Computers in Biology and Medicine 189 (2025) 109898

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Brain imaging-to-graph generation using adversarial hierarchical diffusion models for MCI causality analysis

Qiankun Zuo [a,b,c], Hao Tian [a,b,*], Yudong Zhang [d], Jin Hong [e,*]

[a] Hubei Key Laboratory of Digital Finance Innovation, Hubei University of Economics, Wuhan 430205, Hubei, China
[b] School of Information Engineering, Hubei University of Economics, Wuhan 430205, Hubei, China
[c] Hubei Internet Finance Information Engineering Technology Research Center, Hubei University of Economics, Wuhan 430205, Hubei, China
[d] School of Computing and Mathematical Sciences, University of Leicester, Leicester LE1 7RH, United Kingdom
[e] School of Information Engineering, Nanchang University, Nanchang, 330031, China

ARTICLE INFO

Dataset link: http://adni.loni.usc.edu/

Keywords:
Adversarial diffusive denoising
Multi-resolution transformer
Spatiotemporal enhanced feature
Brain effective connectivity
MCI

ABSTRACT

Effective connectivity can describe the causal patterns among brain regions. These patterns have the potential to reveal the pathological mechanism and promote early diagnosis and effective drug development for cognitive disease. However, the current methods utilize software toolkits to extract empirical features from brain imaging to estimate effective connectivity. These methods heavily rely on manual parameter settings and may result in large errors during effective connectivity estimation. In this paper, a novel brain imaging-to-graph generation (BIGG) framework is proposed to map functional magnetic resonance imaging (fMRI) into effective connectivity for mild cognitive impairment (MCI) analysis. The proposed BIGG framework is based on the diffusion denoising probabilistic models (DDPM), where each denoising step is modeled as a generative adversarial network (GAN) to progressively translate the noise and conditional fMRI to effective connectivity. By introducing the diffusive factor, the denoising inference with a large sampling step size is more efficient and can maintain high-quality results. Evaluations of the ADNI dataset demonstrate the feasibility and efficacy of the proposed model. The proposed model not only achieves superior prediction performance compared with other competing methods but also predicts MCI-related causal connections that are consistent with clinical studies.

1. Introduction

Brain effective connectivity (BEC) describes the causal interaction from one brain region to another and helps perform various cognitive and perceptual tasks [1]. It aims to transmit and analyze functional information between distant regions of the brain with the characteristics of high speed and precision [2]. Recent studies have shown that the abnormal changes of BEC can reflect the underlying pathology of brain diseases [3,4]. These changes are probably accompanied by alterations in the brain's microstructure [5,6], such as those associated with Alzheimer's disease (AD) and its early-stage mild cognitive impairment (MCI). Investigation of the BEC is helpful for researchers to understand the underlying mechanisms associated with neurodegenerative diseases and develop potential treatments or new drugs for rehabilitation [7,8]. Therefore, constructing BEC from brain imaging (i.e., functional magnetic resonance imaging, fMRI) becomes a very promising way to analyze cognitive disease and identify possible biomarkers for MCI diagnosis.

Constructing BEC involves learning a mapping network to predict directional connections from one neural unit to another by analyzing their functional signals. Since the fMRI has the advantages of being noninvasive and having high temporal resolution, it has drawn great attention to extracting the complex connectivity features for brain disease diagnosis [9]. The brain functional connectivity (BFC) gives the temporal correlation between any pair of neural units, while the BEC is an asymmetric matrix representing the directional information of neural transmission. The directed graph can be used to analyze BEC, including a set of vertices (named regions-of-interest, or ROIs) and a set of directed edges (effective connections). Previous researchers focused on BEC estimation by using traditional learning methods, including the dynamic causal models (DCM) [10], the Bayesian network (BN) method [11], the correlation analysis method [12], the transfer entropy (TE) method [13], and so on. These methods utilize shallow network structures or prior knowledge, which are unable to extract complex causal features from fMRI and bring low performance in

* Corresponding authors.
E-mail addresses: th@hbue.edu.cn (H. Tian), hongjin@ncu.edu.cn (J. Hong).

https://doi.org/10.1016/j.compbiomed.2025.109898
Received 18 September 2024; Received in revised form 22 January 2025; Accepted 18 February 2025
Available online 28 February 2025
0010-4825/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

disease analysis.

Deep learning methods have been widely applied in the exploration of BEC estimation [14,15]. They not only achieve excellent performance on image recognition tasks in Euclidean space but also show good results on brain network generation in non-Euclidean space. The primary characteristic is the strong ability to perform high-level and complex feature extraction. Increasingly new methods based on deep learning have been explored to construct BEC from functional MRI data [16]. However, the current methods heavily rely on the software toolkit to preprocess the fMRI for extracting empirical features (empirical ROI-based time series). The main drawback is that the manual parameter settings from different researchers may result in large errors when using these empirical features to estimate BEC.

As the most popular and powerful generative model, the generative adversarial network (GAN) [17] implicitly characterizes the distribution of synthetic samples through a two-player adversarial game [18]. It can generate high-quality samples with efficient computation while producing homogeneous samples because of training instability and mode collapse. An alternative way to solve this issue is the emergence of diffusion denoising probabilistic models (DDPM) [19] that have received great attention in generating tasks [20]. The main advantage of DDPM is its powerful ability to generate high-quality and diverse samples, and the disadvantage lies in its expensive computation. Inspired by the above observations, we intend to combine the advantages of the GAN and the DDPM for generation performance improvement. A novel brain imaging-to-graph generation (BIGG) model is proposed to generate BEC from fMRI to analyze MCI causality. Specifically, the DDPM is split into several diffusive steps, and each step is modeled as a conditional GAN to denoise the Gaussian sample and generate clean samples. The transformer-based generator and discriminator are combined to capture hierarchical spatial–temporal features for noise removal. During denoising inference, a diffusive factor is introduced to increase sampling step size for efficient and high-quality sample generation. As a result, the estimated BEC reflects a more sophisticated causal relationship between brain regions and captures the MCI-related features. To the best of our knowledge, the proposed BIGG is a unified AIGC framework that firstly translates fMRI into BEC. The main contributions of this work are summarized as follows:

- The proposed BIGG model is a unified framework to map fMRI onto effective connectivity in an end-to-end manner. It leverages the adversarial strategy to model each denoising step of DDPM by introducing conditional fMRI, which is high-quality, diverse, and efficient for effective connectivity generation.
- The hierarchical denoising transformer is designed to learn multi-scale features for noise removal. By focusing on both spatial and temporal domains, the quality of the denoising process is greatly improved, and the performance of causal inference is significantly enhanced.
- The multi-resolution consistent transformer is devised to approximate the temporal distribution of diffusive samples to that of denoised samples at different scales. It promotes the generator to capture global and local patterns for good generation in terms of diversity and stability.

The rest of this paper is structured as follows: The related works are introduced in Section 2. The overall architecture of the proposed BIGG model is presented in Section 3. The experimental results, including generation evaluation and classification performance, are described in Section 4. The reliability of our results is discussed in Section 5, and Section 6 draws the main conclusions.

2. Related work

2.1. BEC learning methods

Functional connectivity network analysis is usually used to identify abnormal connectivity patterns that are associated with different cognitive functions, such as memory, attention, and emotion. The BEC belongs to the functional connectivity network and can bridge causal connections between brain regions. Many studies have focused on exploring the BEC for better diagnostic performance and good interpretability. There are two categories of methods for learning BEC from functional data: shallow learning methods and deep learning methods.

The frequently used shallow method is the dynamic causal model (DCM) [10]. For example, Park et al. [21] employed the parametric empirical Bayes method to model the directed effects of sliding windows. And the Granger causality (GC) [22,23] is the most commonly used shallow method. For example, DSouza et al. [24] utilize the multiple regression algorithm to process historical information from functional time series for causal interaction prediction. These methods cannot extract deep and complex connectivity features from fMRI.

To explore the deep features of brain imaging, deep learning methods show great success in causal modeling between brain regions. The work in [25] employed nonlinear causal relationship estimation with an artificial neural network to predict causal relations between brain regions by analyzing both linear and nonlinear components of electroencephalogram (EEG) data. Also, Abbasvandi et al. [26] combined the recurrent neural network and Granger causality to estimate effective connectivity from EEG data. They greatly improved the prediction accuracy of the simulation data and the epileptic seizure dataset. Considering the great ability of GANs to characterize data distribution, Liu et al. [27] designed a GAN-based network to infer directed connections from fMRI data. To capture temporal features, they [28] employed recurrent generative adversarial networks for effective connectivity learning. Presently, Zou et al. [29] introduced the graph convolutional network (GCN) to mine both temporal and spatial topological relationships among distant brain regions for learning BECs. Zhang et al. [30] established a causal connectivity model using fMRI to analyze brain diseases. Due to the ability of generative models in learning complex brain connectivity causality, the work in [31] took account of the dynamic features of EC and constructed a nonparametric state estimation model for capturing the abnormal patterns of EC. Liu et al. [32] investigated multimodal imaging data to estimate EC and applied it in a visual categorization task. Besides, Ji et al. [33] introduced the meta-knowledge into an autoencoder to model the causal relationships between brain regions. Although the above deep learning methods achieved promising prediction performance in BEC estimation, they heavily rely on the software toolkit to preprocess the fMRI for extracting empirical time-series data. That may result in large errors due to different manual parameter settings during preprocessing procedures.

2.2. Generative learning models

Generative adversarial networks (GANs) have ruled generative approaches since they were first proposed by Goodfellow. The primary advantage is implicitly modeling the distribution of the generated samples through a two-player game. Many variants of GANs have been proposed and applied to many generation tasks, such as generating super-resolution, synthesizing cross-modal data, segmenting images, and so on. To satisfy specific generating tasks, conditional GAN [34] and related variants also achieve quite good performance efficiently [35]. The problem of instability and mode collapse in training has not been completely addressed yet. This may lead to homogeneously generated data and hinder its wider applications. Recently, denoising diffusion probabilistic models (DDPM) have attracted much attention in image generation [36] because of their ability to generate high-quality and diverse samples. DDPM aims to denoise the Gaussian sample gradually and recover the clean sample. However, the denoising process requires a Gaussian distribution assumption with only a small denoising step, which leads to slow reverse diffusion in about thousands of steps to approach clean data. Based on the above observations, we try to combine the advantages of GAN and DDPM in generation, such as efficiency, high quality, and diversity. We propose the novel BIGG model to precisely estimate BEC from fMRI for MCI causality analysis.

