# STCBN-EC:A Spatio-Temporal Constrained Bayesian Causal Network for Multimodal Brain Effective Connectivity Learning

**Zhihao Su**[1] , **Junzhong Ji**[1] , **Minqi Yu**[1] , **Jinduo Liu**[1*]

[1]Beijing University of Technology

suzhihao@emails.bjut.edu.cn, jjz01@bjut.edu.cn
yuminqi@emails.bjut.edu.cn, jinduo@bjut.edu.cn

## Abstract

Brain effective connectivity (EC) characterizes directional causal interactions among brain regions. However, learning stable and directionally explicit EC networks from multimodal data remains challenging. In practice, functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) differ substantially in spatio-temporal resolution and noise characteristics. Existing methods often rely on manual spatio-temporal alignment and tend to recover only partial causal structures under high noise and Bayesian equivalence classes. To address these challenges, we propose a spatio-temporal constrained Bayesian causal network for multimodal brain effective connectivity learning (STCBN-EC). First, STCBN-EC constructs an anatomically guided EEG–fMRI spatial mapping and derives a unified spatio-temporal representation through slice-level alignment and adaptive modality fusion. Then, a Bayesian causal network is employed to model nonlinear inter-regional dependencies, where uncertainty-driven surrogate scoring is used to evaluate candidate structures. Finally, multimodal representation learning and EC structure estimation are jointly optimized via a gradient-free global optimization strategy. Experiments on simulated and real EEG–fMRI datasets demonstrate that STCBN-EC outperforms state-of-the-art methods and effectively captures state-dependent directional interactions among brain regions.

## 1 Introduction

Brain effective connectivity (EC) characterizes causal and directional interactions among brain regions and is fundamental for understanding cognition, behavior, and the neural mechanisms of neurological disorders [Mamoon *et al.*, 2025; Yu *et al.*, 2025]. Traditional statistical approaches, including structural equation modeling (SEM), dynamic causal modeling (DCM), and Granger causality analysis (GCA), have been widely used for EC learning. However, these methods rely on strong assumptions such as linearity and stationarity [Friston, 2011], which limit their ability to capture the nonlinear and dynamic properties commonly observed in real brain networks [Marinazzo *et al.*, 2011; Stramaglia *et al.*, 2014].

With the advancement of machine learning [Tank *et al.*, 2021; Zhao *et al.*, 2022], data-driven approaches have substantially expanded the scope of EC analysis [Chen and Zhang, 2023; Dai *et al.*, 2025]. Graph neural networks (GNNs) leverage brain network topology to model causal dependencies, recurrent neural networks (RNNs) capture temporal dynamics and nonlinear lag effects, and generative models such as GANs and diffusion models have been explored for data augmentation and causal structure discovery [Liang *et al.*, 2024; Ji *et al.*, 2025; Chen *et al.*, 2025]. Despite these advances, most existing studies remain focused on unimodal data, typically functional magnetic resonance imaging (fMRI), and fail to fully exploit the complementary information provided by multimodal neuroimaging.

Functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) provide highly complementary representations of brain activity [Ciccarelli *et al.*, 2023; Warbrick, 2022], forming a critical basis for accurate EC modeling [Wei *et al.*, 2025]. fMRI offers high spatial resolution for characterizing large-scale inter-regional interactions, but its temporal resolution is limited by hemodynamic effects. In contrast, EEG directly captures neural activity with millisecond-level temporal precision [Herwig *et al.*, 2003], albeit with relatively limited spatial localization accuracy [Zhang *et al.*, 2024a; Sakkalis, 2011]. Coordinating these modalities within a unified modeling framework that explicitly accounts for their spatio-temporal differences enables the simultaneous preservation of fine-grained spatial organization and fast neural dynamics [Xiong *et al.*, 2025a].

However, existing multimodal EC methods still struggle to fully exploit this complementarity at the modeling level [Li *et al.*, 2024; Su *et al.*, 2025]. Some approaches directly apply Granger causality to multimodal data without enforcing consistent spatial or temporal representations across modalities [Muthuraman, 2016]. Other methods employ linear state-space models [Yun *et al.*, 2025] to integrate temporal information [Valdes-Sosa *et al.*, 2009; Abdullah and Rahim, 2026], but their linear formulations restrict the modeling of complex nonlinear causal dependencies [Tu *et al.*, 2019]. In addition, several cross-modal mapping strategies rely primar-

*Corresponding Author

ily on feature concatenation or linear transformations, neglecting anatomical correspondence between EEG electrodes and brain regions as well as the intrinsic temporal structure of neural activity [Liu *et al.*, 2024]. These designs often decouple multimodal representation learning from causal structure inference, reducing interpretability and robustness, and highlighting the need for principled multimodal EC frameworks that explicitly enforce spatio-temporal consistency and model nonlinear cross-modal causal interactions.

To address these limitations, we propose STCBN-EC, a spatio-temporal constrained Bayesian causal network framework for multimodal effective connectivity learning from EEG–fMRI data. STCBN-EC consists of three key components. First, the Spatio-Temporal Fusion Module constructs a unified spatio-temporal representation by explicitly accounting for anatomical correspondence and temporal alignment across modalities. Second, the Bayesian EC Learning Module enables uncertainty-aware exploration of candidate causal graphs in the discrete and non-differentiable structure space. Third, the Structure-Guided Joint Optimization Module couples multimodal representation with EC learning through a structure-guided optimization strategy. By integrating the Spatio-Temporal Fusion Module, the Bayesian Neural Surrogate for EC Search, and the Structure-Guided Joint Optimization Module into a unified framework, STCBN-EC enables stable and directionally consistent EC learning under heterogeneous and noisy multimodal observations.

The main contributions of this work are summarized as follows:

- We introduce anatomical geometric constraints and slice-level temporal alignment to construct a unified spatio-temporal representation for EEG–fMRI EC learning under heterogeneous observations.
- We design learnable ROI-wise fusion weights $\alpha$ to adaptively integrate EEG and fMRI representations, improving robustness under multimodal noise and temporal mismatch.
- We employ a Bayesian neural surrogate for EC structure scoring to model nonlinear causal dependencies with uncertainty awareness, enhancing the stability of causal direction identification.

## 2 Preliminary and Related Work

### 2.1 Brain Effective Connectivity

EC describes directed causal influences among brain regions and reflects the directionality of neural information flow [Cai *et al.*, 2018; Huang *et al.*, 2018; Liu *et al.*, 2022]. EC is commonly modeled as a directed graph $G = (V, E, W)$, where $V$ denotes regions of interest (ROIs), $E$ represents directed causal edges, and $W$ is an asymmetric adjacency matrix encoding causal strengths. Given multivariate neural time series $\mathbf{X} = \{\mathbf{x}(t)\}_{t=1}^{T}$, EC learning aims to recover the latent causal structure $G$, which can be viewed as a causal structure learning problem over directed acyclic graphs.

