Estimating Effective Connectivity by Recurrent Generative Adversarial Networks

Junzhong Ji, Jinduo Liu, Lu Han, and Feipeng Wang

***Abstract*— Estimating effective connectivity from functional magnetic resonance imaging (fMRI) time series data has become a very hot topic in neuroinformatics and brain informatics. However, it is hard for the current methods to accurately estimate the effective connectivity due to the high noise and small sample size of fMRI data. In this paper, we propose a novel framework for estimating effective connectivity based on recurrent generative adversarial networks, called EC-RGAN. The proposed framework employs the generator that consists of a set of effective connectivity generators based on recurrent neural networks to generate the fMRI time series of each brain region, and uses the discriminator to distinguish between the joint distributions of the real and generated fMRI time series. When the model is well-trained and generated fMRI data is similar to real fMRI data, EC-RGAN outputs the effective connectivity by means of the causal parameters of the effective connectivity generators. Experimental results on both simulated and real-world fMRI time series data demonstrate the efficacy of our proposed framework.**

***Index Terms*— Effective connectivity, generative adversarial networks, recurrent neural networks, fMRI time series.**

I. Introduction

IN RECENT years, there has been an increasing interest in brain network analysis. One of the main end goals of brain network analysis is to determine the set of causal relations that describe the effective connectivity (EC) within a set of brain regions from neuroimaging data, e.g., functional magnetic resonance imaging (fMRI) data. EC characterizes the neural influence between two brain regions [1], and its impairment is associated with some brain diseases, e.g., Alzheimer's disease (AD) [2], [3], Parkinson's disease (PD) [4], [5], and Schizophrenia [6].

With the continuous development in the field of machine learning theory, lots of computational methods have been developed to estimate the causal relations between variables [7]–[10]. In nature, estimating effective connectivity can be represented as a problem of learning the causal relationships between brain regions from fMRI data [11], [12]. More precisely, a brain effective connectivity network can be represented by a causal graph (directed graph) where nodes denote brain regions, and the directed arcs denote effective connectivity between brain regions [13].

In the last decade, more and more methods have been widely used to estimate the effective connectivity involved in human brain [8]–[28]. In general, these methods can be categorized into two types, i.e., the model-driven method and the data-driven method. The model-driven approaches include the dynamic causal model (DCM) [14] and the structural equation model (SEM) [15], [16], which have been applied extensively for estimating effective connectivity. However, this kind of method requires prior assumptions on the models and is commonly used to construct relatively small-scale networks [11].

Data-driven methods include Bayesian network (BN) methods [17]–[20], Granger causality (GC) methods [21], [22], and Linear non-Gaussian acyclic model (LiNGAM) methods [23], [24], which can directly estimate effective connectivity from data without any prior knowledge. BN methods search for effective connectivity under the assumption that the true effective connectivity network forms a directed acyclic graph (DAG), which cannot model cyclic or bidirectional connections of the effective connectivity networks. Granger causality (GC) methods infer effective connectivity in fMRI time series by the multiple regression of time-indexed variables on lagged values of variables and require the time series to be wide-sense stationary and have a zero mean. LiNGAM uses the assumption of non-normality to determine the unique set of directed paths that best describe the given subject's data set. However, LiNGAM requires a large number of data points, so it performs poorly when the fMRI data sample is small. Current studies show that data-driven methods are fairly efficient at determining the presence of an association between variables, but less effective at orienting the association into a causal relation. In summary, cyclic or bidirectional structure,

Manuscript received February 9, 2021; revised April 28, 2021; accepted May 19, 2021. Date of publication May 26, 2021; date of current version November 30, 2021. This work was supported in part by the NSFC Research Program under Grant 61672065, in part by the Beijing Postdoctoral Research Foundation, in part by the Fundamental Research Funds of the Beijing University of Technology, in part by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health) under Grant U01 AG024904, and in part by the Department of Defense (DOD) ADNI under Award W81XWH-12-2-0012. *(Corresponding authors: Junzhong Ji; Jinduo Liu.)*

The authors are with the Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology, Beijing Artificial Intelligence Institute, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: jjz01@bjut.edu.cn; jinduo@bjut.edu.cn; hanlu1@emails.bjut.edu.cn; wangfeipeng1998@emails.bjut.edu.cn).

Digital Object Identifier 10.1109/TMI.2021.3083984

high noise, and small samples are main factors that affect the performance of this kind of methods.

In 2020, Liu *et al.* presented a generative adversarial networks model [27] to infer the effective connectivity from fMRI data (EC-GAN) [28]. The EC-GAN employs effective connectivity generators to generate the fMRI time series of each brain region, and employs a discriminator to measures the difference between generated fMRI data and real fMRI data. The experimental results show that EC-GAN performs well under bidirectional structure and high noise. However, EC-GAN only considers the distribution properties of fMRI data, and ignores the temporal information. Previous studies have shown that temporal information is very important for fMRI data analysis. Therefore, exploring novel models that incorporate the temporal information with EC-GAN to estimate effective connectivity from fMRI time series data may further improve the accuracy of the model.

In this paper, we further extend our previous work [28] to estimate effective connectivity via recurrent neural networks (RNN) [29] and generative adversarial networks (GAN), called EC-RGAN. The new method employs the RNN to capture the temporal information from the fMRI time series, which improves the authenticity of the generated data. When the generated fMRI data is similar to real fMRI data, then EC-RGAN can get effective connectivity from the causal parameters of the effective connectivity generators. The experimental results on both simulated fMRI data and real-world fMRI data show that EC-RGAN is more effective for estimating effective connectivity than our previous model and other state-of-the-art methods.

In a nutshell, the main contributions of this paper, compared with our preliminary model [28], are as follows:

- To capture the temporal information from fMRI time series, the proposed model applies RNN in the effective connectivity analysis of fMRI data.
- To accurately estimate the effective connectivity from the high noise and small sample size fMRI data, the proposed model employs the RNN and SEM as the effective connectivity generators.
- Systematic experiments show that EC-RGAN achieves better performance compared with other state-of-the-art methods and our preliminary model EC-GAN.

II. Methodology

In this section, we put forward our novel model, i.e., EC-RGAN to estimate effective connectivity from fMRI time series data. We first show the notation and problem formulation, and then give an overview of the proposed EC-RGAN, last describe the details of the main components.

A. Notation and Problem Formulation

In this paper, we employ capital letter X_i to indicate a variable (brain region), and i is the label of X_i. The bold letter $\mathbf{X}_i$ represents the time series of X_i, and $\mathbf{X}$ denotes a collection of fMRI time series with N brain regions and T length, where $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N) \in \mathbb{R}^{N\times T}$. $\Pi(X_i)$ represents the parent nodes of variable X_i. $X_i \not\perp\!\!\!\perp X_j$ and $X_i \perp\!\!\!\perp X_j$ indicate dependence and independence between the two corresponding variables X_i and X_j, respectively.