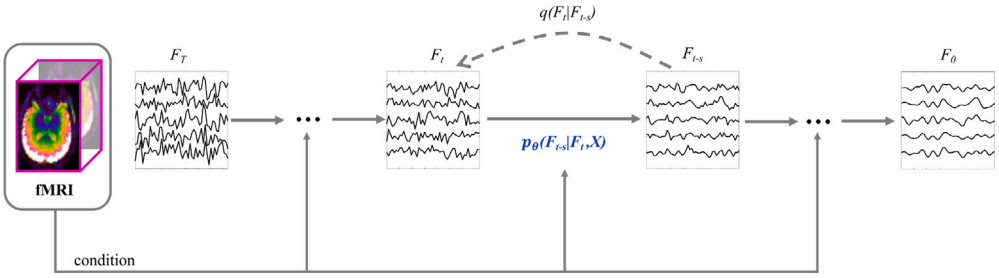

Fig. 1. The architecture of the proposed BIGG framework. In the forward process, the empirical sample F_0 is transformed to the normal Gaussian noisy sample F_T by gradually adding noise. In the reverse process, the fMRI is considered as a condition to guide the network p_θ to remove the noise from F_T.

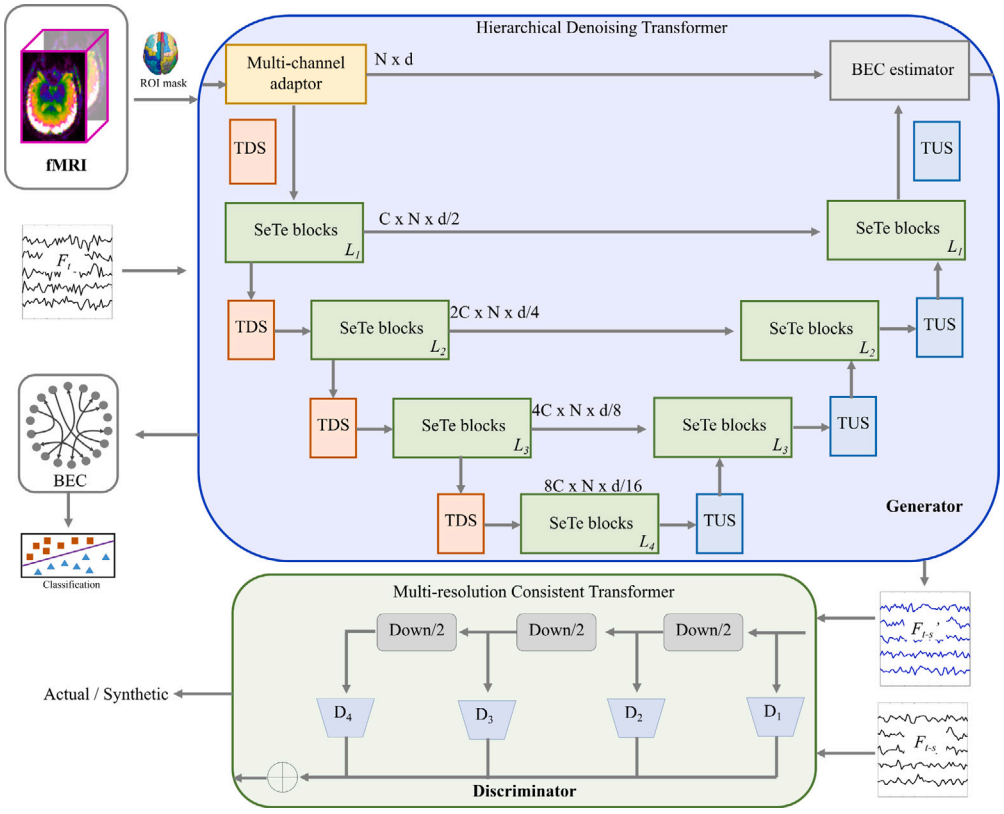

Fig. 2. Detailed network structure of one denoising step. It includes one generator, one discriminator, and one classifier. The input is the conditional fMRI and noisy sample F_t, and the output is the denoised sample F'_{t-s} and BEC.

3. Method

As shown in Fig. 1, the proposed BIGG model is a diffusion-based model to denoise a Gaussian sample into a clean sample. At each denoising step, the fMRI is treated as a condition to guide a GAN-based network to remove noise and estimate BEC. The details of one denoising step are shown in Fig. 2, including one generator and one discriminator. At the final denoising step, the obtained BEC is used to analyze MCI causality. In the following, we first introduce conventional diffusion models and then describe the adversarial hierarchical diffusion models with transformer-based generators and discriminators. Finally, hybrid objective functions are devised to optimize the proposed BIGG model. There are 210 subjects used in this study to determine the model's parameters. We select the traditional AAL90 atlas with 90 brain regions. The length of the time series is 187. The mentioned symbols in the method are listed in Table 1.

3.1. Conventional denoising diffusion model

The basic principle of the diffusion model is to learn the information attenuation caused by noise and then use the learned noise to generate a clean sample. It is usually divided into the forward process and the inverse process. In the diffusion process, Gaussian noise is constantly added to the input sample F_0 for sufficiently large T (hundreds or thousands) steps. Under the rules of the Markov chain, the probability distribution of a noisy sample F_T will approach the stationary distribution (such as the Gaussian distribution) at the Tth step. Here is the diffusion formula:

$$q\left(\mathbf{F}_{1:T} \mid \mathbf{F}_0\right) = \prod_{t=1}^{T} q\left(\mathbf{F}_t \mid \mathbf{F}_{t-1}\right) \tag{1}$$

$$q\left(\mathbf{F}_t \mid \mathbf{F}_{t-1}\right) = \mathcal{N}\left(\mathbf{F}_t; \sqrt{1-\beta_t}\mathbf{F}_{t-1}, \beta_t \mathbf{I}\right) \tag{2}$$

where $t = \{1, 2, \ldots, T\}$. β_t is the noise variance that is defined before the model's training. $\mathcal{N}$ is the assumed Gaussian distribution, and $\mathbf{I}$ is an identity matrix. The reverse process also follows the Markov chain to

Table 1
Summary of symbols.

Symbol	Meaning
T	Maximum steps during the diffusion process
F_0	Empirical (clean) sample
F_T	Noisy sample at Tth step
F_t	Noisy sample at tth step during the diffusion process
F'_t	Denoised sample at tth step during the reverse process
s	Sampling interval
β_t	Predefined noise variance at tth step
$\mathcal{N}$	Gaussian distribution
q	Probability distribution during the diffusion process
p_θ	Probability distribution during the reverse process
θ	Model's parameters
$\boldsymbol{\mu}_\theta(\mathbf{F}_t, t)$	Mean value of the denoised sample $\mathbf{F}'_{t-1}$
σ_t^2	Variance of the denoised sample $\mathbf{F}'_{t-1}$
$\boldsymbol{\epsilon}$	Gaussian noise
$\boldsymbol{\epsilon}_\theta$	Estimated noise
$\mathbf{X}$	Rough sample (the condition during reverse process)
$\hat{\mathbf{F}}_0^{[t/s]}$	The initial clean sample at t/s step
N	Number of brain regions, in our experiment, $N = 90$
d	Length of time series, in our experiment, $d = 187$
A	The brain effective connectivity matrix

translate the noisy sample F_T to a cleaned sample F_0. In the reversed diffusion, we defined the intermediate sample as denoised sample. We let $F_T = F'_T$. The assumption of a large step T and a small β_t can model the denoising probability in a Gaussian distribution:

$$p_\theta\left(\mathbf{F}'_{0:T}\right) = p\left(\mathbf{F}'_T\right)\prod_{t=1}^{T} p_\theta\left(\mathbf{F}'_{t-1} \mid \mathbf{F}'_t\right) \tag{3}$$

$$p_\theta\left(\mathbf{F}'_{t-1} \mid \mathbf{F}'_t\right) = \mathcal{N}\left(\mathbf{F}'_{t-1}; \boldsymbol{\mu}_\theta\left(\mathbf{F}'_t, t\right), \sigma_t^2\mathbf{I}\right) \tag{4}$$

where $\boldsymbol{\mu}_\theta(\mathbf{F}'_t, t)$ and σ_t^2 are the mean and variance of the denoised sample $\mathbf{F}'_{t-1}$. θ indicates the network's parameters.

We use the deep learning model $p_\theta(\mathbf{F}'_{t-1}|\mathbf{F}'_t)$ to approximate the true distribution $q(\mathbf{F}'_{t-1}|\mathbf{F}'_t)$, and get the reparameterization of μ and σ_t^2 in the following form:

$$\boldsymbol{\mu}_\theta\left(\mathbf{F}'_t, t\right) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{F}'_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta\left(\mathbf{F}'_t, t\right)\right) \tag{5}$$

$$\sigma_t^2 = \frac{\left(1-\bar{\alpha}_{t-1}\right)}{\left(1-\bar{\alpha}_t\right)}\beta_t \tag{6}$$

here, $\alpha_t = 1-\beta_t, \bar{\alpha}_t = \prod_{i=1}^{t}\alpha_i$. By applying a variational evidence lower bound (ELBO) constraint, the added noise $\epsilon_\theta(\mathbf{F}'_t, t)$ can be calculated by minimizing the MSE loss:

$$\mathbb{E}_{t,\mathbf{F}_0,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{F}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t\right)\right\|^2\right] \tag{7}$$

where, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. After the model's optimization, new samples can be derived from Gaussian noise $F_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ by reversely diffusing T steps based on Eqs. (3) and (4).

3.2. Adversarial hierarchical diffusion models

Conventional DDPM can generate high-quality samples but suffers from low efficiency in sampling because of the thousands of denoising steps. While the GAN can make up for this shortcoming by having fast-generating ability, Thus, we utilize adversarial hierarchical diffusion models by combining both DDPM and GAN for efficient and high-fidelity sampling generation. Apart from this, there are two other differences compared with the conventional DDPM: (1) fMRI is treated as a condition to guide the denoising process to generate a clean subject-specific sample using our model; the conventional diffusion process likely generates uncontrollable samples that cannot reflect subject-specific disease information; (2) the step size is reduced by a diffusive factor of s ($s \gg 1$), which speeds the generation process and keeps the sample generation high-quality. Before the diffusion process, the conditional fMRI is parcellated to obtain the rough sample $\mathbf{X}$ using the non-parameter method. In this method, the voxels in the corresponding brain regions of the fMRI and the anatomical atlas (N ROIs) are dotted and summed, and the output is the rough ROI-based two-dimensional time-series (rough sample $\mathbf{X}$).