### 2.2 Multimodal Effective Connectivity

Multimodal effective connectivity extends EC by integrating complementary information from multiple neuroimaging modalities, such as EEG and fMRI, to infer a shared causal brain network [Daunizeau *et al.*, 2011]. Let $\mathbf{X}^{(m)} = \{\mathbf{x}^{(m)}(t)\}_{t=1}^{T_m}$, $m \in \{\mathrm{EEG}, \mathrm{fMRI}\}$, denote modality-specific observations with distinct temporal resolution, noise characteristics $\varepsilon^{(m)}$, and observation operators $\mathcal{O}^{(m)}(\cdot)$. Multimodal EC assumes that these heterogeneous observations are generated from a common latent neural process governed by a shared directed causal structure $G$, and aims to infer a cross-modality consistent connectivity pattern from multi-source data [Liu *et al.*, 2024; Li *et al.*, 2025].

## 3 Motivations

Achieving reliable and stable inference of directed EC from multimodal data faces core challenges. The spatial mismatch between EEG electrodes and fMRI regions complicates anatomically-grounded representation. The temporal resolution gap, compounded by hemodynamic response, obscures causal timing. Modality-specific noise also renders fixed fusion strategies ineffective, often producing unstable connectivity estimates. Critically, many methods decouple representation learning from causal search, leading to direction identification that is sensitive to noise. Our framework addresses this by explicitly modeling anatomical constraints, enabling adaptive temporal alignment, and jointly optimizing fusion with discovery for robust directional EC learning.

## 4 Method

We present STCBN-EC, a spatio-temporal constrained Bayesian causal network for learning multimodal EC from EEG-fMRI data. The framework incorporates three modules: a spatio-temporal fusion module, a Bayesian EC learning module, and a structure-guided joint optimization module. Together, these modules enable stable and directionally consistent EC learning. By combining the spatial information from fMRI and the temporal data from EEG, STCBN-EC overcomes the challenges of their differences in resolution and noise characteristics, providing a more reliable solution for EC learning.

### 4.1 Spatio-Temporal Fusion Module

To enable stable and identifiable EC learning from multimodal data, we begin by constructing a spatio-temporal constrained representation that explicitly aligns EEG and fMRI at the ROI level. Let $\mathbf{X}_{\mathrm{EEG}}(t) \in \mathbb{R}^{E}$ denote EEG observations from $E$ electrodes, and let $\{\mathbf{p}_e\}$ and $\{\mathbf{q}_r\}$ denote the 3D coordinates of EEG electrodes and fMRI ROIs, respectively. The EEG-to-ROI mapping is defined as

$$W_{re} = \frac{\exp\left(-\|\mathbf{q}_r - \mathbf{p}_e\|_2^2 / 2\sigma^2\right)}{\sum_{e'} \exp(-\|\mathbf{q}_r - \mathbf{p}_{e'}\|_2^2 / 2\sigma^2)}, \tag{1}$$

where $\sigma$ controls spatial decay and $W$ is row-normalized. This yields ROI-level EEG representations $\tilde{\mathbf{H}}_{\mathrm{EEG}}(t) = W\mathbf{X}_{\mathrm{EEG}}(t)$.

To reconcile temporal resolution differences, EEG signals are aggregated within each fMRI repetition time (TR), while fMRI ROI time series are convolved with a canonical hemodynamic response function (HRF) to account for temporal

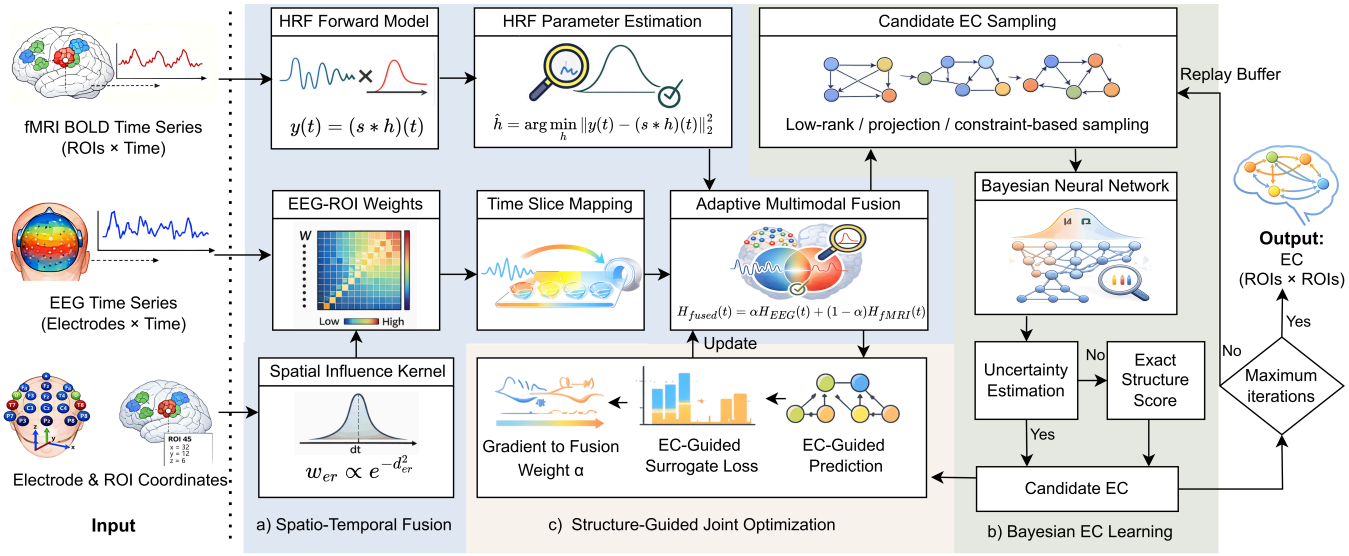


Figure 1: Overview of the proposed STCBN-EC framework. The method consists of three modules: (a) the Spatio-Temporal Fusion Module, which constructs a unified spatio-temporal representation from EEG and fMRI; (b) the Bayesian EC Learning Module, which performs uncertainty-aware exploration of candidate causal graphs; (c) the Structure-Guided Joint Optimization Module, which couples multimodal representation learning with EC learning.

smoothing effects. Based on the temporally aligned EEG and fMRI representations, we construct a unified multimodal representation via adaptive ROI-wise fusion:

$$\mathbf{H}(k) = \boldsymbol{\alpha} \odot \mathbf{H}_{\text{fMRI}}(k) + (1 - \boldsymbol{\alpha}) \odot \mathbf{H}_{\text{EEG}}(k), \quad (2)$$

where $\boldsymbol{\alpha} \in (0,1)^R$ denotes learnable fusion weights parameterized by a sigmoid function, which adaptively balance the contributions of EEG and fMRI across brain regions under varying noise conditions.

The resulting fused sequence $\mathbf{H} = \mathbf{H}(k)_{k=1}^{T}$ offers a unified spatio-temporal representation that maintains anatomical structure, temporal consistency, and multimodal complementarity, providing a stable foundation for EC learning.

## 4.2 Bayesian EC Learning Module

Given the fused multimodal representation $\mathbf{H}$, we proceed to learn EC by modeling directed interactions among ROIs within a structural causal framework.