Next, we describe the problem setting considered in this paper. Let $\mathcal{G}$ represent a directed graph, $\mathcal{D}$ represent the fMRI time series data, and $\mathcal{P}$ represent the distribution of the fMRI data. Then, a brain effective connectivity network can be expressed as a directed graph $\mathcal{G} =< \mathbf{V}, \mathbf{E} >$, where $\mathbf{V}$ is a set of nodes with each node $X_i \in \mathbf{V}$ indicating a brain region or region of interest (ROI); and $\mathbf{E}$ is a set of arcs with each arc $X_i \to X_j \in \mathbf{E}$ representing the effective connectivity.

B. EC-RGAN Architecture

The EC-RGAN is made up of a generator (G) and a discriminator (D). The generator takes noise variable and real fMRI time series data as input, and generates the samples which are similar to the realistic fMRI time series data. The discriminator takes the real fMRI time series data and samples generated by the generator as inputs and tries to find a mapping that tells us the input data's probability of being real. The structure of the proposed EC-RGAN is shown in Fig. 1.

From Fig.1 we can find that EC-RGAN takes the fMRI time series as input and adopts a generator to generate samples. In detail, the generator is designed based on the RNN and SEM, where RNN is used to capture the temporal information and SEM is employed to measure the causal relations between brain regions. If the samples generated by the generator are very similar to the real input data, then we can get the effective connectivity via the causal parameters from the part of SEM. The network structure of the discriminator is a decoder. In other words, the discriminator is also made up of one-layer RNN and a fully connected layer (without SEM model). The task of the discriminator is to distinguish between generated samples and real samples. The output of the discriminator is a probability that indicates the degree of authenticity.

C. Effective Connectivity Generators

To estimate effective connectivity and learn the statistical and temporal information of the original fMRI time series data, we develop a set of effective connectivity generators as a generator based on the RNN and SEM.

Given a fMRI time series with N brain regions X_i ($i = 1, \ldots, N$) and T length, the input data $\mathcal{D}$ can be represented as:

$$\mathcal{D} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N)^\top \in \mathbb{R}^{T\times N}. \quad (1)$$

For the data $\mathcal{D}$, we can estimate the effective connectivity by SEM model in the following way:

$$X_i = \sum_{X_j \in \Pi(X_i)} F_{X_j, X_i}(X_j) + \varepsilon_i, \quad for\ i = 1, \ldots, N, \quad (2)$$

where $\Pi(X_i)$ denotes the set of parent nodes of brain region X_i, F_{X_j,X_i} is the strength of effective connectivity ($X_j \to X_i$) between brain regions X_i and X_j, and $\varepsilon_1, \ldots, \varepsilon_N$ are random noise in nodes X_i ($i = 1, \ldots, N$), which are mutually independent. In detail, the main idea of SEM is to express every ROI time series ($\mathbf{X}_i$) in a network by a linear combination ($y = f(x) + b$) of all the time series ($\mathcal{D}$), and the effective

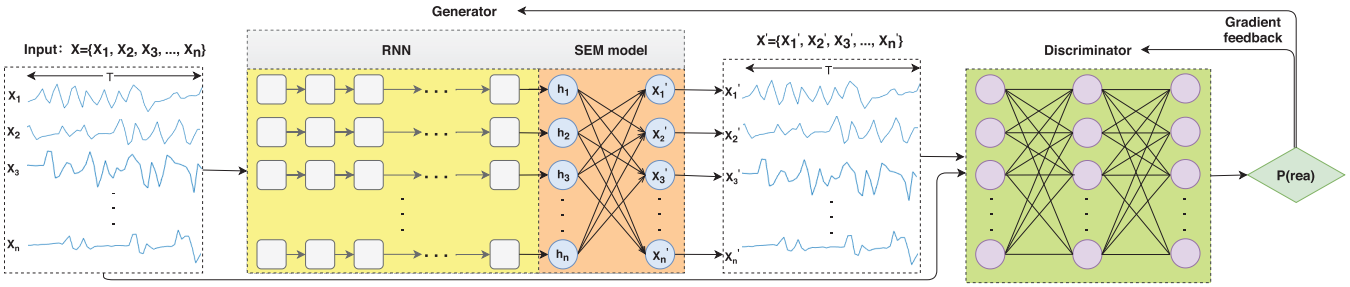


Fig. 1. The structure of EC-RGAN.

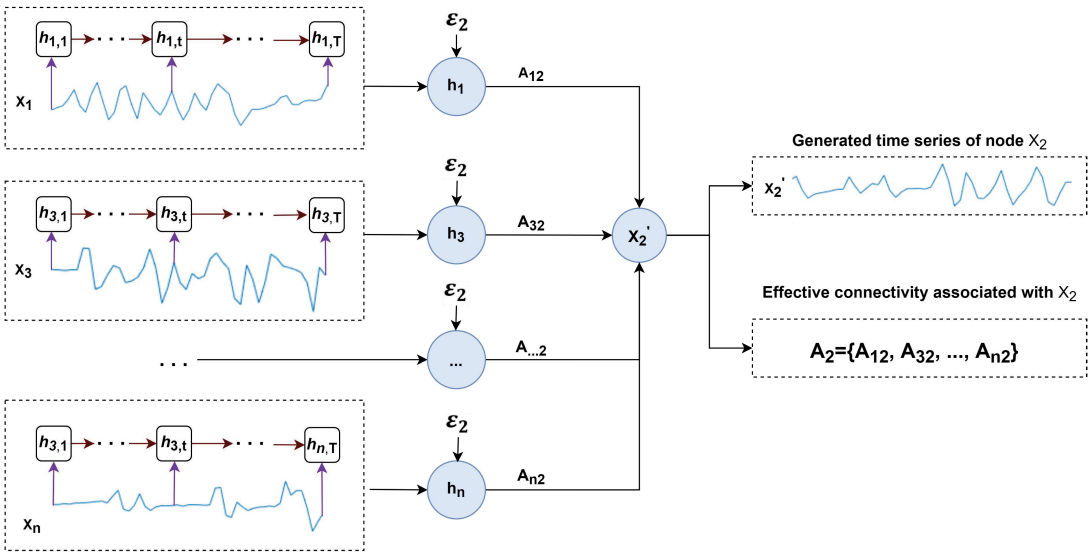


Fig. 2. An example of estimating effective connectivity by an effective connectivity generator. The input of the effective connectivity generator is the fMRI time series of brain regions, and the output is the causal parameters and the generated fMRI time series.

connectivity between underlying neuronal activity ($X_j \to X_i$) can be identified by analyzing the observed fMRI data ($\mathbf{X}_i$ and $\mathbf{X}_j$). To estimate the causal parameters of SEM, the common strategy is using the maximum likelihood algorithm to fit the model based on the covariance of observed variables. However, SEM is a linear model that cannot fit well with fMRI data by iteratively modifying the model parameters (the strength of effective connectivity). To address this issue, we employ RNN to capture the temporal information from fMRI time series data to help the SEM model determinate the model parameters. In this way, the F_{X_i,X_j} can more realistic response to the effective connectivity between two brain regions.