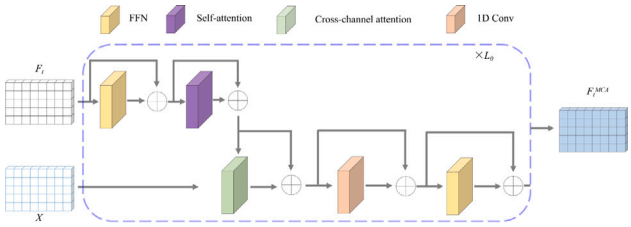

Fig. 3. The structure of the multi-channel adaptor. The inputs are the noisy sample $\mathbf{F}_t$ and the rough sample $\mathbf{X}$, and the output is the fused sample. The three samples share the same dimension.

The adversarial hierarchical diffusion models have several denoising steps. Each step adopts a conditional GAN to remove the noise from previous noisy sample. As shown in Fig. 2, it consists of two parts: one generator and one discriminator. In the forward direction, the empirical sample $\mathbf{F}_0$ is transformed to the normal Gaussian noisy sample $\mathbf{F}_T$ by gradually adding noise. The computation formula is the same as Eq. (2). In the reverse direction, we incorporate the conditional $\mathbf{X}$ into the denoising procedure. First, the rough sample $\mathbf{X}$ is considered as a condition to guide the generator $\mathbf{G}_\theta$ predict the initial sample $\hat{\mathbf{F}}_0$ from $\mathbf{F}_T$, then the posterior sampling is utilized to synthesize the denoised sample $\mathbf{F}'_{t-s}$; meanwhile, the transformer-based discriminator $\mathbf{D}_\theta$ distinguishes the actual ($\mathbf{F}_{t-s}$) or synthetic ($\mathbf{F}'_{t-s}$) for the denoising process. Specifically, at the t step, we aim to predicte $\mathbf{F}'_{t-s}$ from $\mathbf{F}'_t$. Firstly, a generator $\mathbf{G}_\theta(\mathbf{F}'_t, \mathbf{X}, t)$ is utilized to predict the initial sample $\hat{\mathbf{F}}_0^{[t/s]}$, then $\mathbf{F}'_{t-s}$ is sampled using the posterior distribution $q(\mathbf{F}'_{t-s}|\mathbf{F}'_t, \hat{\mathbf{F}}_0^{[t/s]})$ by giving $\mathbf{F}'_t$ and $\hat{\mathbf{F}}_0^{[t/s]}$. Finally, after T/s steps, the ultimate denoised sample $\mathbf{F}'_0$ (equal to $\hat{\mathbf{F}}_0^{[1]}$) sampled from the estimated distribution $p_\theta(\mathbf{F}'_0 \mid \mathbf{F}'_s, \mathbf{X})$. The denoising process can be expressed as follows:

$$p_\theta\left(\mathbf{F}'_{t-s} \mid \mathbf{F}'_t, \mathbf{X}\right) = q\left(\mathbf{F}'_{t-s} \mid \mathbf{F}'_t, \hat{\mathbf{F}}_0^{[t/s]} = \mathbf{G}_\theta^{[t/s]}\left(\mathbf{F}'_t, \mathbf{X}, t\right)\right) \tag{8}$$

$$q\left(\mathbf{F}'_{t-s} \mid \mathbf{F}'_t, \hat{\mathbf{F}}_0^{[t/s]}\right) = q\left(\mathbf{F}'_t \mid \mathbf{F}'_{t-s}, \hat{\mathbf{F}}_0^{[t/s]}\right)\frac{q\left(\mathbf{F}'_{t-s} \mid \hat{\mathbf{F}}_0^{[t/s]}\right)}{q\left(\mathbf{F}'_t \mid \hat{\mathbf{F}}_0^{[t/s]}\right)} \tag{9}$$

The sampling probability from $p_\theta\left(\mathbf{F}'_{t-s} \mid \mathbf{F}'_t, \mathbf{X}\right)$ is defined as:

$$\mathbf{F}'_{t-s} = \frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-s}\right)}{1-\bar{\alpha}_t}\mathbf{F}'_t + \frac{\sqrt{\bar{\alpha}_{t-s}}\beta_t}{1-\bar{\alpha}_t}\mathbf{G}_\theta^{[t/s]} + \epsilon\sqrt{\beta_t} \tag{10}$$

3.2.1. Generator with the hierarchical denoising transformer

The aim of the transformer-based generator ($\mathbf{G}_\theta$) is to remove the noise from noisy samples to obtain clean samples by conditional guidance. Specifically, at t-the step, the input is the noisy sample $\mathbf{F}'_t$ and the rough sample $\mathbf{X}$, and the output is the initial denoised sample $\hat{\mathbf{F}}_0$. All of them share the same size, $N \times d$. The generator consists of four modules, including the multi-channel adaptor (MCA), the spatial-enhanced temporal-enhanced (SeTe) blocks, the temporal down- and up-sampling (TDS and TUS), and the brain effective connectivity (BEC) estimator.

The MCA adaptively fuses the noisy sample $\mathbf{F}'_t$ and the rough sample $\mathbf{X}$, Different from the traditional way of concatenating the two samples, we designed a cross-channel attention mechanism when fusing them. As shown in Fig. 3, first, the sample $\mathbf{F}'_t$ is passed through the feed-forward network (FFN) and self-attention network; then, it is projected on the $\mathbf{X}$ to compute weighting scores for itself. Next, the 1×3 convolutional kernel is applied for each ROI to extract local features. Finally, FFN is

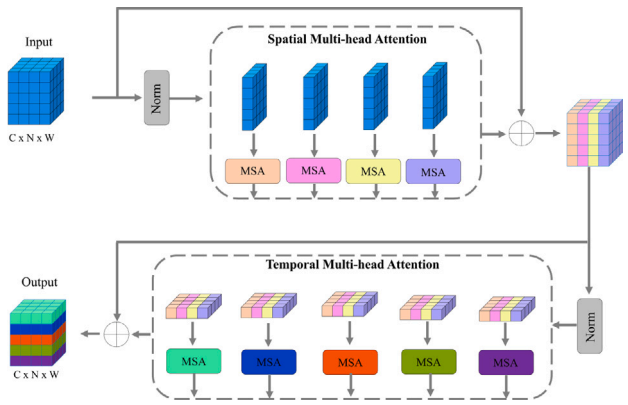

Fig. 4. The detailed structure of the SeTe block. The input and output share the same dimension. It passes successively through the spatial multi-head attention module and the temporal multi-head attention module.

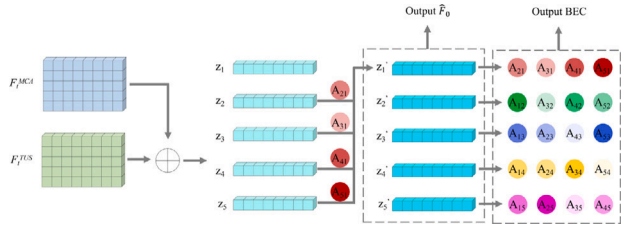

Fig. 5. The structure of the BEC estimator. The outputs are a denoised sample and an asymmetric brain connectivity matrix.

used to adjust temporal features among different channels. The output of every sub-network has the same size, $N \times d$. The output of the MCA module is $\mathbf{F}_t^{MCA}$.

The TDS is used to halve the feature dimension and add the channels. Taking the first TDS as an example, for each ROI feature, we apply a 1D convolutional kernel with the size 1×3 to extract local temporal features, and through C convolutional kernels with a step size of 2, each ROI feature is translated to a sequence of vectors with the length $d/2$. The final output $\mathbf{F}_t^{TDS}$ has the size $C \times N \times d/2$.

The SeTe blocks are designed to extract both spatial and temporal features. The conventional transformed-based method focuses on the spatial correlation between pairs of ROIs while ignoring the temporal continuity. The SeTe benefits from multi-head attention (MSA), which enhances long-term dependence both spatially and temporally. As shown in Fig. 4, it is comprised of a spatial multi-head attention (SMA) and a temporal multi-head attention (TMA). The difference between these two attention networks is that the former is operated in the spatial direction, and the latter is operated in the temporal direction. The input is a tensor with the size $C \times N \times d/2$, and the output $\mathbf{F}_t^{SeTe}$ can be defined by:

$$\mathbf{F}_t'^{SeTe} = \mathrm{SMA}\left(\mathrm{Norm}\left(\mathbf{F}_t^{TDS}\right)\right) + \mathbf{F}_t^{TDS} \tag{11}$$

$$\mathbf{F}_t^{SeTe} = \mathrm{TMA}\left(\mathrm{Norm}\left(\mathbf{F}_t'^{SeTe}\right)\right) + \mathbf{F}_t'^{SeTe} \tag{12}$$

here, the $\mathbf{F}_t'^{SeTe}$ and $\mathbf{F}_t^{SeTe}$ have the same size, $C \times N \times d/2$. Specifically, the SMA splits the input $\mathbf{F}_t^{SeTe}$ into several parallel parts and applies MSA to each of them. We denote H as the head number. The detailed computation of SMA is expressed below:

$$\mathbf{F}_t^{SeTe} = \{\mathbf{F}_t^{SeTe(1)}, \mathbf{F}_t^{SeTe(2)}, \ldots, \mathbf{F}_t^{SeTe(d)}\} \tag{13}$$

$$\begin{aligned}\hat{\mathbf{F}}_t^{SeTe(i)} &= \mathrm{Att}\left(\mathbf{F}_t^{SeTe(i)}\mathbf{W}_h^Q, \mathbf{F}_t^{SeTe(i)}\mathbf{W}_h^K, \mathbf{F}_t^{SeTe(i)}\mathbf{W}_h^V\right) \\ &= \mathrm{Att}(\mathbf{Q}_h^i, \mathbf{K}_h^i, \mathbf{V}_h^i) \\ &= \mathrm{softmax}\left(\frac{\mathbf{Q}_h^i(\mathbf{K}_h^i)^T}{\sqrt{C/H}}\right)\mathbf{V}_h^i\end{aligned} \tag{14}$$

$$\mathbf{F}_t'^{SeTe} = \{\hat{\mathbf{F}}_t^{SeTe(1)}, \hat{\mathbf{F}}_t^{SeTe(2)}, \ldots, \hat{\mathbf{F}}_t^{SeTe(d)}\} \tag{15}$$

where, $\mathbf{F}_t^{SeTe(i)} \in \mathbb{R}^{N \times C}$, $i = \{1, 2, \ldots, d\}$, $h = \{1, 2, \ldots, H\}$. Att means the attention operation, $\mathbf{W}_h^Q, \mathbf{W}_h^K, \mathbf{W}_h^V \in \mathbb{R}^{C \times (C/H)}$ project each part of $\hat{\mathbf{F}}_t^{SeTe}$ onto the matrices of queries ($\mathbf{Q}$), keys ($\mathbf{K}$), and values ($\mathbf{V}$) for the hth head, respectively. The outputs of all attention mechanisms are concatenated to get the final result of this module in this layer. The TMA also has a similar structure and definitions as described above.