We model the activity of each Region of Interest (ROI) using a Structural Causal Model (SCM). The causal relationships among ROIs are represented as an EC graph $G = (V, E)$, where $V = \{1, \ldots, R\}$ represents the set of $R$ ROIs. In this EC, an edge $(i \to j) \in E$ indicates that ROI $i$ has a direct causal influence on ROI $j$. The parent set of node $j$ in this graph is denoted by $\text{pa}_G(j) = \{i \in V \mid (i \to j) \in E\}$, representing all the ROIs that causally influence ROI $j$.

The data-generating process for the activities of ROI $j$ at time $t$ is modeled as follows:

$$h_j^{(t)} = f_j\big(\mathbf{h}_{\text{pa}_G(j)}^{(t-1)}, \epsilon_j^{(t)}\big), \quad \epsilon_j^{(t)} \sim P(\epsilon), \quad (3)$$

where $f_j(\cdot)$ represents the causal mechanism governing the evolution of ROI $j$ over time, and $\epsilon_j^{(t)}$ is an exogenous noise term.

**Score-based causal discovery.** We adopt a score-based formulation and seek

$$G^* = \arg\max_{G \in \mathcal{G}_{\text{DAG}}} \mathcal{S}(\mathcal{D}, G), \quad (4)$$

where $\mathcal{G}_{\text{DAG}}$ is the space of DAGs over $R$ nodes and $\mathcal{S}$ is a consistent score. In practice, we instantiate $\mathcal{S}$ using a BIC-style score that decomposes over nodes. Let $\widehat{f}_j$ be the fitted predictor/regressor for node $j$ given $\text{pa}_G(j)$, and define the node-wise mean squared error

$$\text{MSE}_j\big(\text{pa}_G(j)\big) = \frac{1}{T-1} \sum_{t=2}^{T} \Big(h_j^{(t)} - \widehat{f}_j\big(\mathbf{h}_{\text{pa}_G(j)}^{(t-1)}\big)\Big)^2. \quad (5)$$

Then a common non-equal-variance BIC form is

$$\begin{aligned} \mathcal{S}_{\text{BIC}}(\mathcal{D}, G) = &-(T-1) \sum_{j=1}^{R} \ln \text{MSE}_j\big(\text{pa}_G(j)\big) \\ &- |G| \ln(T-1), \end{aligned} \quad (6)$$

where $|G|$ is the number of edges in $G$. This score rewards better predictive fit while penalizing dense graphs.

**Parametrized DAG generation with low-rank potentials.** Directly searching $\mathcal{G}_{\text{DAG}}$ is combinatorial and non-differentiable. To enable efficient global exploration, we map an unconstrained continuous representation to a DAG through a *DAG generator* $\tau(\cdot)$. Following recent constraint-free DAG parameterizations, we associate each node with (i) a scalar *node potential* vector $\mathbf{p} \in \mathbb{R}^R$ that induces an ordering, and (ii) a low-rank embedding matrix $\mathbf{R} \in \mathbb{R}^{R \times k}$ with $k \ll R$ that controls connectivity. Define the gradient-flow operator

$$\text{grad}(\mathbf{p})_{ij} = p_j - p_i, \quad (7)$$

and let $\mathbb{H}(\cdot)$ be the entry-wise Heaviside step function. We construct a binary adjacency matrix by

$$\mathbf{A} = \tau(\mathbf{p}, \mathbf{R}) = \mathbb{H}(\mathrm{grad}(\mathbf{p})) \odot \mathbb{H}\big(\mathbf{R}\mathbf{R}^\top\big), \tag{8}$$

where $\odot$ denotes the Hadamard product. The first term enforces acyclicity by permitting edges only from lower to higher potential; the second term provides a low-rank connectivity prior. We denote by $\mathbf{z} \in \mathbb{R}^{R(1+k)}$ the concatenation of $\mathbf{p}$ and the flattened $\mathbf{R}$, and write $\tau(\mathbf{z})$ for brevity. The structure learning objective (4) becomes

$$\mathbf{z}^* = \arg\max_{\mathbf{z}\in\mathbb{R}^{R(1+k)}} \mathcal{S}_{\mathrm{BIC}}\big(\mathcal{D}, \tau(\mathbf{z})\big), \qquad G^* = \tau(\mathbf{z}^*). \tag{9}$$

**Bayesian optimization for discrete-structure search.** The score $\mathcal{S}_{\mathrm{BIC}}(\mathcal{D}, \tau(\mathbf{z}))$ is expensive to evaluate (it requires fitting node-wise predictors for each candidate DAG) and is non-smooth due to $\mathbb{H}(\cdot)$ in (8). We therefore employ BO, which iteratively: (i) proposes candidates using an acquisition function, (ii) evaluates the true score for selected candidates, and (iii) updates a probabilistic surrogate model.

**Bayesian neural surrogate as the BO surrogate model.** Instead of Gaussian processes, we adopt a Bayesian neural network (BNN) surrogate to improve scalability with the growing number of evaluations. The surrogate models the posterior over the unknown score function

$$f(\mathbf{z}) := \mathcal{S}_{\mathrm{BIC}}\big(\mathcal{D}, \tau(\mathbf{z})\big), \tag{10}$$

using a parameter distribution $p(\boldsymbol{\theta} \mid \mathcal{H})$ conditioned on the evaluation history

$$\mathcal{H} = \big\{(\mathbf{z}^{(i)},\, y^{(i)})\big\}_{i=1}^{n}, \quad y^{(i)} = f(\mathbf{z}^{(i)}). \tag{11}$$

With approximate Bayesian inference (e.g., Monte-Carlo dropout), for any candidate $\mathbf{z}$ we draw $M$ stochastic forward passes

$$\big\{\hat{y}^{(m)}(\mathbf{z})\big\}_{m=1}^{M}, \quad \hat{y}^{(m)}(\mathbf{z}) \sim q\big(y \mid \mathbf{z}, \mathcal{H}\big), \tag{12}$$

and compute predictive mean and epistemic uncertainty:

$$\mu(\mathbf{z}) = \frac{1}{M}\sum_{m=1}^{M} \hat{y}^{(m)}(\mathbf{z}), \tag{13}$$

$$\sigma^2(\mathbf{z}) = \frac{1}{M}\sum_{m=1}^{M} \big(\hat{y}^{(m)}(\mathbf{z}) - \mu(\mathbf{z})\big)^2. \tag{14}$$

This uncertainty is critical for robust exploration in the highly non-convex, partially observed structure landscape.

**Uncertainty-aware acquisition.** We employ an exploration exploitation trade-off via an upper confidence bound (UCB) acquisition:

$$\mathrm{AF}(\mathbf{z}) = \mu(\mathbf{z}) + \beta\,\sigma(\mathbf{z}), \qquad \beta > 0, \tag{15}$$

or Thompson sampling as a stochastic alternative:

$$\mathrm{AF}(\mathbf{z}) \sim q\big(y \mid \mathbf{z}, \mathcal{H}\big). \tag{16}$$

At each BO iteration, we sample $C$ preliminary candidates $\{\mathbf{z}^{(j)}\}_{j=1}^{C}$ in a trust region around the current best $\mathbf{z}_{\text{best}}$, compute $\mathrm{AF}(\mathbf{z}^{(j)})$, select the top-$B$ candidates, and evaluate their true scores.