In detail, given data $\mathcal{D}$ and the noise variable ε, we can estimate the time series of each brain regions:

$$\begin{aligned}\hat{\mathbf{X}}_i &= \sum_{X_j \in \Pi(X_i)} F_{X_j,X_i}(\mathbf{X}_j) + \varepsilon_i, \\ &= \sum_{j=1}^{N} A_{ji}\mathbf{X}_j + \varepsilon_i, \end{aligned} \tag{3}$$

where A_{ji} is the causal parameter of the effective connectivity $X_j \to X_i$ from brain region X_j to X_i, thus A is a matrix of the causal parameters (effective connectivity strength). In other words, we can generate a time series of brain region X_i by its parents ($\Pi(X_i)$) time series and the strength of effective connectivity between X_i and $\Pi(X_i)$. In particular, $F_{X_j,X_i} = 0$ means $X_j \perp\!\!\!\perp X_i$ and there is no directed arcs from X_j to X_i (i.e., $X_j \nrightarrow X_i$). Besides, there is no meaningful that one brain region is generated by itself, thus we do not consider the effective connectivity of a brain region itself, which means that if $i = j$ then $F_{X_j,X_i} = 0$.

We use the following example to further illustrate the structure of the effective connectivity generator: suppose we have N brain regions and we want to estimate the effective connectivity between X_2 and other brain regions. Fig.2 illustrates the structure of the effective connectivity generator of the brain region X_2.

For the RNN model in Fig. 2, we employ the Long Short-Term Memory (LSTM), which can relieve the problem of vanishing and exploding gradients of the normal RNN. The core of the LSTM architecture is the memory cell, which is controlled by three different gates: input gate, forget gate, and output gate. In detail, the input gate determines the information to be updated, the forget gate decides whether to retain or discard the previous state, and the output gate is to control how much information the LSTM wants to output at the current time step. Through the gates, the cell of LSTM can store and update long-term memory. Given the T length of fMRI time series $\{x_1, x_2, \ldots, x_t, \ldots, x_T\}$, the operation of

gates and cell update at time t are as follows:

$$\tilde{C}_t = tanh(W_c \times (h_{t-1}, x_t) + b_c), \tag{4}$$
$$i_t = \sigma(W_i \times (h_{t-1}, x_t) + b_i), \tag{5}$$
$$f_t = \sigma(W_f \times (h_{t-1}, x_t) + b_f), \tag{6}$$
$$o_t = \sigma(W_o \times (h_{t-1}, x_t) + b_o), \tag{7}$$
$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{C}_t, \tag{8}$$
$$h_t = o_t \cdot tanh(c_t), \tag{9}$$

where i_t, f_t, o_t, c_t denote the states of the corresponding gates and cells at time t, x_t represents the input at time t, h_{t-1} and h_t indicate the output at time $t-1$ and t, respectively. $\tilde{C}_t$ is a temporary sate, and σ is the sigmoid activation function. W and b are the weight matrix and bias vector, respectively.

As shown in Fig. 2, the generator takes the original fMRI time series (without X_2) as input, and employs the RNN to estimate the time series of X_2. When the model is well-trained, and the generated samples $\mathbf{X}_{gen}$ of X_2 are very similar to real data $\mathbf{X}_{real}$ of X_2. We can get the causal parameters $(F_{X_i,X_2}, i = 1, 3, \ldots, N)$ from the effective connectivity generator. Finally, we can get the strength of effective connectivity by the causal parameters.

In a word, we adopt a set of effective connectivity generators as a generator to generate samples (the number of effective connectivity generators is the same as the number of brain regions). Each effective connectivity generator is employed to generate the fMRI time series of one brain region based on the causal parameters between one brain region and another brain region. If the samples generated by the generator are very similar to the real input data, then we can get the strength of effective connectivity.

D. Regularization for Sparsity

To overcome overfitting and infer sparse effective connectivity networks (causal graphs), we present a new loss function by adding a sparsity penalty to the effective connectivity generators. First, we define the sparsity penalty as:

$$L_p = \frac{\lambda}{2} \log T \, \|A\|, \tag{10}$$

where T is the length of time series, λ is the hyper-parameter that controls the sparsity, and $\|A\|$ denotes the network complexity, that is, the sum of causal parameters A for the effective connectivity network and is calculated as:

$$\|A\| = \sum_{i=1,j=1}^{N} A_{ij}, \tag{11}$$

where N is the number of nodes (brain regions). In particular, if $i = j$, we do not consider the effective connectivity for one brain region itself, and set $A_{ij} = 0$.

E. Loss Function

Finally, we define EC-RGAN Loss Function as:

$$\begin{aligned} &\min_G \max_D V(G, D) \\ &= \mathbb{E}_{\mathbf{X}\sim\mathcal{P}_{\mathcal{D}}(\mathbf{X})}[\log D(\mathbf{X})] \\ &\quad + \sum_{i=1}^{N} \mathbb{E}_{\tilde{\mathbf{X}}_i\sim\mathcal{P}_{\mathcal{D}}(\tilde{\mathbf{X}}_i),\ \varepsilon\sim\mathcal{P}(\varepsilon)}[\log(1 - D(G_i(\tilde{\mathbf{X}}_i, \varepsilon)))] \\ &\quad + L_p, \end{aligned} \tag{12}$$

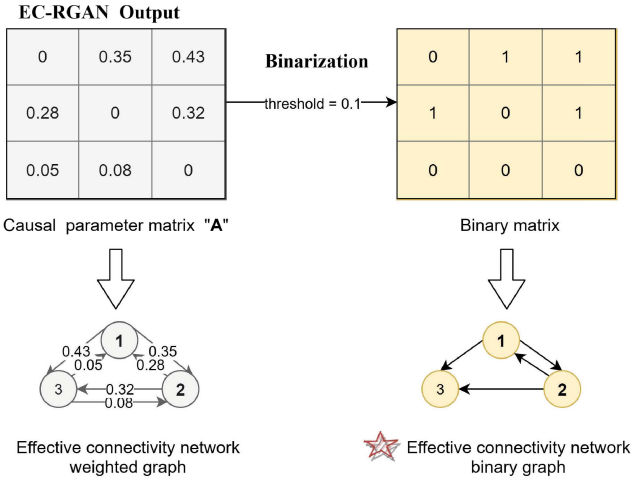


Fig. 3. The detailed process of network binarization (the threshold is 0.1).

where n is the number of brain regions, and $\tilde{\mathbf{X}}_i$ is the subset of real fMRI time series set $\mathbf{X}$ without $\mathbf{X}_i$ ($\tilde{\mathbf{X}}_i = \mathbf{X} \backslash \mathbf{X}_i$).

As the number of effective connectivity generators is the same as the number of brain regions, the hyper-parameters of EC-RGAN are decided by the number of brain regions. Therefore, when we employ EC-RGAN to estimate effective connectivity in practice, we should generate the synthetic fMRI data with the known ground-truth of the effective connectivity network to select the hyper-parameters with the best performance. When the hyper-parameters are determined by the synaptic data, we can employ the EC-RGAN to estimate the effective connectivity from fMRI time series data.