The TUS is the reverse of the TDS, where the dimension is doubled and the channel is halved. Taking the last TUS as an example, with the concatenated sequence $\mathbf{F}_t^{SeTe} \in \mathbb{R}^{C \times N \times d}$, we apply several 1D transposed convolutions to reduce the channel and increase the dimension. At last, the output, $\mathbf{F}_t^{TUS}$ has the same size as $\mathbf{F}_t^{MCA}$.

The BEC aims to generate a denoised sample $\hat{\mathbf{F}}_0$ and estimate the causal direction between pairs of ROIs. As shown in Fig. 5, the inputs are the noisy samples: $\mathbf{F}_t^{MCA}$ and $\mathbf{F}_t^{TUS}$. After the element adding operation, we can obtain the denoised sample at t/s step:

$$\hat{\mathbf{F}}_0 = \mathbf{F}_t^{MCA} + \mathbf{F}_t^{TUS} \tag{16}$$

here, we separate the $\hat{\mathbf{F}}_0$ into multiple rows, where each row represents the corresponding ROI's feature. To mine the causal relationship among ROIs, we introduce the structural equation model (SEM) to predict the direction from one region to another. The causal parameters of SEM can be estimated by:

$$z_i' = \sum_{j=1}^{N} \mathbf{A}_{ji} z_j + n_i \tag{17}$$

where $i = \{1, 2, \ldots, N\}$ $\mathbf{A}_{ji}$ indicates the causal effect on ith brain region from jth brain region. n_i is the independent random noise. The matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ is asymmetric, representing the causal parameters of SEM. The diagonal elements of $\mathbf{A}$ are set to 0 because it is meaningless to consider the effective connectivity of the brain region itself. Therefore, the BEC matrix $\mathbf{A}$ can represent the effective connectivity between any pair of brain regions.

3.2.2. Multi-resolution consistent discriminator

To constrain the generated denoised sample $\mathbf{F}_{t-s}'$ to be consistent with the real sample $\mathbf{F}_{t-s}$ in distribution, we downsample the input sample into four different resolutions and devise four corresponding sub-discriminators to distinguish the synthetic and actual samples. Each discriminator has the conventional transformer structure: layer normalization, self-attention, a feed-forward layer, and the classification head. The output of each discriminator is a scalar in the range of 0~1. Averaging all the discriminator's outputs is the final output of the multi-resolution consistent discriminator.

3.3. Hybrid loss functions

To guarantee the model generates high-quality denoised samples, adversarial loss is introduced to optimize the generator's parameters. There are five kinds of loss functions: the spatial–temporal enhanced generative loss (L_{SEG}), the multi-resolution consistent discriminative loss (L_{MDD}), the reconstructed loss (L_{REC}), the sparse connective penalty loss (L_{SCP}), and the classification loss (L_{CLS}). We treat the generator as a conditional GAN; when inputting a noisy sample and conditional sample, the generator $\mathbf{G}_\theta(\mathbf{F}_t', \mathbf{X}, t)$ outputs a synthetic sample $\mathbf{F}_{t-s}'$. And the discriminator $\mathbf{D}_\theta$ discriminates whether the sample is from the generator or the forward diffusion. Here are the non-saturating generative and discriminative losses (adversarial diffusive losses):

$$L_{SEG} = \mathbb{E}_{t, q(\mathbf{F}_t|\mathbf{F}_0, \mathbf{X}), p_\theta(\mathbf{F}_{t-s}'|\mathbf{F}_t', \mathbf{X})}\left[-\log\left(D_\theta\left(\mathbf{F}_{t-s}'\right)\right)\right] \tag{18}$$

Table 2
The subject information in the experiment. (Mean ± SD).

Group	NC(81)	EMCI(68)	LMCI(61)
Male/Female	40M/41F	35M/33F	31M/30F
Age (mean ± SD)	76.1 ± 7.9	75.8 ± 7.3	75.9 ± 6.9

$$L_{MDD} = \mathbb{E}_{t,q(\mathbf{F}_t|\mathbf{F}_0,\mathbf{X})}\left\{\mathbb{E}_{p_\theta(\mathbf{F}'_{t-s}|\mathbf{F}'_t,\mathbf{X})}\left[-\log\left(D_\theta\left(\mathbf{F}_{t-s}\right)\right)\right] + \mathbb{E}_{p_\theta(\mathbf{F}'_{t-s}|\mathbf{F}'_t,\mathbf{X})}\left[-\log\left(1-D_\theta\left(\mathbf{F}'_{t-s}\right)\right)\right]\right\} \tag{19}$$

after T steps, the final denoised sample $\mathbf{F}'_0$ should recover the clean sample $\mathbf{F}_0$ for every element. The reconstruction loss is defined by:

$$L_{REC} = \mathbb{E}_{p_\theta(\mathbf{F}_0|\mathbf{F}'_s,\mathbf{X})}[\|\mathbf{F}'_0 - \mathbf{F}_0\|_1] \tag{20}$$

moreover, the sparse effectivity connection between brain regions can be interpreted in brain functional activities. We introduce a penalty on the obtained $\mathbf{A}$ for sparse constraints. Besides, the obtained $\mathbf{A}$ is sent to the classifier C to predict the disease label y. The losses are expressed by:

$$L_{SCP} = \gamma \parallel \sum_{i=1,j=1}^{N} \mathbf{A}_{i,j} \parallel \tag{21}$$

$$L_{CLS} = \mathbb{E}_{p_\theta(\mathbf{A}|\mathbf{F}'_s,\mathbf{X}),C(y|\mathbf{A})}(-log(y|\mathbf{A})) \tag{22}$$

4. Experiments

4.1. Datasets

In our study, we tested our model on the public Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset[1] for classification evaluation. During downloading the data, we set three criteria: (1) Both fMRI and T1-weighted MRI are scanned; (2) the scanning TR of the fMRI is 3. (3) The field strength of the scanning machine is 3.0 T. After successfully preprocessing these data, we have obtained 210 subjects scanned with functional magnetic resonance imaging (fMRI), including 61 subjects with late mild cognitive impairment (LMCI), 68 subjects with early mild cognitive impairment (EMCI), and 81 normal controls (NCs). As shown in Table 2, the average ages of LMCI, EMCI, and NC range from 75 to 76. The sex ratio between males and females is nearly balanced. The fMRI is scanned by Siemens with the following scanning parameters: TR = 3.0 s, field strength = 3.0 T, turning angle = 80.0 degrees, and the EPI sequence is 197 volumes. There are two ways to preprocess the fMRI. Both of them require the anatomical automatic labeling (AAL90) atlas [37] for ROI-based time series computing. One is the routine precedence using the GRETNA software to obtain the empirical ROI-based time series, which is treated as the ground truth $\boldsymbol{F}_0$ in the proposed model. The detailed computing steps using GRETNA [38] are described in the work [39]. Another one adopted in this paper is using the dot product of a nonparametric matrix. The standard Atlas file *aal.nii* splits each volume of fMRI into 90 brain regions and all the voxels of each brain region are averaged. We discard the first 10 volumes and obtain a matrix with a size of 90×187, which is the rough sample (rough ROI-based time series) $\boldsymbol{X}$ input in the model.

4.2. Training settings and evaluation metrics

In the training process, the input of our model is the 4D functional MRI, and the output is the clean ROI-based time series and the BECs. We set the parameters as follows: $T = 1000$, $s = 250$. $C = 2$. $N = 90$, $d = 187$, $L_i = 2$ $(i = 0, 1, 2, 3, 4)$, $\gamma = 1.9$. The Pytorch framework is used to optimize the model's weightings under an Ubuntu 18.04 system. The batch size is 16, and the total epochs are 600. The learning rates for the generator and discriminator are 0.001 and 0.0002, respectively.

We adopt stratified 5-fold cross-validation in our model's validation. Specifically, the subjects in each category are randomly divided into five parts. The model is trained on the four parts of them and tested on the rest. The final accuracy is computed by averaging the results from the five parts. After obtaining the BECs, we conduct three binary classification tasks (i.e., NC vs. EMCI, NC vs. LMCI, and EMCI vs. LMCI) for the model's performance evaluation. Two commonly used classifiers are adopted to evaluate the classification performance, including the support vector machine (SVM) and the BrainNetCNN [40]. The evaluation metrics are the area under the receiver operating characteristic curve (AUC), the prediction accuracy (ACC), the positive sensitivity (SEN), and the negative specificity (SPE). One demo of our model is released at the website (https://github.com/zookhbue/BIGG).

The compared methods are used to construct brain networks and then classify these brain networks using SVM or BrainNetCNN classifiers. They are described as follows: (1) The empirical method preprocesses the fMRI through the GRETNA toolbox and outputs the empirical time series and the functional connectivity. (2) The DCM method [10] models the empirical time series and generates causal connectivity. The sample cross spectra are determined using a fourth-order autoregressive model, and the frequency bins are set to 64. (3) The BN method [11] constructs the effective connectivity from empirical time series. The interaction between brain regions is presumed to be Gaussian. The Markov Chain Monte Carlo (MCMC) search strategy is used to find the optimal network structure. (4) The TE method [13] utilizes the empirical time series to estimate the effective connectivity. The mass of the nearest-neighbors search is set to 4, and the Theiler correction is set to 1. (5) The GCCA method [22] is performed using the MATLAB toolbox; the input is the empirical time series, and the output is the granger causal matrix. The parameters are selected with the default setting. The VAR model estimation regression mode is ordinary least squares (OLS), the maximum model order for model order estimation is 20, and the significance level for the significance test is 0.05. (6) The EC-GAN method [27] consists of 90 generators and discriminators. The input is the empirical time series, and the output is the effective connectivity network. The default parameter configurations are: the number of units is 100, the sparsity parameter is 5, and the maximum number of parents is 5. (7) The STGCM method [29] consists of two graph-based layers, two convolutional layers, and one prediction layer. The input is the empirical time series, and the output is the brain effective connectivity network. The adjustment coefficient is 0.083; the weight coefficients of the prediction loss function and the graph regularization loss function are 1 and 0.5, respectively. The threshold of brain region activation is 0.3. (8) The nPSE-GAN method [31] takes the empirical time series as input and outputs the dynamic effective connectivity. The occurrence probability of each state is 0.3, and the number of states is 2. (9) The MetaCAE method [33] contains a causal autoencoder and meta knowledge transfer, including the dilated causal convolution layer. The input is the empirical time series, and the output is the brain-effective connectivity network. The filter size is 3 and the dilation factor size is 4. The importance of brain effective connectivity sparsity is 0.05, and the directed acyclic graph loss ratio is 20. (10) Our model directly constructs effective connectivity from fMRI by learning the causal relationship between brain regions.