**Indirect surrogate modeling via node-wise local scores.** A key practical enhancement is to exploit the decomposability of (6). Rather than modeling $f(\mathbf{z})$ directly, we model node-wise local quantities and recombine them. Specifically, for each node $j$, define the local target

$$\ell_j(G) := \ln \mathrm{MSE}_j\big(\mathrm{pa}_G(j)\big), \tag{17}$$

which is observable once $G$ is evaluated. We train $R$ independent dropout networks $\{\mathrm{DropoutNN}_j\}_{j=1}^{R}$ to predict $\ell_j(G)$ from a binary parent-indicator vector $\mathbf{u}_j(G) \in \{0,1\}^R$, where $u_{j,i}(G) = \mathbb{I}\{i \in \mathrm{pa}_G(j)\}$. For a candidate $G = \tau(\mathbf{z})$, we sample

$$\hat{\ell}_j^{(m)}(G) \sim \mathrm{DropoutNN}_j\big(\mathbf{u}_j(G)\big), \tag{18}$$

and combine them to obtain a sampled global score:

$$\hat{\mathcal{S}}^{(m)}(\mathcal{D}, G) = -(T-1)\sum_{j=1}^{R} \hat{\ell}_j^{(m)}(G) - |G|\ln(T-1). \tag{19}$$

Repeating (19) for $m = 1, \ldots, M$ yields $\mu(\mathbf{z})$ and $\sigma(\mathbf{z})$ used in (15). This *local-to-global* surrogate improves sample-efficiency and reduces spurious correlations by avoiding a single shared network across nodes.

**Surrogate update.** After evaluating the selected batch, we update the experience set:

$$\mathcal{H} \leftarrow \mathcal{H} \cup \Big\{\big(\mathbf{z}^{(j)},\, \mathcal{S}_{\mathrm{BIC}}(\mathcal{D}, \tau(\mathbf{z}^{(j)}))\big)\Big\}_{j=1}^{B}, \tag{20}$$

and additionally store node-wise local targets $\{\ell_j(\tau(\mathbf{z}^{(j)}))\}_{j=1}^{R}$ to update $\mathrm{DropoutNN}_j$ with replay-based training. This continual refinement yields an increasingly accurate uncertainty-aware surrogate, enabling efficient recovery of a high-scoring DAG, which we take as the estimated EC graph $\widehat{G}$ for Sec. 4.3.

## 4.3 Structure-Guided Joint Optimization

Let $\mathbf{H} = \{\mathbf{H}(k)\}_{k=1}^{T}$ denote the fused multimodal representation obtained via Eq. (2), parameterized by fusion weights $\boldsymbol{\alpha}$. Given a candidate causal graph $G^* \in \{0,1\}^{R\times R}$ inferred by the Bayesian neural surrogate, our goal is to update $\boldsymbol{\alpha}$ such that the resulting representation $\mathbf{H}$ is maximally consistent with the causal dependencies encoded by $G^*$.

Treating $G^*$ as fixed, we impose a linear structural equation model on the fused representation:

$$\mathbf{H}_j(k) = \sum_{i\in \mathrm{Pa}_{G^*}(j)} w_{ij}\,\mathbf{H}_i(k-1) + \varepsilon_j(k), \tag{21}$$

where $\mathrm{Pa}_{G^*}(j)$ denotes the parent set of node $j$ under $G^*$, and $\varepsilon_j(k)$ is a noise term. The regression coefficients $\mathbf{w}_j = \{w_{ij}\}$ are estimated independently for each node.

To ensure numerical stability and differentiability with respect to $\mathbf{H}$, we estimate $\mathbf{w}_j$ via ridge regression:

$$\begin{aligned}\mathbf{w}_j = \arg\min_{\mathbf{w}} &\big\|\mathbf{H}_j(2{:}T) - \mathbf{H}_{\mathrm{Pa}(j)}(1{:}T-1)\mathbf{w}\big\|_2^2 \\ &+ \lambda_{\mathrm{ridge}}\|\mathbf{w}\|_2^2.\end{aligned} \tag{22}$$

whose closed-form solution

$$\mathbf{w}_j = \big(\mathbf{X}_j^\top\mathbf{X}_j + \lambda_{\mathrm{ridge}}\mathbf{I}\big)^{-1}\mathbf{X}_j^\top\mathbf{y}_j \tag{23}$$

is differentiable with respect to $\mathbf{H}$, where $\mathbf{X}_j = \mathbf{H}_{\mathrm{Pa}(j)}(1{:}T-1)$ and $\mathbf{y}_j = \mathbf{H}_j(2{:}T)$.

**EC-guided surrogate loss.** Using the fitted regression model, we define a structure-consistency loss that penalizes violations of the inferred causal dependencies:

$$\mathcal{L}_{\mathrm{dag}}(\mathbf{H}, G^*) = \frac{1}{R}\sum_{j=1}^{R}\left\|\mathbf{H}_j(2{:}T) - \hat{\mathbf{H}}_j(2{:}T)\right\|_2^2, \quad (24)$$

where $\hat{\mathbf{H}}_j$ denotes the ridge-based prediction induced by $G^*$. This loss measures how well the current fused representation admits a dynamical factorization consistent with the EC.

The fusion parameters $\boldsymbol{\alpha}$ are optimized by minimizing the following composite objective:

$$\mathcal{L}_{\mathrm{fusion}} = \mathcal{L}_{\mathrm{align}} + \lambda_{\mathrm{smooth}}\mathcal{L}_{\mathrm{smooth}} + \lambda_{\alpha}\mathcal{L}_{\alpha} + \lambda_{\mathrm{dag}}\mathcal{L}_{\mathrm{dag}}, \quad (25)$$

where $\mathcal{L}_{\mathrm{align}}$ enforces multimodal consistency, $\mathcal{L}_{\mathrm{smooth}}$ regularizes temporal dynamics, and $\mathcal{L}_{\alpha}$ prevents degenerate fusion solutions. Importantly, $\mathcal{L}_{\mathrm{dag}}$ provides causal supervision without requiring gradients through the DAG search process.

**Alternating optimization scheme.** Training proceeds via block-coordinate descent: (i) update fusion parameters $\boldsymbol{\alpha}$ with $G^*$ fixed by minimizing Eq. (25); (ii) update the causal structure by re-running Bayesian neural surrogate–guided search on the updated representation $\mathbf{H}$. This alternating procedure progressively aligns multimodal representations with structurally stable causal graphs, yielding robust and directionally identifiable effective connectivity estimates.

# 5 Experiments

We evaluate STCBN-EC on simulated and real EEG-fMRI datasets to assess its performance. Three simulated datasets validate EC recovery against ground truth, while two real datasets test robustness. As real data lacks ground-truth EC, inferred networks are evaluated via brain-state classification. All methods use the same feature extraction and a 10-fold cross-validated $k$-Nearest Neighbors (KNN) classifier.