F. Network Binarization

The binary processing is necessary when we want to compare the inferred effective connectivity networks with the binarization ground-truth networks. It is worth noting that this step is optional, which can be regarded as post-processing. In detail, if the causal parameter A_{ij} of an arc $(X_i \to X_j)$ is large than the threshold, then we set it to 1, otherwise, set it to 0. The threshold is usually selected from the generated data set with the ground-truth network, and then used in real data set. To clearly illustrate the processing of the network binarization, we give a detailed description in Fig. 3.

III. Experiments

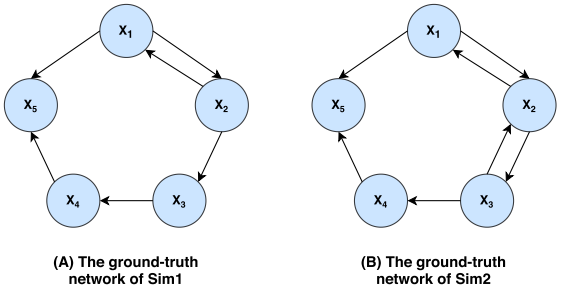
In this section, we first compare the proposed method with our previous model EC-GAN to assess the performance of EC-RGAN. And then, we compare EC-RGAN with other state-of-the-art methods on simulated fMRI data generated from known ground-truth networks. Finally, to illustrate the application potential of EC-RGAN, we apply it to the real fMRI data.

A. Data Description

1) Benchmark Simulated Data: The benchmark simulated data we used are supported by Sanchez-Romero *et al.* (2019) [19], which are widely used for detecting methods' performance on estimating effective connectivity. In our experiments, we run each method on a single subject to testing the performance of different methods on small sample data.

TABLE I
DESCRIPTION OF THE BENCHMARK SIMULATION DATA

Dataset	Nodes	Sessions (s)	TR (s)	Number of Subjects
Sim1	5	600	1.2	60
Sim2	5	600	1.2	60



(A) The ground-truth network of Sim1 (B) The ground-truth network of Sim2

Fig. 4. The ground-truth networks of the two simulated data sets.

TABLE II
THE ROIS OF THE REAL TASK FMRI DATA

NO.	ROI	Detailed description
1	LOCC	Left Occipital Cortex
2	LACC	Left Anterior Cingulate Cortex
3	LIFG	Left Inferior Frontal Gyrus
4	LIPL	Left Inferior Parietal
5	ROCC	Right Occipital Cortex
6	RACC	Right Anterior Cingulate Cortex
7	RIFG	Right Inferior Frontal Gyrus
8	RIPL	Right Inferior Parietal
9	IV	Input Variable

The Sanchez simulated fMRI data contain different bidirectional structures with different degrees of complexity. In our experiments, We chose two simulations from Sanchez data sets, which have one bidirectional arc and two bidirectional arcs, respectively. The data is available at https://github.com/cabal-cmu/feedback-discovery, and its detailed description is shown in Table I. In detail, the ground-truth networks of two simulations are shown in Fig. 4.

2) Real Task fMRI Data: We employ a set of real task fMRI dataset [19] to test the performance of the EC-RGAN. The real task fMRI data is acquired with a 3T scanner, TR = 2s, the number of data points is 160, and the number of subjects is 9. The data is available at the OpenNeuro project (https://openneuro.org/datasets/ds000003/versions/00001). In the experiments, we employ eight brain regions and one input variable as the ROIs. The detailed information of ROIs is shown in Table II.

3) Real Alzheimer's Disease Data: Real data used in this article is obtained from the ADNI database (adni.loni.usc.edu). In this study, we employ all the healthy controls (HC) and AD patients available with fMRI data (53 HC subjects and 66 AD patients). The data is acquired using a 3T MRI scanner (Siemens) and an 8-channel receive only head coil. fMRI sequence parameters include: Slices = 48; volumes = 140; TR/TE = 3000/30ms; FA = 90°; Matrix = 64×64. For more data information, please see www.adni-info.org.

fMRI data preprocessing is performed using the Data Processing Assistant for Resting-State fMRI (DPARSF, http://www.restfmri.net), which is based on Statistical Parametric Mapping (SPM8, http://www.fil.ion.ucl.ac.uk/spm) and Resting-State fMRI Data Analysis Toolkit (REST).

TABLE III
THE PARAMETER CONFIGURATIONS OF THE SIX METHODS

Algorithm	Category	Parameters
GIMME	SEM-based	groupcutoff=0.6, subcutoff=0.6
PC	BN-based	α=0.1
GES	BN-based	penalty=2.4
AIAEC	BN-based	N=80,P_s=0.5, P_c=0.6 P_m=0.4, M=70, l=10
MVARp	GC-based	α=0.001, nperm=1000
PcLiNGAM	LiNGAM-based	method=1, threshold = 0.4

B. Evaluation Metrics

We compared the learned result to ground-truth networks on the three most common graph metrics: 1) Precision, 2) Recall, and 3) F1-measure (F1). Let L_n express the learned network and G_n denote the ground-truth network. The three evaluation metrics are then given by:

$$Precision = \frac{SA}{T_{L_n}}, \tag{13}$$

$$Recall = \frac{SA}{T_{G_n}}, \tag{14}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{15}$$

where SA represents the number of same directed arcs which are both in L_n and G_n; T_{L_n} and T_{G_n} respectively denote the total number of directed arcs in L_n and G_n. In our experiments, the bidirectional arc is equal to two arcs (e.g., $X_i \leftrightarrow X_j$ means $X_i \rightarrow X_j$ and $X_j \rightarrow X_i$).

C. Baseline Methods

To intuitively show the competitiveness of the EC-RGAN, we compare EC-RGAN with the other seven methods, some of them are classical methods, and some of them are state-of-the-art methods. These methods include: GIMME [16], PC [20], GES [20], AIAEC [19], MVARp [22], and PcLiNGAM [24], respectively.

The parameters of the algorithms under comparison are selected according to the existed literature [16]–[24]. For a more fair comparison, we also randomly select 10 subjects from the 60 subjects (as training data) to select the best parameters for the seven comparison methods. The parameter configurations of the corresponding methods are shown in Table III.

D. Comparative Evaluations With EC-GAN

Before conducting the comparative experiments, we first generated some simulated data with known ground-truth network (as the training dataset) to optimize the hyper-parameters of the EC-RGAN. The generation model of simulations is referenced to the method in [8], which used the dynamic causal modeling (DCM) to generate the fMRI time series data. For a five nodes fMRI time series data, the hyper-parameters of EC-GAN and EC-RGAN are set as: the learning rate of

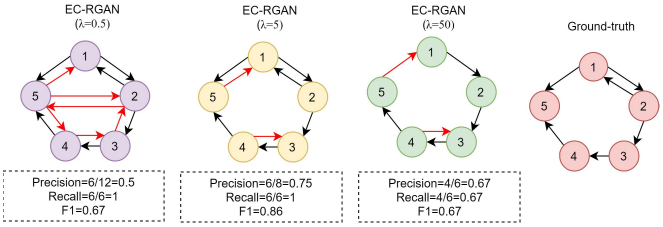


Fig. 5. The effect of the sparsity parameter λ. The black arrow indicates that the direction of effective connectivity is consistent with the ground-truth network, while the red arrow indicates that it is inconsistent.