4.3. Prediction results

In order to show the superior performance of the generated BECs, we introduce nine other methods to compare the classification performance. We repeated the classification experiment 10 times. For each repetition, we randomly divided the dataset into five parts and adopted the stratified 5-fold cross-validation strategy to train the model. The final mean and standard deviation values are computed from the results of 10 repetition experiments. The classification results are shown in

[1] http://adni.loni.usc.edu/.

Table 3
Classification performance based on the generated brain networks by different methods (%).

Method	Classifier	NC vs. EMCI				NC vs. LMCI				EMCI vs. LMCI			
		ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Empirical		69.46 (±1.87)	70.14 (±2.77)	68.88 (±3.28)	69.98 (±3.48)	76.26 (±1.81)	74.91 (±2.79)	77.28 (±2.03)	76.69 (±3.45)	76.66 (±1.65)	76.72 (±1.69)	76.61 (±2.13)	77.99 (±2.95)
DCM		72.89 (±2.90)	71.91 (±3.21)	73.70 (±4.84)	72.88 (±2.91)	80.21 (±2.64)	81.48 (±4.44)	79.26 (±4.19)	80.22 (±2.66)	79.77 (±2.02)	78.69 (±3.71)	80.74 (±3.36)	79.76 (±2.03)
BN		73.02 (±2.16)	71.91 (±4.72)	73.95 (±4.05)	73.02 (±2.18)	81.13 (±2.65)	80.49 (±3.90)	81.60 (±3.84)	81.13 (±2.65)	79.53 (±1.97)	80.16 (±3.32)	78.97 (±2.20)	79.54 (±1.98)
TE		73.29 (±2.30)	74.41 (±4.76)	72.35 (±4.20)	73.25 (±2.28)	80.14 (±1.69)	79.67 (±4.65)	80.49 (±4.61)	80.16 (±1.68)	80.31 (±1.91)	81.48 (±3.79)	79.26 (±3.63)	80.31 (±1.91)
GCCA		75.30 (±1.57)	75.14 (±2.13)	75.43 (±2.13)	76.20 (±2.48)	83.16 (±1.60)	82.62 (±2.21)	83.58 (±1.93)	83.03 (±1.83)	82.48 (±1.42)	81.97 (±1.89)	82.93 (±1.85)	82.98 (±2.72)
EC-GAN	SVM	83.08 (±1.17)	82.93 (±1.85)	83.21 (±2.11)	82.69 (±2.23)	90.35 (±1.19)	90.32 (±1.21)	90.37 (±2.08)	90.53 (±1.33)	87.59 (±1.21)	87.21 (±1.85)	87.94 (±1.81)	86.65 (±1.83)
STGCM		84.09 (±0.95)	85.14 (±0.83)	83.21 (±1.66)	85.15 (±1.63)	92.25 (±1.04)	92.12 (±1.50)	92.34 (±1.51)	92.50 (±1.12)	90.54 (±1.40)	91.47 (±2.01)	89.71 (±0.98)	89.67 (±1.83)
nPSE-GAN		83.08 (±1.10)	83.23 (±0.80)	82.96 (±1.61)	83.08 (±1.10)	91.19 (±0.73)	91.52 (±1.23)	90.94 (±0.64)	91.17 (±0.69)	90.38 (±0.88)	88.52 (±1.63)	92.06 (±0.80)	90.38 (±0.88)
MetaCAE		84.42 (±0.99)	84.40 (±0.80)	84.44 (±1.40)	84.40 (±1.01)	92.53 (±0.80)	92.45 (±0.89)	92.59 (±1.23)	92.51 (±0.82)	91.16 (±0.88)	89.83 (±1.37)	92.35 (±1.22)	91.16 (±0.91)
Ours		**85.43 (±0.78)**	**86.31 (±0.99)**	**84.69 (±0.86)**	**86.91 (±1.03)**	**93.45 (±0.88)**	**93.93 (±1.10)**	**93.08 (±1.04)**	**93.68 (±1.13)**	**91.70 (±0.97)**	**91.63 (±1.21)**	**91.76 (±1.23)**	**91.65 (±1.26)**
Empirical		70.93 (±1.34)	70.73 (±2.01)	71.11 (±1.76)	71.87 (±1.83)	78.87 (±1.62)	78.19 (±2.79)	79.38 (±2.02)	78.20 (±2.17)	77.51 (±1.71)	78.03 (±1.58)	77.05 (±2.70)	78.89 (±2.71)
DCM		74.70 (±2.00)	73.68 (±3.76)	75.56 (±2.00)	74.70 (±1.98)	81.69 (±2.82)	85.41 (±3.74)	78.89 (±2.99)	81.72 (±2.83)	81.40 (±1.71)	81.48 (±3.00)	81.32 (±2.69)	81.43 (±1.72)
BN		74.50 (±2.39)	75.15 (±3.21)	73.95 (±3.66)	74.52 (±2.37)	83.17 (±2.36)	85.08 (±4.67)	81.73 (±4.23)	83.17 (±2.38)	81.94 (±2.10)	82.30 (±3.17)	81.62 (±3.62)	81.98 (±2.08)
TE		75.57 (±2.33)	74.41 (±4.45)	76.54 (±2.97)	75.57 (±2.33)	82.25 (±2.65)	84.92 (±3.77)	80.25 (±3.99)	82.25 (±2.63)	82.25 (±2.14)	80.98 (±3.21)	83.38 (±3.92)	82.25 (±2.17)
GCCA		76.24 (±1.38)	76.61 (±2.54)	75.92 (±1.86)	77.48 (±2.62)	83.73 (±1.42)	83.77 (±1.80)	83.70 (±1.91)	84.13 (±1.77)	82.63 (±1.22)	82.95 (±1.92)	82.35 (±1.69)	82.45 (±2.64)
EC-GAN	BrainnetCNN	83.69 (±1.09)	83.08 (±1.86)	84.20 (±1.27)	84.35 (±1.41)	90.56 (±0.94)	90.48 (±1.29)	90.61 (±1.19)	90.24 (±1.53)	88.45 (±1.28)	88.03 (±1.73)	88.82 (±1.42)	89.71 (±1.71)
STGCM		85.09 (±0.88)	84.85 (±1.55)	85.31 (±1.58)	85.41 (±1.15)	93.52 (±0.98)	94.26 (±1.39)	92.96 (±1.17)	93.06 (±1.56)	91.47 (±0.89)	92.62 (±1.15)	90.44 (±1.24)	90.76 (±1.47)
nPSE-GAN		84.83 (±0.76)	84.70 (±0.80)	84.94 (±1.03)	84.80 (±0.74)	92.81 (±0.91)	92.78 (±0.89)	92.84 (±1.35)	92.78 (±0.95)	91.21 (±0.80)	90.70 (±0.84)	91.67 (±1.51)	91.34 (±0.91)
MetaCAE		86.04 (±0.56)	86.17 (±0.80)	85.92 (±0.67)	85.97 (±0.57)	94.08 (±0.80)	93.76 (±1.37)	94.32 (±0.67)	94.07 (±0.82)	91.62 (±0.64)	91.14 (±1.46)	92.06 (±1.67)	91.65 (±0.69)
Ours		**86.58 (±0.44)**	**86.02 (±0.77)**	**87.03 (±0.87)**	**87.11 (±0.86)**	**94.57 (±0.74)**	**94.75 (±1.29)**	**94.44 (±0.64)**	**94.49 (±1.16)**	**92.24 (±0.63)**	92.45 (±0.84)	**92.06 (±0.75)**	**92.45 (±0.58)**

Table 3. Compared with the empirical method, the other nine methods achieve better classification performance in both classifiers by generating BECs. This indicates that effective connectivity contains the causal information that is correlated with MCI. Among the nine BEC-based methods, our model achieves the best mean values for ACC, SEN, SPE, and AUC with 86.58%, 86.02%, 87.03%, and 87.11% for NC vs. EMCI, respectively. The best mean values of ACC, SEN, SPE, and AUC for the NC vs. LMCI task are 94.57%, 94.75%, 94.44%, and 94.49%. The values of 92.24%, 92.45%, 92.06%, and 92.45% are obtained using our model in the EMC vs. LMCI prediction task. The results demonstrate that the generated effective connectivity (our model) gives better classification performance than the functional connectivity (empirical method), indicating its greater ability to capture MCI-related information.

4.4. Effective connectivity analysis

In addition to the fact that brain regions play an important role in disease diagnosis, the causal relationship between them can uncover the underlying pathological mechanism of MCI. In this section, we analyze the generated BECs and predict the abnormal directional connections for further study. To make the results statistically significant, we average all the BECs for each category and filter out the values that fall below the threshold of 0.1. The averaged BEC at three different stages is shown in Fig. 6 by modifying the circularGraph packages.[2] We compute the altered effective connectivity by subtracting the averaged BEC matrix at the early stage from the later stage. As a result, a total of six altered effective connectivity matrices are obtained, consisting of the enhanced and diminished connectivities from NC to EMCI, from NC to LMCI, and from EMCI to LMCI. The altered effective connectivities are shown in Fig. 7. The top row represents the enhanced connections, and the bottom row represents the diminished connections. Each matrix is asymmetric, and the element values range from −0.35∼0.35. The positive value means the directional connection strength is enhanced, while the negative value means the directional connection strength is diminished.

These altered effective connections probably contribute to the cause of MCI. To find the important effective connections during the MCI progression, we sort the altered effective connections and calculate the top 10 enhanced and diminished effective connections. The results are

[2] https://github.com/paul-kassebaum-mathworks/circularGraph.

