective connectivity estimation via fourier spatiotemporal attention. arXiv preprint arXiv:2503.11283 (2025)
31. Zhang, X., Shi, E., Yu, S., Zhang, S.: Dtca: Dual-branch transformer with cross-attention for eeg and eye movement data fusion. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 141–151. Springer (2024)
32. Zhang, Z., Ji, J., Liu, J.: Metarlec: Meta-reinforcement learning for discovery of brain effective connectivity. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 10261–10269 (2024)