TABLE IV
THE RESULTS OF EC-GAN AND EC-RGAN ON SIMULATED DATA

Data	Metrics	Methods	
		EC-GAN	EC-RGAN
Sim1	Precision	0.67±0.10	0.70±0.12
	Recall	0.87±0.14	0.87±0.12
	F1	0.75±0.08	0.77±0.09
Sim2	Precision	0.77±0.12	0.79±0.11
	Recall	0.82±0.14	0.83±0.15
	F1	0.79±0.09	0.80±0.09

generator and discriminator are 0.01, the number of units m is 100, the number of effective connectivity generators n is 5, sparsity parameter λ is 5.

As the sparsity parameter is an important parameter, we give an example to show how to tune it. We first run EC-RGAN on the generated data and change the value of the sparsity parameter λ. Then we compare the networks (estimated by EC-RGAN with different λ) to the ground-truth network and choose the sparsity parameter λ whose F-1 value is the highest. Fig. 5 shows the effect of the sparsity parameter λ on five nodes generated data.

Then we employ the benchmark simulated dataset as the test dataset to evaluate the performance of EC-RGAN and other methods. The results of EC-GAN and EC-RGAN on simulated data are shown in Table IV. Compared to EC-GAN, we can find that the new method EC-RGAN has an improvement in Precision and F1.

To clearly show the performance and further analyze the advantages of the new method, we also give the detailed results of EC-GAN and EC-RGAN on each subject. The results of the two methods on each subject through 60 subjects are shown in Fig. 6 and Fig. 7. From Fig. 6 (A), we can find that EC-RGAN performs better than EC-GAN on most of the subjects, one reason is that EC-RGAN has fewer redundant arcs. In particular, EC-RGAN achieves 1 on subject 6, 18, 38, and 58, while EC-GAN only achieves 1 on subject 18. Fig. 6 (B) shows that EC-GAN and EC-RGAN have similar results in recall. However, the worst performance of EC-RGAN is 0.67, while EC-GAN gets 0.5 of recall on subject 2 and subject 53. Fig. 6 (C) shows the F1 value of the two methods. As F1 is a comprehensive evaluation of precision and recall, we can find that EC-RGAN has a better performance on more subjects.

Fig. 7 shows the results on Sim2, and the results are similar to that of Sim1. In detail, in Fig. 7 (A), the worst performance in precision of EC-GAN is 0.5 (subject 24), and EC-RGAN is 0.6. From Fig. 7 (B), we can find that EC-GAN get the

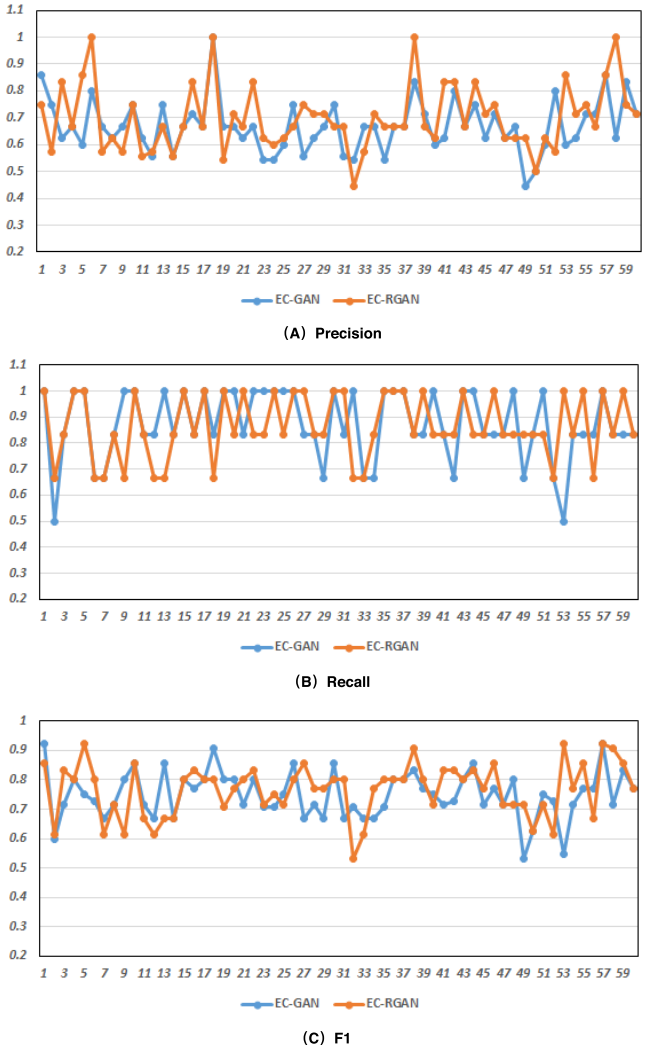


Fig. 6. The results of EC-GAN and EC-RGAN on Sim1. The horizontal axis is the serial number of the subjects (1-60), and the vertical axis is the value of the evaluation metrics (A) Precision, (B) Recall, and (C) F1.

worst performance on subject 38 with the recall of 0.43, but the recall of EC-RGAN is higher 0.5. Fig. 7 (C) shows the F1 value of EC-GAN and EC-RGAN on 60 subjects which are higher than the results in Fig. 6 (C).

Next, we discuss what makes the new method EC-RGAN more accurate than EC-GAN. We give two main reasons that affect the performance of the corresponding two methods in Fig. 8. The results show that even if we employ the same hyper-parameter and threshold, the EC-RGAN has a better ability to distinguish the directions among brain regions.

From Fig. 8, we can also find that EC-RGAN has fewer bidirectional arcs and redundant arcs compare to EC-GAN. As the EC-GAN and the EC-RGAN employ the same hyper-parameters and the same structure of the neural networks, the increase in precision may due to the use of temporal information.

In a word, EC-RGAN performs better than EC-GAN in precision and F1 value, and have a comparative result in recall. Then we compare EC-RGAN with other state-of-the-art methods in the following.