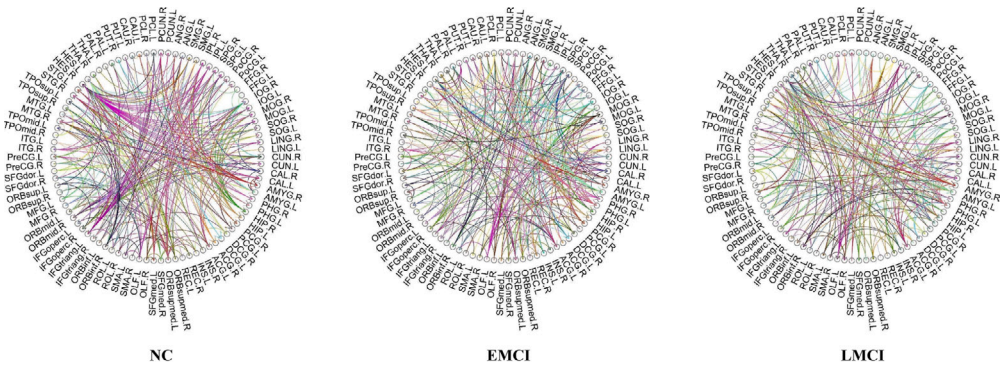

Fig. 6. The generated brain effective connectivities at NC, EMCI, and LMCI, respectively. The arcs with arrows are the effective connection, and the colors have no meaning. The circles on the outside represent the brain regions.

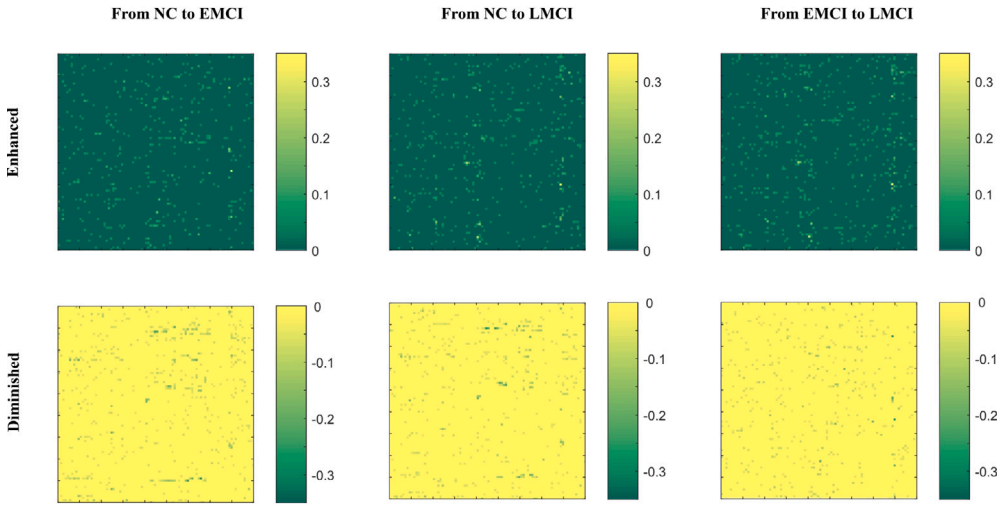

Fig. 7. The altered brain effective connectivities from NC to EMCI, from NC to LMCI, and from EMCI to LMCI, respectively. The top row represents the enhanced connections, and the bottom row represents the diminished connections.

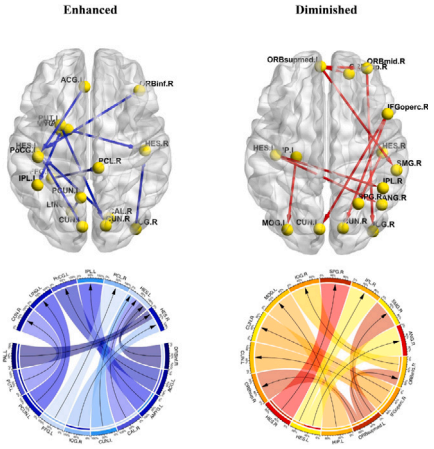

Fig. 8. The top 10 enhanced and diminished effective connections from NC to EMCI.

shown in Figs. 8, 9, and 10. The enhanced and diminished connections are displayed on both the brain atlas view and the circularGraph view. From NC to EMCI, the 10 effective connections with the greatest enhancement in connection strength are ORBinf.R → HES.L, ACG.L → PoCG.L, AMYG.L → IPL.L, CAL.R → HES.L, CUN.L → HES.L, IOG.R → HES.R, FFG.L → PCL.R, PCUN.L → LING.L, PUT.L → CUN.R, PAL.L → HES.R; The 10 effective connections with the greatest diminishment in connection strength are ORBmid.R → IOG.R, IFGoperc.R → CUN.L, IFGoperc.R → CUN.R, ORBsupmed.L → ORBsup.R, ORBsupmed.L → ORBmid.R, ORBsupmed.L → SMG.R, HIP.L → MOG.L, HES.L → IPL.R, HES.L → ANG.R, HES.R → SPG.R. As the EMCI progresses to the LMCI, the top 10 enhanced effective connections are: ROL.L → AMYG.R, SMA.R → HES.L, SOG.R → PCG.R, SOG.R → HIP.L, SPG.R → HES.L, IPL.R → HES.L, THA.L → OLF.R, THA.R → AMYG.R, TPOsup.R → AMYG.L, MTG.R → AMYG.R; and the top 10 diminished effective connections are ORBinf.R → HES.L, OLF.L → SOG.R, ACG.L → PoCG.L, CAL.R → HES.L, CUN.L → HES.L, IOG.R → HES.R, FFG.L → PCL.R, FFG.R → IOG.L, PUT.L → CUN.R, PAL.L → HES.R. The enhanced effective connections between NC and LMCI groups includes ORBmid.L → TPOsup.L, ROL.L → AMYG.R, SMA.R → HES.L, SOG.R → PCG.R, SPG.R → HES.L, PAL.R → HES.L, THA.L → OLF.R, THA.R → AMYG.R, TPOsup.L → ORBmid.R, TPOsup.R → AMYG.L; and the diminished effective connections are ORBmid.R → IOG.R, IFGoperc.R → CUN.L, IFGoperc.R → CUN.R, IFGoperc.R → SOG.L, IFGoperc.R → SOG.R, ORBinf.R → ANG.L, HIP.L → MOG.L, HIP.R → IOG.R, CAL.R → AMYG.L, HES.R → SPG.R. These connection-related ROIs are consistent with the top nine important ROIs identified in the above section.

4.5. Ablation study

In our experiment, the BEC is obtained by optimizing the generator and the discriminator. To investigate whether the proposed generator and discriminator are effective, we design three variants of

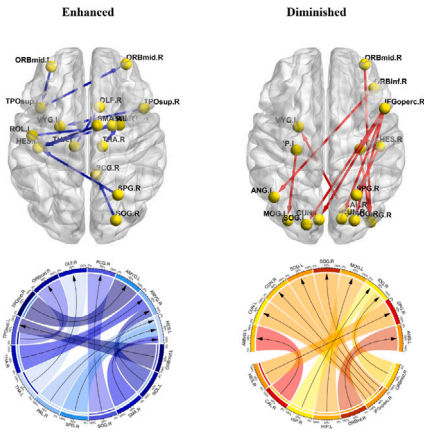

Fig. 9. The top 10 enhanced and diminished effective connections from NC to LMCI.

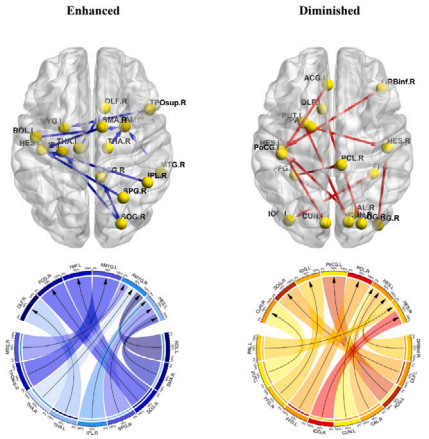

Fig. 10. The top 10 enhanced and diminished effective connections from EMCI to LMCI.

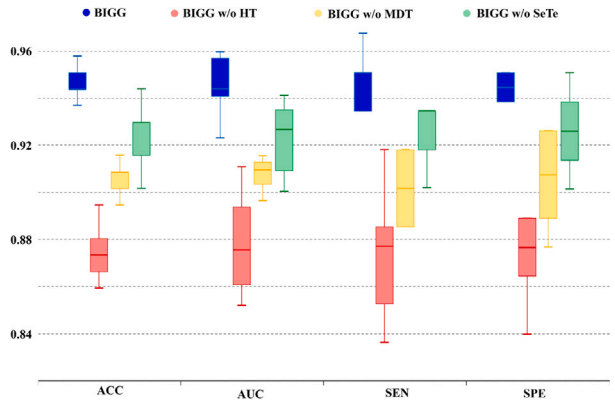

Fig. 11. The impact of the proposed generator and discriminator on the model's performance.

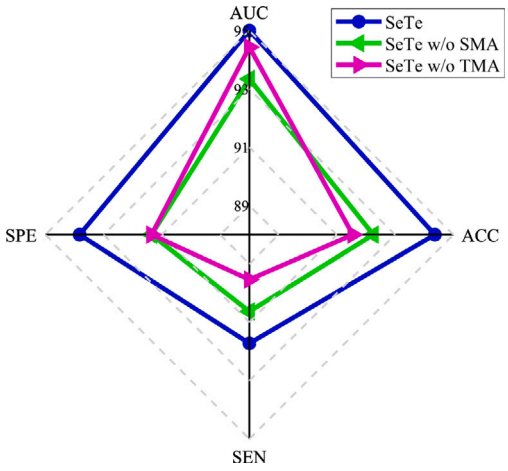

Fig. 12. The influence of SeTe structure on the model's classification performance.

the proposed model and repeat ten times the stratified 5-fold cross-validations. (1) BIGG without hierarchical transformer (BIGG w/o HT). We removed the TDS and TUS and only kept one SeTe block in the generator. (2) BIGG without SeTe blocks (BIGG w/o SeTe). In this case, we replace the SeTe with conventional 1D convolution with a kerner size of 1×3. (3) BIGG without multiresolution diffusive transformer (BIGG w/o MDT). We remove the downsampling operation in the discriminator and keep the D_1. For each variant, we compute the ACCs, AUCs, SENs, and SPEs. The results are shown in Fig. 11. It can be observed that removing the hierarchical structure greatly reduces the classification performance, which shows the effectiveness and necessity of the hierarchical structure for BEC generation. The SeTe block and the discriminator structure also lower the model's classification performance to some extent. Besides, we analyze the impact of the SMA and TMA in the SeTe block. The results are displayed in Fig. 12. Both of these two modules contribute a lot to classification performance. It probably indicates that they capture the spatial and temporal characteristics that are correlated with MCI. Therefore, the SeTe block can effectively integrate the spatial and temporal features during the denoising process.