## 5.1 Baseline Methods and Evaluation Metrics

We compare STCBN-EC with a range of classical and state-of-the-art EC learning methods, including Patel [Patel *et al.*, 2006], pwLiNGAM [Hyvarinen, 2010], lsGC [DSouza *et al.*, 2017], DiffAN [Sanchez *et al.*, 2023], MetaRLEC [Zhang *et al.*, 2024b], MCAN [Liu *et al.*, 2024], FSTA-EC [Xiong *et al.*, 2025b], and DrBO [Duong *et al.*, 2025].

For all baselines, we follow the parameter settings recommended in the original literature and further tune hyperparameters on a validation subset.

We report standard EC evaluation metrics throughout the experiments, including Accuracy (ACC), F1-score (F1), Structural Hamming Distance (SHD), Precision (Prec), Recall (Rec), and ROC-AUC.

## 5.2 Dataset Details

**Simulated Datasets.** We evaluate STCBN-EC using three simulated datasets with varying network sizes ($R \in \{5, 10, 30\}$). Each dataset includes EEG and fMRI signals, with ground-truth effective connectivity structures for controlled EC recovery tests. The fMRI data is simulated using DCM, and EEG signals are generated via a state-space model.

**Sleep–Rest Dataset.** This dataset[1] includes EEG and fMRI data from 33 healthy participants, total 248 samples [Gu *et al.*, 2022]. EEG was recorded from 32 channels, with fMRI data acquired on a 3T Siemens Prisma scanner (TR = 2.1s, TE = 25 ms, resolution = $3 \times 3 \times 3$ mm$^3$). Time series were extracted using the AAL3 atlas and DPABI preprocessing.

**XP2 Motor Imagery Neurofeedback Dataset.** The XP2 dataset[2] [Lioi *et al.*, 2020] contains simultaneous EEG–fMRI data from 40 dNF and 28 MI participants. EEG was recorded from 64 channels, and fMRI was obtained using an EPI sequence on a 3T scanner (TR = 1s, TE = 23 ms, resolution = $3 \times 3 \times 3$ mm$^3$). Time series were extracted using the AAL atlas and DPABI preprocessing.

## 5.3 Results on Simulated Datasets

### Parameter Sensitivity Analysis

We evaluate the sensitivity of STCBN-EC to its key hyperparameters on the Sim-1 setting, including the fusion learning rate, number of training evaluations, and surrogate hidden size. As shown in Table 1, the model demonstrates stable performance across a broad range of parameter values. For the fusion learning rate, $5\times10^{-3}$ achieves the best overall performance, yielding the highest ACC and F1 as well as the lowest SHD. Increasing the training evaluation budget substantially improves EC recovery compared to the low-budget setting, highlighting the importance of uncertainty-aware structure exploration. In addition, a medium hidden size consistently outperforms both smaller and larger configurations. Overall, these results indicate that STCBN-EC is robust to hyperparameter variations and does not require extensive tuning to achieve strong performance.

### Ablation Study

Table 3 presents ablation results for STCBN-EC. Removing joint optimization (*w/o Joint*) degrades ACC and F1 while increasing SHD, showing that decoupled representation learning harms causal discovery. Excluding EEG (*w/o EEG*) also reduces performance, confirming that multimodal fusion adds complementary temporal information beyond fMRI. Disabling HRF modeling (*w/o HRF*) or shuffling EEG signals (*EEG shuffled*) further impairs results, underscoring the need for physiological alignment and cross-modal consistency. The full model achieves the best performance across

| Parameter | Setting | ACC ↑ | F1 ↑ | SHD ↓ |
|---|---|---|---|---|
| Fusion lr | 1e−3 | 0.93±0.05 | 0.82±0.13 | 1.50±0.97 |
| | 5e−3 | **0.94±0.03** | **0.83±0.10** | **1.40±0.84** |
| | 1e−2 | 0.91±0.06 | 0.76±0.19 | 1.80±1.13 |
| Train evals | 1k | 0.86±0.05 | 0.73±0.12 | 2.80±1.03 |
| | 5k | 0.92±0.04 | 0.82±0.10 | 1.60±0.84 |
| | 10k | **0.94±0.03** | **0.83±0.10** | **1.40±0.84** |
| Hidden size | 32 | 0.89±0.06 | 0.69±0.19 | 2.30±1.15 |
| | 64 | **0.94±0.03** | **0.83±0.10** | **1.40±0.84** |
| | 128 | 0.92±0.06 | 0.79±0.17 | 1.70±1.25 |

Table 1: Parameter sensitivity analysis of STCBN-EC on Sim-1.

[1]https://openneuro.org/datasets/ds003768/versions/1.0.13

[2]https://openneuro.org/datasets/ds002338/versions/2.0.2

| Method | Sim-1 | | | Sim-2 | | | Sim-3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Acc↑ | F1↑ | SHD↓ | Acc↑ | F1↑ | SHD↓ | Acc↑ | F1↑ | SHD↓ |
| Patel (2006) | 0.82±0.05 | 0.45±0.36 | 3.50±1.39 | 0.90±0.02 | 0.30±0.29 | 9.20±2.04 | 0.91±0.02 | 0.25±0.04 | 76.80±19.70 |
| lsGC (2010) | 0.76±0.07 | 0.46±0.22 | 4.80±1.31 | 0.93±0.01 | 0.63±0.10 | 6.40±1.70 | 0.88±0.01 | 0.16±0.04 | 103.30±9.20 |
| pwLiNGAM (2017) | 0.76±0.03 | 0.66±0.03 | 4.90±0.56 | 0.84±0.02 | 0.35±0.31 | 14.00±4.20 | 0.71±0.01 | 0.14±0.01 | 248.4±15.50 |
| DiffAN (2023) | 0.88±0.04 | 0.63±0.16 | 2.80±1.10 | 0.92±0.00 | 0.55±0.05 | 7.20±0.90 | **0.93±0.00** | 0.21±0.03 | 60.60±4.60 |
| MetaRLEC (2024) | 0.71±0.03 | 0.59±0.05 | 5.80±0.63 | 0.81±0.02 | 0.26±0.27 | 16.60±4.80 | 0.92±0.08 | 0.21±0.07 | 66.70±7.00 |
| FSTA-EC (2025) | 0.84±0.07 | 0.53±0.26 | 3.10±1.40 | 0.90±0.01 | 0.38±0.17 | 8.40±2.30 | 0.90±0.01 | 0.28±0.03 | 76.90±10.10 |
| DrBO (2025) | 0.89±0.06 | 0.75±0.13 | 2.20±1.20 | 0.94±0.00 | 0.72±0.06 | 4.70±1.10 | 0.88±0.01 | 0.22±0.02 | 99.60±11.50 |
| **STCBN-EC (Ours)** | **0.94±0.03** | **0.83±0.10** | **1.40±0.84** | **0.96±0.00** | **0.80±0.02** | **3.50±0.52** | **0.93±0.00** | **0.34±0.03** | **57.00±3.16** |

Table 2: Performance comparison on three simulated datasets. ACC and F1 quantify edge-level recovery performance of EC, while SHD measures structural deviation from the ground-truth EC. Best results are highlighted in **bold**, and second-best results are underlined.