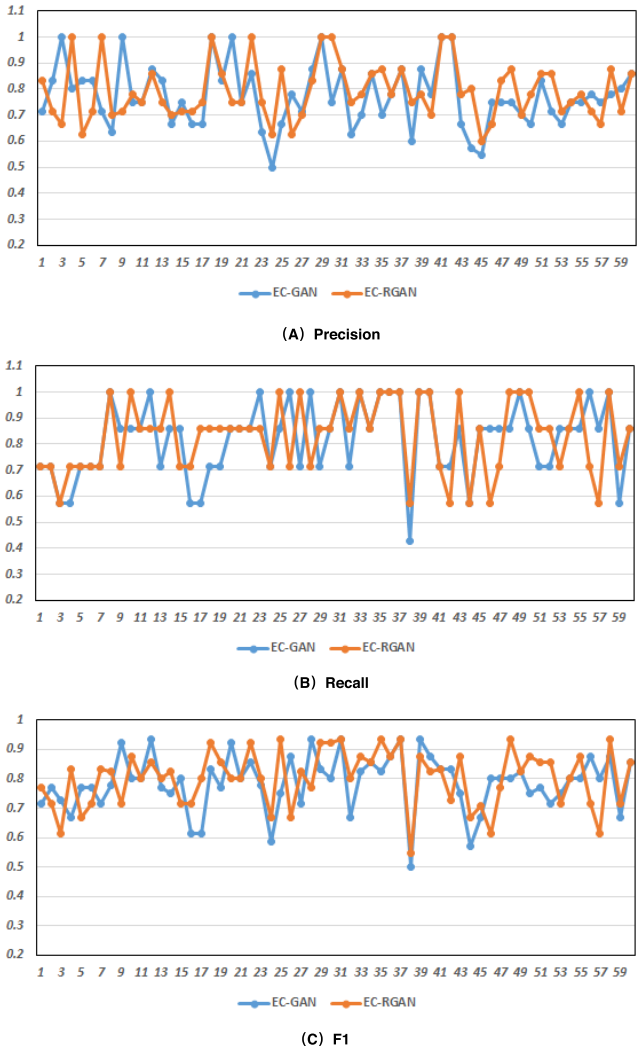


Fig. 7. The results of EC-GAN and EC-RGAN on Sim2. The horizontal axis is the serial number of the subjects (1-60), and the vertical axis is the value of the evaluation metrics (A) Precision, (B) Recall, and (C) F1.

E. Comparisons on Benchmark Simulated fMRI Data

In the experiments, we run each method on each subject, and show the results in Fig. 9 and Fig. 10. In detail, the labels on the out ring show the serial number of the subjects, and the labels in the middle of each graph show the value of each metric. In our experiments, we use single subjects (500 data points $\times$ 60 subjects), mainly because we want to test the performance of different methods on a small sample size of fMRI data. In particular, an algorithm performs well when it gets higher values of precision, recall, and F1.

From Fig. 9 we can find that EC-RGAN performs the best in recall and F1 value, while gets a mediocre performance in precision. Specifically, AIAEC and GES have a good performance in precision in Fig. 9 (A). One reason for these methods performs well maybe that both AIAEC and GES are score-based BN methods. They employ a penalty function in their score functions, which can learn a sparse graph without training a sparse parameter. In detail, the ground-truth network of Sim1 is a causal graph with 5 nodes and 6 arcs, which is a sparse graph. The number of arcs learned by GES and

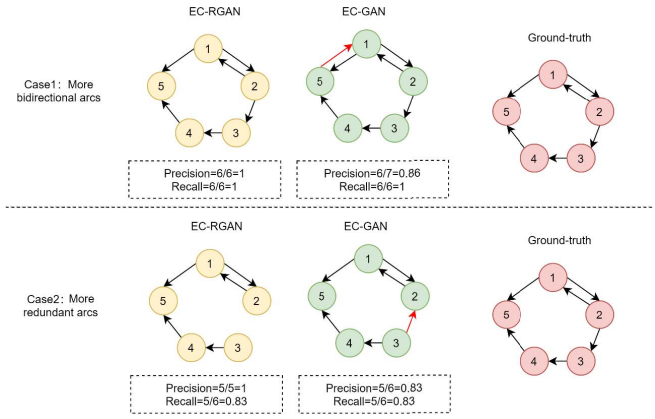


Fig. 8. The two cases that EC-RGAN has a better performance than EC-GAN on precision.

AIAEC is around 6 arcs, while the number of arcs learned by EC-RGAN is around 8 arcs. So the precision of GES and AIAEC is higher than that of EC-RGAN. From Fig. 9 (B), we can find that EC-RGAN has an outstanding performance in recall. In particular, it can get all true positive arcs on more than 20 subjects. Fig. 9 (C) shows that EC-RGAN gets the best performance on most of the subjects, which demonstrates that the EC-RGAN has excellent learning performance under the small sample and bidirectional structure fMRI data.

Compared to the results in Fig. 9, BN methods have an improvement in precision and performance degradation in recall in Fig. 10. In contrast, MVARp and PcLiNGAM perform worse in precision but have a better performance in recall. The reason is that as the ground-truth network becomes dense with more bidirectional connections (from Sim1 to Sim2), BN methods cannot model the bidirectional connections, which lead to the decrease of recall. From Fig.10 we can also find that EC-RGAN performs the best in recall and F1, and has an obvious increase in precision.

Besides, to show the results intuitively, we also give the mean μ and the standard deviation σ results of 7 methods through 60 subjects. The results on two simulated data sets are shown in Table V. From Table V we can find that the baseline methods perform worse in recall (lower than 0.6) and F1 (lower than 0.65) on Sim1, however, our proposed method EC-RGAN achieves the best performance in recall (0.87) and F1 (0.76). Besides, AIAEC is better than EC-RGAN in precision, but has a higher standard deviation. This indicates that EC-RGAN has a more stable performance compared to AIAEC. The results on Sim2 is similar to that of Sim1, EC-RGAN still performs the best in recall and F1. It is worth noting that EC-RGAN gets the highest value in precision which is equal to AIAEC, and has a smaller value of standard deviation. These results show that EC-RGAN can estimate more reliable effective connectivity than other baseline methods.

In summary, the proposed model EC-RGAN performs better than the six comparison methods on the simulation data. Next, We discuss its performance on the real fMRI data in the following section.

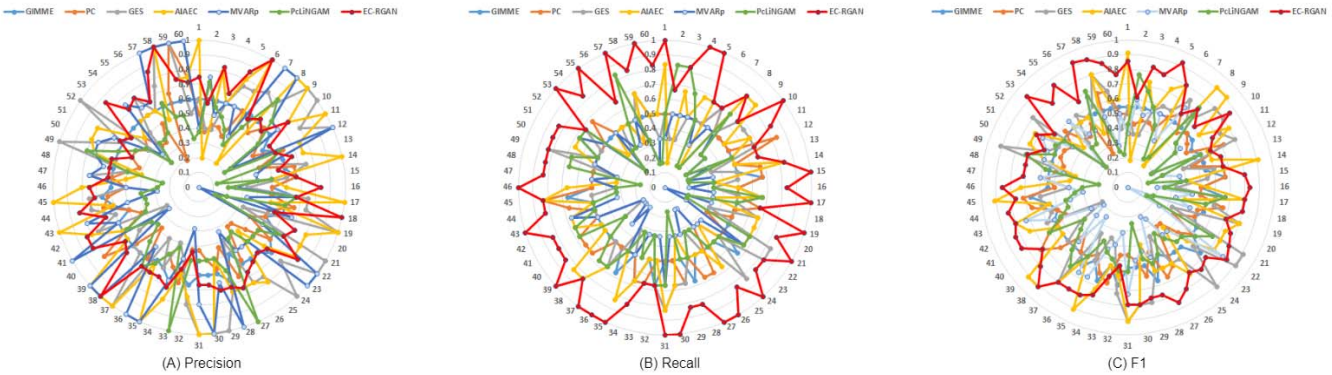


Fig. 9. The results of seven methods on Sim1. The outer ring is the serial number of subjects, and the inner ring shows the value of the evaluation metrics (A) Precision, (B) Recall, (C) F1.