5. Discussion

The proposed model can generate BEC from 4D fMRI in an end-to-end manner for MCI analysis. The ROI mask of the AAL90 atlas helps to parcellate the whole 3D volume into 90 ROIs at each time point. This operation contains many linear interpolations and ignores voxels at the boundaries of adjacent brain regions, which results in a rough ROI-based time series. To denoise the rough ROI-based time series obtained, we employ the conditional DDPM to successively remove the noise and get a clean ROI-based time series because of its powerful ability to generate high-quality and diverse results. As the denoising process needs thousands of steps to approach clean data, we introduce the adversarial strategy to speed the denoising process. As displayed in Fig. 11, when removing the adversarial strategy (BIGG w/o MDT), the classification performance shows a significant decrease. Also, the hierarchical structure and the SeTe block in the generator both ensure good generation results because they focus on the multi-scale spatiotemporal features and thus enhance the denoising quality. Furthermore, we have analyzed the computational complexity of the brain region scale on our model. Considering the runtime is excessively long with a larger number of regions, we choose to analyze brain regions ranging from 10 to 90 and evaluate the proposed model's computational complexity. As shown in Fig. 13, the number of brain regions (ROIs) has a significant impact on the model complexity. The model with a low-resolution template has low complexity and is relatively easy to train, but some subtle information related to brain disease may be lost. The model with a high-resolution template has high complexity and can mine more detailed information, but it faces many challenges in model training. In clinical applications, it is necessary to choose appropriate brain atlas to balance model complexity and diagnostic performance.

The altered effective connections detected are important for the three scenarios. These connections are partly correlated, which is essential for discovering MCI-related biomarkers. We focus on the same

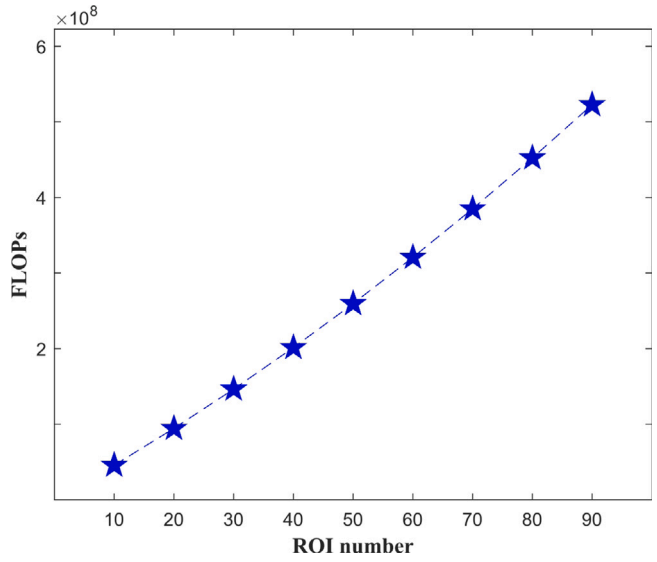

Fig. 13. The computational complexity of the proposed model for different brain region scale.

Table 4
The top twelve important effective connections that are highly correlated with MCI. + means enhanced connection, and − means diminished connection.

Type	Effective connection	Location
+	ORBmid.L→TPOsup.L	Frontal lobe/Temporal lobe
	SOG.R→PCG.R	Occipital lobe/Parietal lobe
	PAL.R→HES.L	Subcortical areas/Temporal lobe
	THA.L→OLF.R	Subcortical areas/Frontal lobe
	THA.R→AMYG.R	Subcortical areas/Temporal lobe
	TPOsup.L→ORBmid.R	Temporal lobe/Frontal lobe
−	IFGoperc.R→SOG.L	Frontal lobe/Occipital lobe
	HIP.L→MOG.L	Temporal lobe/Occipital lobe
	HIP.L→MOG.R	Temporal lobe/Occipital lobe
	HIP.R→IOG.R	Temporal lobe/Occipital lobe
	CAL.R→AMYG.L	Occipital lobe/Temporal lobe
	HES.L→ANG.L	Temporal lobe/Parietal lobe

altered connections between NC vs. EMCI and NC vs. LMCI. The top 12 MCI-related effective connections are shown in Table 4. The enhanced and diminished effective connection-related ROIs are identified in previous studies [41–43]. For example, the HIP has characteristics of decreasing volume and diminishing connection strength as the disease progresses [44,45]. Also, the AMYG has been reported to process both sensory information and punishment/reward-related learning memory [46]. Disruption of AMYG-related connections can bring cognitive decline [47]. The ANG can correlate visual information with language expression. Patients with MCI lose ANG-related connections and cannot read the visual signals [48]. We display two examples for visualizing the effective connection-strength-changing process. As shown in Fig. 14, the effective connection from SOG.R to PCG.R is becoming weaker as NC progresses to LMCI. This perhaps results in the memory problem and is consistent with clinical works [49,50]. The effective connection from HIP.L to MOG.R becomes progressively stronger during disease progression.

The proposed BIGG framework in this paper is modality-independent and can directly map brain imaging to effectively connectivity (imaging-to-graph). This model does not require preprocessing of brain imaging and can directly construct effective connectivity from other brain imaging data, such as EEG and DTI. This framework can not only analyze neurodegenerative diseases but also be applied to analyze other nervous system or psychiatric diseases, such as Parkinson's disease, schizophrenia, or autism spectrum disorder. By constructing brain effective connectivity, the causal connections between brain regions can be used to analyze the pathogenic mechanisms of these diseases. The proposed model still has two limitations, as follows: One is that this work ignores the causal dynamic connections between paired brain regions. The dynamic characteristics are indicative of cognitive and emotional brain activities, which play an important role in detecting the early stage of AD and understanding the pathological mechanisms. In the next work, we will explore the time delay properties of fMRI among brain regions to bridge the dynamic causal connections with biological interpretability. Another is that the input data only concentrates on the fMRI while ignoring other complementary information. Since diffusion tensor imaging (DTI) can characterize the microstructural information, it can enhance the BEC's construction performance and make biological analyses. In the future, we will utilize the GCN and combine it with fMRI to extract complementary information for effective connectivity construction.

6. Conclusion

In this paper, we propose a brain imaging-to-graph generation (BIGG) framework to estimate the brain's effective connectivity from 4-dimensional fMRI in an end-to-end manner. The proposed framework is based on diffusion models to gradually remove noise from a Gaussian sample with several successive denoising steps. Each denoising step is modeled with a transformer-based GAN to translate the noise and conditional fMRI to the clean sample and effective connectivity. The hierarchical structure in the framework can enhance noise removal, and make the generation more high-quality, diverse, and efficient. Results from the ADNI datasets prove the feasibility and efficiency of the proposed model. The proposed model not only achieves superior prediction performance compared with other shallow and deep-learning methods but also identifies MCI-related causal connections for better understanding pathological deterioration and discovering potential MCI biomarkers.

CRediT authorship contribution statement

Qiankun Zuo: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Hao Tian:** Writing – review & editing, Visualization, Supervision, Funding acquisition. **Yudong Zhang:** Writing – review & editing, Visualization, Resources, Project administration. **Jin Hong:** Writing – review & editing, Visualization, Supervision.

Statement

A preprint has previously been published [51].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This paper was partly supported by National Natural Science Foundation of China (62406107), in part by Natural Science Foundation of Hubei Province, China (2023AFB004), in part by Education Department Scientific Research Program Project of Hubei Province of China (Q20232206).

Data availability

Data collection in preparation of this work were funded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/).

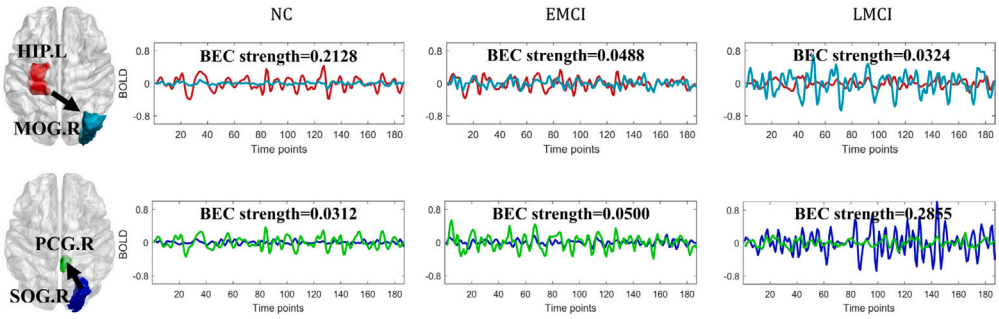

Fig. 14. Visualization of empirical ROI-based BOLD signals from two effective connections. The red arrow is the causal direction. The top row is the diminished connection, and the bottom row is the enhanced connection.

References

[1] A.T. Reid, D.B. Headley, R.D. Mill, R. Sanchez-Romero, L.Q. Uddin, D. Marinazzo, D.J. Lurie, P.A. Valdés-Sosa, S.J. Hanson, B.B. Biswal, et al., Advancing functional connectivity research from association to causation, Nature Neurosci. 22 (11) (2019) 1751–1760.

[2] A. Avena-Koenigsberger, B. Misic, O. Sporns, Communication dynamics in complex brain networks, Nature Rev. Neurosci. 19 (1) (2018) 17–33.

[3] S. Rupprechter, L. Romaniuk, P. Series, Y. Hirose, E. Hawkins, A.-L. Sandu, G.D. Waiter, C.J. McNeil, X. Shen, M.A. Harris, et al., Blunted medial prefrontal cortico-limbic reward-related effective connectivity and depression, Brain 143 (6) (2020) 1946–1956.

[4] M. Mijalkov, G. Volpe, J.B. Pereira, Directed brain connectivity identifies widespread functional network abnormalities in parkinson's disease, Cerebral Cortex 32 (3) (2022) 593–607.

[5] B.M. Hampstead, M. Khoshnoodi, W. Yan, G. Deshpande, K. Sathian, Patterns of effective connectivity during memory encoding and retrieval differ between patients with mild cognitive impairment and healthy older adults, Neuroimage 124 (2016) 997–1008.