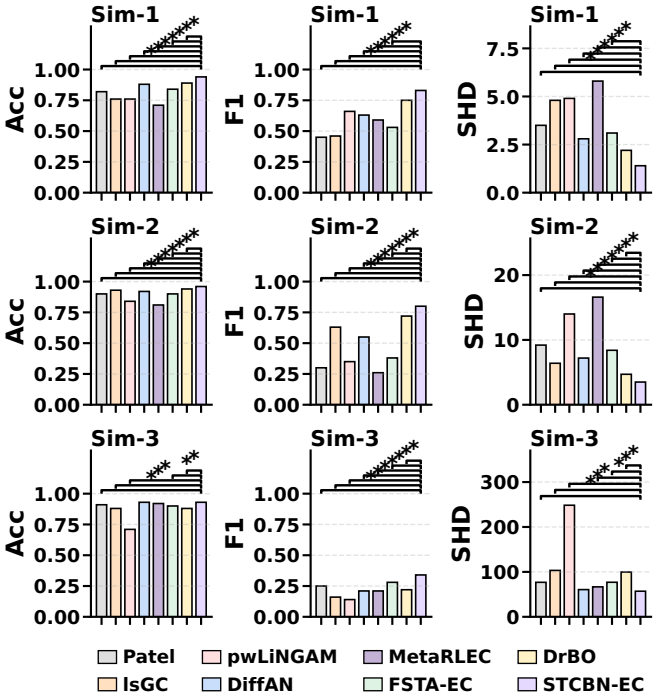


Figure 2: Bars represent the mean performance of each method. Statistical significance is indicated by an asterisk (*).

| Setting | ACC ↑ | F1 ↑ | SHD ↓ |
|---|---|---|---|
| w/o Joint | 0.86 (**+9.3%**) | 0.72 (**+15.3%**) | 2.80 (**-50.0%**) |
| w/o EEG | 0.89 (**+5.6%**) | 0.75 (**+10.7%**) | 2.20 (**-36.4%**) |
| w/o HRF | 0.92 (**+2.2%**) | 0.78 (**+6.4%**) | 1.70 (**-17.6%**) |
| EEG shuffled | 0.92 (**+2.2%**) | 0.78 (**+6.4%**) | 1.70 (**-17.6%**) |
| Full model | **0.94 ± 0.03** | **0.83 ± 0.10** | **1.40 ± 0.84** |

Table 3: Ablation study of STCBN-EC on Sim-1. Relative performance changes (%) are reported in parentheses with respect to each ablated variant.

all metrics, indicating each component's contribution to stable EC learning.

### Comparison with Baseline Methods

We evaluate STCBN-EC on three simulated datasets with increasing network sizes (Sim-1, Sim-2, and Sim-3) to assess its ability to recover ground-truth effective connectivity (EC) structures. Table 2 reports the EC learning results of all compared methods. Overall, STCBN-EC consistently achieves superior performance across all simulation settings. In Sim-1, STCBN-EC attains the highest F1-score and the lowest SHD, indicating accurate edge identification and strong structural consistency. While several baseline methods achieve comparable accuracy, their substantially larger SHD values suggest inferior recovery of the underlying EC.

In Sim-2, STCBN-EC clearly outperforms all competing approaches, achieving the best accuracy, F1-score, and SHD simultaneously. As network complexity increases, classical and recent learning-based baselines exhibit degraded structural recovery, whereas STCBN-EC remains robust, demonstrating good scalability to moderately sized networks.

In the most challenging Sim-3 setting, all methods experience increased SHD due to the difficulty of large-scale EC. Nevertheless, STCBN-EC maintains competitive accuracy and achieves the best F1-score, indicating stable performance under high-dimensional conditions.

## 5.4 Results on Real EEG–fMRI Datasets

We further evaluate STCBN-EC on two real-world EEG–fMRI datasets: XP2 and Sleep–Rest. Quantitative classification results are summarized in Table 4.

### XP2 Dataset

Table 4 shows that STCBN-EC achieves consistently superior classification performance on the XP2 dataset, with marked improvements in ACC, F1, and AUC compared to baseline methods. Notably, the gain is most pronounced in precision and overall F1-score, indicating that the learned EC networks provide highly discriminative and stable features for distinguishing MI and dNF states. This improvement is attributed to the ability of STCBN-EC to recover directional effective connectivity, encoding asymmetric information flow not captured by undirected or weakly directed models.

Beyond quantitative classification performance, the learned EC patterns exhibit clear physiological interpretability. As illustrated in Fig. 3, MI and dNF states are characterized by distinct large-scale interaction profiles at the functional-group level.

In the MI state, strong directed interactions are predominantly observed among parietal and sensorimotor-related systems, particularly involving the parietal group (e.g., Superior

| Method | XP2 (90 ROIs) | | | | | Sleep-Rest (170 ROIs) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Acc↑ | Prec↑ | Rec↑ | F1↑ | AUC↑ | Acc↑ | Prec↑ | Rec↑ | F1↑ | AUC↑ |
| Patel (2006) | 0.43±0.17 | 0.39±0.15 | 0.70±0.23 | 0.49±0.16 | 0.47±0.20 | – | – | – | – | – |
| lsGC (2010) | 0.61±0.06 | 0.10±0.32 | 0.03±0.11 | 0.05±0.16 | 0.53±0.22 | 0.73±0.04 | 0.30±0.46 | 0.07±0.12 | 0.11±0.18 | 0.54±0.12 |
| pwLiNGAM (2017) | 0.61±0.18 | 0.47±0.41 | 0.32±0.30 | 0.35±0.31 | 0.57±0.24 | 0.72±0.06 | 0.35±0.40 | 0.10±0.11 | 0.15±0.17 | 0.57±0.09 |
| DiffAN (2023) | 0.54±0.15 | 0.30±0.36 | 0.22±0.25 | 0.25±0.28 | 0.54±0.18 | 0.64±0.05 | 0.21±0.16 | 0.17±0.15 | 0.19±0.15 | 0.51±0.07 |
| MetaRLEC (2024) | 0.52±0.18 | 0.37±0.40 | 0.30±0.33 | 0.29±0.28 | 0.50±0.21 | – | – | – | – | – |
| FSTA-EC (2025) | 0.57±0.16 | 0.61±0.14 | **0.79±0.19** | 0.68±0.13 | 0.60±0.20 | – | – | – | – | – |
| DrBO (2025) | 0.53±0.14 | 0.35±0.22 | 0.43±0.26 | 0.38±0.23 | 0.49±0.11 | 0.63±0.12 | 0.36±0.21 | **0.32±0.15** | 0.32±0.15 | 0.53±0.14 |
| **STCBN-EC (Ours)** | **0.82±0.20** | **0.78±0.25** | 0.76±0.29 | **0.77±0.03** | **0.82±0.20** | **0.78±0.01** | **0.77±0.15** | 0.30±0.02 | **0.42±0.02** | **0.76±0.11** |

Table 4: Performance comparison on the XP2 and Sleep-Rest datasets. Best results are highlighted in **bold**, and second-best results are underlined.