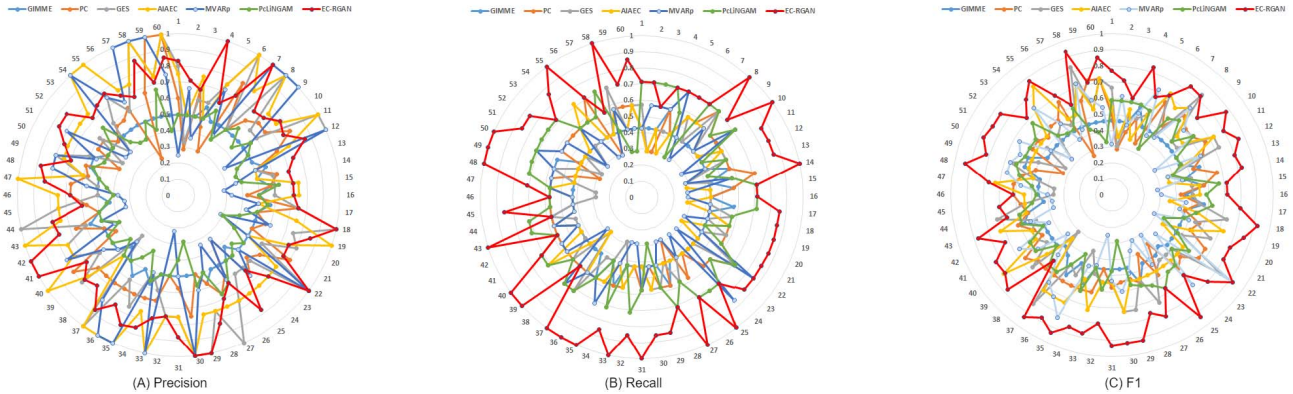


Fig. 10. The results of seven methods on Sim2. The outer ring is the serial number of subjects, and the inner ring shows the value of the evaluation metrics (A) Precision, (B) Recall, (C) F1.

TABLE V
THE MEAN AND THE STANDARD DEVIATION RESULTS OF 7 METHODS ON SMITH SIMULATED DATA

Data	Metrics	Methods						
		GIMME	PC	GES	AIAEC	MVARp	PcLiNGAM	EC-RGAN
Sim1	Precision	0.60±0.04	0.52±0.13	0.70±0.22	0.75±0.21	0.65±0.26	0.48±0.18	0.70±0.12
	Recall	0.52±0.06	0.54±0.10	0.51±0.16	0.57±0.17	0.36±0.15	0.45±0.20	0.87±0.12
	F1	0.56±0.04	0.53±0.10	0.58±0.18	0.61±0.18	0.44±0.16	0.45±0.16	0.76±0.09
Sim2	Precision	0.51±0.03	0.60±0.14	0.71±0.19	0.79±0.15	0.61±0.25	0.48±0.08	0.79±0.11
	Recall	0.45±0.06	0.54±0.10	0.46±0.11	0.46±0.10	0.47±0.15	0.61±0.13	0.82±0.15
	F1	0.48±0.04	0.56±0.10	0.56±0.13	0.58±0.12	0.51±0.15	0.53±0.09	0.80±0.09

F. Results on Real Task fMRI Dataset

To test the performance of the algorithms with fMRI task data, we use data previously published in [13], in which nine subjects judged if a pair of visual stimuli rhymed or not. The task data included an Input variable build by convolving the rhyming task boxcar model with a canonical hemodynamic response function. If the algorithms infer orientations correctly, then edges from the Input variable must feedforward into the regions of interest, and no edge should point backward into the Input variable. In this way, we can model the dynamics of the task with an Input variable for which we expect feedforward edges into the regions of interest and not vice versa. This is a simple limited but gold standard test for the accuracy of orientation algorithms.

In the experiments, we run EC-RGAN on the task fMRI data of 9 subjects concatenated. The effective connectivity networks estimated by EC-RGAN and other baseline methods are shown in Fig. 11.

From Fig. 11 we can find that both EC-RGAN and EC-GAN can infer the effective connectivity $Input \rightarrow LOCC$ and $Input \rightarrow LACC$, which indicates that they can correctly infer the feedforward connections from the Input variable to the brain regions. GIMME and PC have the connections i.e., $Input - LOCC$ and $Input - LACC$, but the directions are opposite. AIAEC and PcLiNGAM only get one effective connectivity $Input \rightarrow LOCC$, and MVARp gets two bidirectional connections. From Fig. 11, we also find that more than half of the methods get the effective connectivity

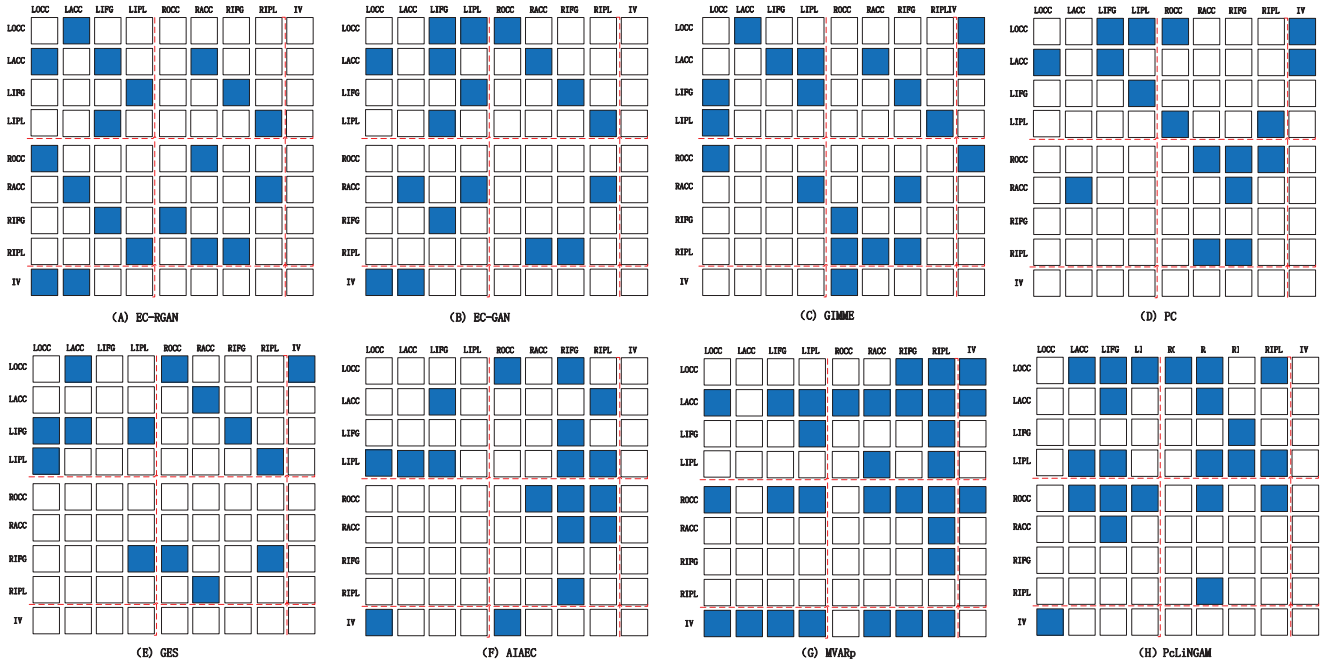


Fig. 11. The effective connectivity estimated by EC-RGAN and the other six methods from the task fMRI data. The horizontal and vertical coordinates indicate the corresponding regions of interest, the blue grid indicates effective connectivity between the two corresponding regions. The red dotted line divides the entire network into nine regions, including the effective connectivity in the three parts (left brain, right brain, and input variable) and the effective connectivity between each of the two corresponding parts.

i.e., $LACC \rightarrow RACC$, $LIFG \rightarrow RIFG$, $LIPL \rightarrow RIPL$. These results may indicate that the left hemisphere of brain regions always activated earlier than the right hemisphere of brain regions under this task (the information flow is from the left brain to the right brain).