[6] S. Sami, N. Williams, L.E. Hughes, T.E. Cope, T. Rittman, I.T. Coyle-Gilchrist, R.N. Henson, J.B. Rowe, Neurophysiological signatures of alzheimer's disease and frontotemporal lobar degeneration: pathology versus phenotype, Brain 141 (8) (2018) 2500–2510.

[7] Y. Shi, H.-I. Suk, Y. Gao, S.-W. Lee, D. Shen, Leveraging coupled interaction for multimodal Alzheimer's disease diagnosis, IEEE Trans. Neural Netw. Learn. Syst. 31 (1) (2019) 186–200.

[8] M. Scherr, L. Utz, M. Tahmasian, L. Pasquini, M.J. Grothe, J.P. Rauschecker, T. Grimmer, A. Drzezga, C. Sorg, V. Riedl, Effective connectivity in the default mode network is distinctively disrupted in Alzheimer's disease-a simultaneous resting-state fdg-pet/fmri study, Hum. Brain Mapp. 42 (13) (2021) 4134–4143.

[9] S. Pei, C. Wang, S. Cao, Z. Lv, Data augmentation for fmri-based functional connectivity and its application to cross-site adhd classification, IEEE Trans. Instrum. Meas. 72 (2022) 1–15.

[10] K.J. Friston, J. Kahan, B. Biswal, A. Razi, A dcm for resting state fmri, Neuroimage 94 (2014) 396–407.

[11] X. Zheng, J.C. Rajapakse, Learning functional structure from fMR images, Neuroimage 31 (4) (2006) 1601–1613.

[12] N. Xu, R.N. Spreng, P.C. Doerschuk, Initial validation for the estimation of resting-state fmri effective connectivity by a generalization of the correlation approach, Front. Neurosci. 11 (2017) 271.

[13] R. Vicente, M. Wibral, M. Lindner, G. Pipa, Transfer entropy—a model-free measure of effective connectivity for the neurosciences, J. Comput. Neurosci. 30 (1) (2011) 45–67.

[14] A. Al-Ezzi, N. Yahya, N. Kamel, I. Faye, K. Alsaih, E. Gunaseli, Severity assessment of social anxiety disorder using deep learning models on brain effective connectivity, IEEE Access 9 (2021) 86899–86913.

[15] S. Bagherzadeh, M.S. Shahabi, A. Shalbaf, Detection of schizophrenia using hybrid of deep learning and brain effective connectivity image from electroencephalogram signal, Comput. Biol. Med. 146 (2022) 105570.

[16] J. Liu, J. Ji, G. Xun, A. Zhang, Inferring effective connectivity networks from fmri time series with a temporal entropy-score, IEEE Trans. Neural Netw. Learn. Syst. 33 (10) (2021) 5993–6006.

[17] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144.

[18] Q. Zuo, Y. Shen, N. Zhong, C.L.P. Chen, B. Lei, S. Wang, Alzheimer's disease prediction via brain structural-functional deep fusing network, IEEE Trans. Neural Syst. Rehabil. Eng. 31 (2023) 4601–4612.

[19] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, Adv. Neural Inf. Process. Syst. 33 (2020) 6840–6851.

[20] Q. Zuo, L. Chen, Y. Shen, M.K.-P. Ng, B. Lei, S. Wang, BDHT: Generative AI enables causality analysis for mild cognitive impairment, IEEE Trans. Autom. Sci. Eng. (2024) 1–13.

[21] H.-J. Park, K.J. Friston, C. Pae, B. Park, A. Razi, Dynamic effective connectivity in resting state fmri, NeuroImage 180 (2018) 594–608.

[22] A.K. Seth, A matlab toolbox for granger causal connectivity analysis, J. Neurosci. Methods 186 (2) (2010) 262–273.

[23] A.K. Seth, A.B. Barrett, L. Barnett, Granger causality analysis in neuroscience and neuroimaging, J. Neurosci. 35 (8) (2015) 3293–3297.

[24] A.M. DSouza, A.Z. Abidin, L. Leistritz, A. Wismüller, Exploring connectivity with large-scale granger causality on resting-state functional mri, J. Neurosci. Methods 287 (2017) 68–79.

[25] N. Talebi, A.M. Nasrabadi, I. Mohammad-Rezazadeh, R. Coben, Ncreann: Nonlinear causal relationship estimation by artificial neural network; applied for autism connectivity study, IEEE Trans. Med. Imaging 38 (12) (2019) 2883–2890.

[26] Z. Abbasvandi, A.M. Nasrabadi, A self-organized recurrent neural network for estimating the effective connectivity and its application to eeg data, Comput. Biol. Med. 110 (2019) 93–107.

[27] J. Liu, J. Ji, G. Xun, L. Yao, M. Huai, A. Zhang, Ec-gan: inferring brain effective connectivity via generative adversarial networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 4852–4859.

[28] J. Ji, J. Liu, L. Han, F. Wang, Estimating effective connectivity by recurrent generative adversarial networks, IEEE Trans. Med. Imaging 40 (12) (2021) 3326–3336.

[29] A. Zou, J. Ji, M. Lei, J. Liu, Y. Song, Exploring brain effective connectivity networks through spatiotemporal graph convolutional models, IEEE Trans. Neural Netw. Learn. Syst. (2022).

[30] Z. Zhang, Z. Zhang, J. Ji, J. Liu, Amortization transformer for brain effective connectivity estimation from fMRI data, Brain Sci. 13 (7) (2023) 995.

[31] J. Ji, L. Han, F. Wang, J. Liu, Dynamic effective connectivity learning based on nonparametric state estimation and GAN, IEEE Trans. Instrum. Meas. 73 (2024) 1C12.

[32] J. Liu, L. Han, J. Ji, MCAN: Multimodal causal adversarial networks for dynamic effective connectivity learning from fMRI and EEG data, IEEE Trans. Med. Imaging (2024) early access.

[33] J. Ji, Z. Zhang, L. Han, J. Liu, MetaCAE: Causal autoencoder with meta-knowledge transfer for brain effective connectivity estimation, Comput. Biol. Med. 170 (2024) 107940.

[34] M. Mirza, S. Osindero, Conditional generative adversarial nets, 2014, arXiv preprint arXiv:1411.1784.

[35] Y. Wang, B. Yu, L. Wang, C. Zu, D.S. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, L. Zhou, 3D conditional generative adversarial networks for high-quality pet image estimation at low dose, Neuroimage 174 (2018) 550–562.

[36] O. Özdenizci, R. Legenstein, Restoring vision in adverse weather conditions with patch-based denoising diffusion models, IEEE Trans. Pattern Anal. Mach. Intell. (2023).

[37] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain, Neuroimage 15 (1) (2002) 273–289.

[38] J. Wang, X. Wang, M. Xia, X. Liao, A. Evans, Y. He, Gretna: a graph theoretical network analysis toolbox for imaging connectomics, Front. Hum. Neurosci. 9 (2015) 386.

[39] Q. Zuo, H. Wu, C.P. Chen, B. Lei, S. Wang, Prior-guided adversarial learning with hypergraph for predicting abnormal connections in alzheimer's disease, IEEE Trans. Cybern. 54 (6) (2024) 3652–3665.

[40] J. Kawahara, C.J. Brown, S.P. Miller, B.G. Booth, V. Chau, R.E. Grunau, J.G. Zwicker, G. Hamarneh, Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment, NeuroImage 146 (2017) 1038–1049.

[41] S.-Y. Lin, C.-P. Lin, T.-J. Hsieh, C.-F. Lin, S.-H. Chen, Y.-P. Chao, Y.-S. Chen, C.-C. Hsu, L.-W. Kuo, Multiparametric graph theoretical analysis reveals altered structural and functional network topology in Alzheimer's disease, NeuroImage: Clin. 22 (2019) 101680.

[42] B. Lei, N. Cheng, A.F. Frangi, E.-L. Tan, J. Cao, P. Yang, A. Elazab, J. Du, Y. Xu, T. Wang, Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease, Med. Image Anal. 61 (2020) 101652.

[43] B. Chen, Abnormal cortical regions and subsystems in whole brain functional connectivity of mild cognitive impairment and alzheimer's disease: a preliminary study, Aging Clin. Exp. Res. 33 (2021) 367–381.

[44] N. Schuff, N. Woerner, L. Boreta, T. Kornfield, L. Shaw, J. Trojanowski, P. Thompson, C. Jack Jr., M. Weiner, A.D.N. Initiative, Mri of hippocampal volume loss in early Alzheimer's disease in relation to apoe genotype and biomarkers, Brain 132 (4) (2009) 1067–1077.

[45] M. Tahmasian, L. Pasquini, M. Scherr, C. Meng, S. Förster, S.M. Bratec, K. Shi, I. Yakushev, M. Schwaiger, T. Grimmer, et al., The lower hippocampus global connectivity, the higher its local metabolism in Alzheimer disease, Neurology 84 (19) (2015) 1956–1963.

[46] T. Yang, K. Yu, X. Zhang, X. Xiao, X. Chen, Y. Fu, B. Li, Plastic and stimulus-specific coding of salient events in the central amygdala, Nature (2023) 1–10.

[47] M. Ortner, L. Pasquini, M. Barat, P. Alexopoulos, T. Grimmer, S. Förster, J. Diehl-Schmid, A. Kurz, H. Förstl, C. Zimmer, et al., Progressively disrupted intrinsic functional connectivity of basolateral amygdala in very early Alzheimer's disease, Front. Neurol. 7 (2016) 132.

[48] E.-S. Lee, K. Yoo, Y.-B. Lee, J. Chung, J.-E. Lim, B. Yoon, Y. Jeong, Default mode network functional connectivity in early and late mild cognitive impairment, Alzheimer Dis. Assoc. Disord. 30 (4) (2016) 289–296.

[49] J. Xue, H. Guo, Y. Gao, X. Wang, H. Cui, Z. Chen, B. Wang, J. Xiang, Altered directed functional connectivity of the hippocampus in mild cognitive impairment and Alzheimer's disease: a resting-state fmri study, Front. Aging Neurosci. 11 (2019) 326.

[50] D. Berron, D. van Westen, R. Ossenkoppele, O. Strandberg, O. Hansson, Medial temporal lobe connectivity and its associations with cognition in early Alzheimer's disease, Brain 143 (4) (2020) 1233–1248.

[51] Q. Zuo, H. Tian, C.P. Pun, H. Wang, Y. Zhang, J. Hong, Brain imaging-to-graph generation using adversarial hierarchical diffusion models for MCI causality analysis, 2023, arXiv.