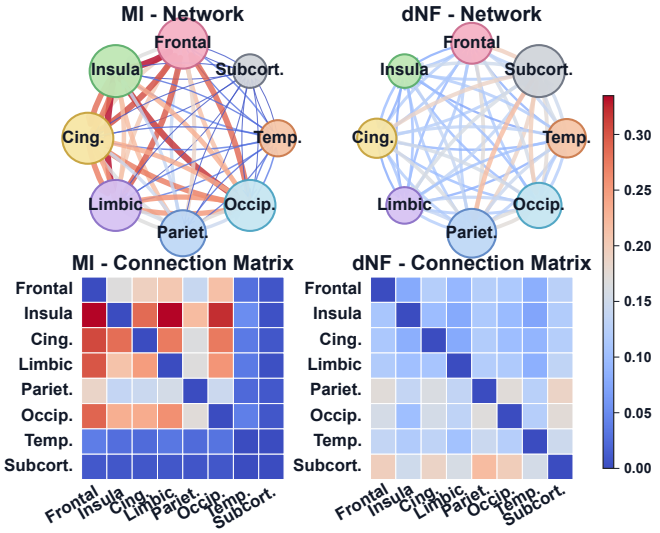


Figure 3: Group EC on the XP2 dataset under MI and dNF states.

Parietal Gyrus and Precuneus) and frontal–parietal pathways. These regions are known to support sensorimotor integration and spatial attention, suggesting that MI relies on coordinated information integration across association cortices.

In contrast, the dNF state shows enhanced EC strength within subcortical and temporal systems, including thalamic, basal ganglia, and auditory-related regions. The prominence of directed interactions involving subcortical groups indicates increased thalamo–cortical coordination, which is associated with global regulation and state monitoring. Meanwhile, stronger temporal interactions may reflect sensory processing during neurofeedback. Compared with MI, the dNF EC exhibits a more distributed interaction pattern, suggesting reduced centralization and a shift toward global modulation.

These findings show that STCBN-EC improves classification accuracy and learns physiologically meaningful directional interaction patterns differentiating cognitive states. The explicit directed EC modeling captures state-dependent reorganization of information flow, explaining its superior performance in both quantitative metrics and neuroscientific interpretability.

### Sleep–Rest Dataset

We compared the learned EC between the sleep and resting-state conditions. Across both states, common hub regions were observed, indicating a stable network backbone. Frontal and temporal cortices consistently appeared as dominant hubs, suggesting their central role in large-scale information integration. Thalamic nuclei also prominently featured in both states, highlighting the importance of thalamocortical interactions in network coordination.

In the sleep condition, the network showed stronger involvement of frontal and orbitofrontal regions, suggesting internally driven regulatory processes. The cerebellum also emerged as a key hub, contributing to large-scale cortical coordination. This pattern reflects the role of the cerebellum in sleep-dependent synchronization and information processing.

In contrast, the resting-state network revealed a stronger focus on posterior regions such as the Precuneus, indicating a self-referential processing mode associated with the default mode network. Additionally, subcortical regions related to neuromodulatory control, including the brainstem and basal ganglia, were more prominent, suggesting that resting-state EC is influenced by global arousal regulation. This indicates that, while sleep emphasizes frontal and cerebellar modulation, the resting state is shaped by posterior integration and neuromodulatory regulation.

These results show that sleep and rest share a common fronto-temporo-thalamic network, while exhibiting state-specific reorganization of hub regions.

## 6 Conclusion and Limitations

We propose STCBN-EC that learns EC from complementary EEG and fMRI data while capturing physiologically meaningful differences across cognitive states. Experimental results on two real-world datasets of different scales demonstrate STCBN-EC's reliability in identifying state-dependent EC patterns, highlighting its practical value for multimodal neuroimaging analysis.

Despite these strengths, the method's computational efficiency remains limited, particularly for large-scale networks and multimodal data. Future work will focus on enhancing efficiency and extending the framework to additional neuroimaging modalities.

## Acknowledgements

This work was sponsored by the Beijing Nova Program (20240484635).

## References

[Abdullah and Rahim, 2026] Dahlan Abdullah and Robbi Rahim. Multimodal integration of eeg and fmri data for understanding visual attention networks. *Advances in Cognitive and Neural Studies*, 2(1):27–33, 2026.

[Cai *et al.*, 2018] Ruichu Cai, Jie Qiao, Zhenjie Zhang, and Zhifeng Hao. Self: structural equational likelihood framework for causal discovery. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[Chen and Zhang, 2023] Dongdong Chen and Lichi Zhang. 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*, pages 67–76. Springer, 2023.

[Chen *et al.*, 2025] Dongdong Chen, Mengjun Liu, Sheng Wang, Zheren Li, Lu Bai, Qian Wang, Dinggang Shen, and Lichi Zhang. Guiding fusion of dynamic functional and effective connectivity in spatio-temporal graph neural network for brain disorder classification. *Knowledge-Based Systems*, 309:112856, 2025.

[Ciccarelli *et al.*, 2023] Giuseppina Ciccarelli, Giovanni Federico, Giulia Mele, Angelica Di Cecca, Miriana Migliaccio, Ciro Rosario Ilardi, Vincenzo Alfano, Marco Salvatore, and Carlo Cavaliere. Simultaneous real-time eeg-fmri neurofeedback: A systematic review. *Frontiers in Human Neuroscience*, 17:1123014, 2023.

[Dai *et al.*, 2025] Peishan Dai, Zhuang He, Yilin Ou, Jialin Luo, Shenghui Liao, and Xiaoping Yi. Estimating brain effective connectivity from time series using recurrent neural networks. *Physical and Engineering Sciences in Medicine*, 48(2):785–795, 2025.

[Daunizeau *et al.*, 2011] Jean Daunizeau, Olivier David, and Klaas E Stephan. Dynamic causal modelling: a critical review of the biophysical and statistical foundations. *Neuroimage*, 58(2):312–322, 2011.

[DSouza *et al.*, 2017] Adora M DSouza, Anas Z Abidin, Lutz Leistritz, and Axel Wismüller. Exploring connectivity with large-scale granger causality on resting-state functional mri. *Journal of neuroscience methods*, 287:68–79, 2017.

[Duong *et al.*, 2025] Bao Duong, Sunil Gupta, and Thin Nguyen. Causal discovery via bayesian optimization. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu, editors, *International Conference on Learning Representations*, volume 2025, pages 54318–54351, 2025.

[Friston, 2011] Karl J Friston. Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36, 2011.

[Gu *et al.*, 2022] Yameng Gu, Feng Han, Lucas E Sainburg, Margeaux M Schade, Orfeu M Buxton, Jeff H Duyn, and Xiao Liu. An orderly sequence of autonomic and neural events at transient arousal changes. *NeuroImage*, 264:119720, 2022.

[Herwig *et al.*, 2003] Uwe Herwig, Peyman Satrapi, and Carlos Schönfeldt-Lecuona. Using the international 10-20 eeg system for positioning of transcranial magnetic stimulation. *Brain topography*, 16(2):95–99, 2003.

[Huang *et al.*, 2018] Biwei Huang, Kun Zhang, Yizhu Lin, Bernhard Schölkopf, and Clark Glymour. Generalized score functions for causal discovery. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1551–1560, 2018.

[Hyvarinen, 2010] Aapo Hyvarinen. Pairwise measures of causal direction in linear non-gaussian acyclic models. In *Proceedings of 2nd Asian Conference on Machine Learning*, pages 1–16. JMLR Workshop and Conference Proceedings, 2010.

[Ji *et al.*, 2025] Junzhong Ji, Jingdong Fan, and Jinduo Liu. Msad-lec: Estimating large-scale brain effective connectivity network based on multi-subgraph attention diffusion. *Knowledge-Based Systems*, 309:112858, 2025.

[Li *et al.*, 2024] Loka Li, Ignavier Ng, Gongxu Luo, Biwei Huang, Guangyi Chen, Tongliang Liu, Bin Gu, and Kun Zhang. Federated causal discovery from heterogeneous data. In *The Twelfth International Conference on Learning Representations*, 2024.

[Li *et al.*, 2025] Shengrong Li, Qi Zhu, Liang Sun, Kai Ma, Yixin Ji, Shile Qi, and Daoqiang Zhang. Multi-modal brain network fusion for intelligent diagnostic devices. *IEEE Transactions on Consumer Electronics*, 71(2):3654–3666, 2025.

[Liang *et al.*, 2024] Guojun Liang, Prayag Tiwari, Sławomir Nowaczyk, Stefan Byttner, and Fernando Alonso-Fernandez. Dynamic causal explanation based diffusion-variational graph neural network for spatiotemporal forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 36(5):9524–9537, 2024.

[Lioi *et al.*, 2020] 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(1):173, 2020.

[Liu *et al.*, 2022] Jinduo Liu, Junzhong Ji, Guangxu Xun, and Aidong Zhang. 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, 2022.

[Liu *et al.*, 2024] Jinduo Liu, Lu Han, and Junzhong Ji. 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.

[Mamoon *et al.*, 2025] Saqib Mamoon, Zhengwang Xia, Amani Alfakih, and Jianfeng Lu. Dynamic brain effective connectivity network for identifying neurological disorders. *Applied Intelligence*, 55(12):850, 2025.

[Marinazzo *et al.*, 2011] Daniele Marinazzo, Wei Liao, Huafu Chen, and Sebastiano Stramaglia. Nonlinear connectivity by granger causality. *Neuroimage*, 58(2):330–338, 2011.

[Muthuraman, 2016] Muthuraman Muthuraman. Effective connectivity of cortical sensorimotor networks during finger movement tasks: a simultaneous fnirs, fmri, eeg study. *Brain topography*, 29:645–660, 2016.

[Patel *et al.*, 2006] Rajan S Patel, F DuBois Bowman, and James K Rilling. A bayesian approach to determining connectivity of the human brain. *Human brain mapping*, 27(3):267–276, 2006.

[Sakkalis, 2011] Vangelis Sakkalis. Review of advanced techniques for the estimation of brain connectivity measured with eeg/meg. *Computers in biology and medicine*, 41(12):1110–1117, 2011.

[Sanchez *et al.*, 2023] Pedro Sanchez, Xiao Liu, Alison Q O'Neil, and Sotirios A Tsaftaris. Diffusion models for causal discovery via topological ordering. In *The Eleventh International Conference on Learning Representations*, 2023.

[Stramaglia *et al.*, 2014] Sebastiano Stramaglia, Jesus M Cortes, and Daniele Marinazzo. Synergy and redundancy in the granger causal analysis of dynamical networks. *New Journal of Physics*, 16(10):105003, 2014.

[Su *et al.*, 2025] Zhihao Su, Jihao Zhai, Junzhong Ji, and Jinduo Liu. MSGFlowNet: Learning Effective Connectivity Network based on Sparse Generative Flow Network from fMRI and EEG Data . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume LNCS 15960. Springer Nature Switzerland, September 2025.

[Tank *et al.*, 2021] Alex Tank, Ian Covert, Nicholas Foti, Ali Shojaie, and Emily B Fox. Neural granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4267–4279, 2021.

[Tu *et al.*, 2019] Tao Tu, John Paisley, Stefan Haufe, and Paul Sajda. A state-space model for inferring effective connectivity of latent neural dynamics from simultaneous eeg/fmri. *Advances in Neural Information Processing Systems*, 32, 2019.

[Valdes-Sosa *et al.*, 2009] Pedro A Valdes-Sosa, Jose Miguel Sanchez-Bornot, Roberto Carlos Sotero, Yasser Iturria-Medina, Yasser Aleman-Gomez, Jorge Bosch-Bayard, Felix Carbonell, and Tohru Ozaki. Model driven eeg/fmri fusion of brain oscillations. *Human brain mapping*, 30(9):2701–2721, 2009.

[Warbrick, 2022] Tracy Warbrick. Simultaneous eeg-fmri: what have we learned and what does the future hold? *Sensors*, 22(6):2262, 2022.

[Wei *et al.*, 2025] Xinxu Wei, Kanhao Zhao, Yong Jiao, Nancy B Carlisle, Hua Xie, Gregory A Fonzo, and Yu Zhang. Multi-modal cross-domain self-supervised pre-training for fmri and eeg fusion. *Neural Networks*, 184:107066, 2025.

[Xiong *et al.*, 2025a] Wen Xiong, Junzhong Ji, and Jinduo Liu. Brainec-llm: Brain effective connectivity estimation by multiscale mixing llm. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen, editors, *Advances in Neural Information Processing Systems*, volume 38, pages 123110–123143. Curran Associates, Inc., 2025.

[Xiong *et al.*, 2025b] Wen Xiong, Jinduo Liu, Junzhong Ji, and Fenglong Ma. Brain effective connectivity estimation via fourier spatiotemporal attention. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pages 1657–1668, 2025.

[Yu *et al.*, 2025] Minqi Yu, Jinduo Liu, and Junzhong Ji. Causal invariance-aware augmentation for brain graph contrastive learning. In *International Conference on Machine Learning*, 2025.

[Yun *et al.*, 2025] Xudong Yun, Weidong Cai, Junjun Fun, Xiaoliang Zhu, and Xin Zhao. State-space modeling uncovers brain-behavior dynamics of inhibitory control. *Cerebral Cortex*, 35(7):bhaf197, 2025.

[Zhang *et al.*, 2024a] Xiaoshan Zhang, Enze Shi, Sigang Yu, and Shu Zhang. 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*, pages 141–151. Springer, 2024.

[Zhang *et al.*, 2024b] Zuozhen Zhang, Junzhong Ji, and Jinduo Liu. Metarlec: Meta-reinforcement learning for discovery of brain effective connectivity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 10261–10269, 2024.

[Zhao *et al.*, 2022] Min Zhao, Weizheng Yan, Na Luo, Dongmei Zhi, Zening Fu, Yuhui Du, Shan Yu, Tianzi Jiang, Vince D Calhoun, and Jing Sui. An attention-based hybrid deep learning framework integrating brain connectivity and activity of resting-state functional mri data. *Medical image analysis*, 78:102413, 2022.