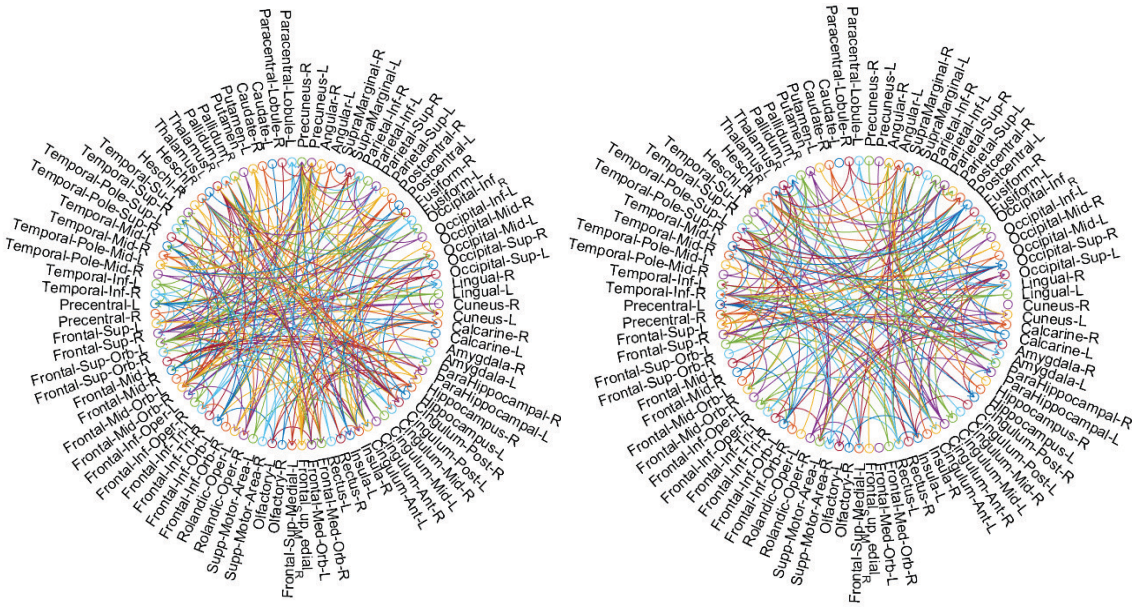
The results on real task data show that the new method EC-RGAN can provide a new and reliable perspective for the analysis of effective connectivity in task data. To further compare the performance of EC-RGAN and EC-GAN on real-world data, we carry out more experiments on the ADNI dataset.

G. Results on the ADNI Database

In this section, we first run EC-RGAN on the ADNI data (group analysis). In detail, there are two common ways to estimate effective connectivity on the real-world data, one is to estimate voxel-wise connectivity, and the other is to estimate region-wise connectivity [30], [31]. In particular, EC-RGAN can estimate the effective connectivity from fMRI data on both voxel-level and region-level. However, EC-RGAN needs to estimate all causal parameters and time series for each voxel (node). Therefore, EC-RGAN has a lower time efficiency on voxel-based estimation. In fact, region-based analysis is more reasonable and common for effective connectivity analysis, thus we employ region-based connectivity estimation on ADNI data analysis.

The brain effective connectivity networks estimated by EC-RGAN are graphically rendered in a circular diagram format in Fig. 12, where the outermost rings represent the brain regions and the center is a representation of brain effective connectivity. In this section, we employ the Automated Anatomical Labeling (AAL) template as the parcellation map to define the brain regions (90 ROIs). Each brain region is represented by a circle with different colors (some may be the same), and the color of arrows is the same as the parent nodes.

To analyze the differences between the two groups, we calculated the number of effective connectivity in each group. In detail, the number of EC in HC and AD groups is 259 and 204, respectively (with the same threshold of 0.1). In particular, there are significant decreases of EC in the Precuneus, Temporal, Precentral, Cingulum, and Hippocampus, etc. These findings once again verify previous studies implicating that there are changes in these regions related to AD, which can help explain and predict the progression and evolution of AD disease [2], [3]. Aside from having different amounts of brain effective connectivity at the global scale, AD groups also have a different pattern of connectivity across the brain compared with HC subjects. In particular, we can also find that the directions of the effective connectivity are different in AD patients compared to HC subjects. For instance, the directions of effective connectivity in HC subjects are from other regions to Frontal-Sup-Orv-L, but they have opposite directions (from Frontal-Sup-Orv-L to other regions) in AD patients. This may indicate that the neural influence and information flow may be different between HC groups and AD groups. In the current study, lots of work have found that the reduction of effective connectivity or functional connectivity is closely related to brain disease, while the situation of reverse connection is rarely reported. Therefore, causal learning is an important and relevant aspect of fMRI research, and the proposed EC-RGAN can provide a new perspective and application for the analysis of AD disease. Besides, we also find that AD



(A) HC (B) AD

Fig. 12. The effective connectivity estimated by EC-RGAN from ADNI data of HC and AD groups.

groups have more EC in the SupraMarginal-R than HC groups. This phenomenon may be related to compensatory processes for AD patients, however, the implicative mechanism behind the phenomenon remains unclear, and it needs to be further discussed and studied.

As the effective connectivity networks on the ADNI dataset have no ground-truth networks, we also add a classification experiment to further compare the EC-RGAN and the other state-of-the-art methods. In the classification experiment, we can employ the label (AD or HC) as the ground-truth [25], [32]. If the brain network effective connectivity network estimated by a method can be successfully classified, then we think that this method performs well on effective connectivity estimation. In detail, we compare our proposed method EC-RGAN with several other effective connectivity estimation methods (EC-GAN, PcLiNGAM, MVARp, and PC) for AD classification on the ADNI dataset, and all methods employ the same classifier (random forest with 1000 decision trees). We divide the data set into 10 parts, nine of them were taken as training data in turn, and one was used as test data for testing. Then all methods run on the same training data and test data. In particular, the classification accuracy of GES and AIAEC is terrible, which lacks practical or clinical significance, so the results are not reported. The classification results of 10 runs are shown in Fig. 13.

From Fig. 13, we can find that the proposed method EC-RGAN outperforms the competing state-of-the-art methods. The classification accuracy (mean value) by EC-RGAN scheme of 0.72, which is the highest among the competing methods. Besides, the t-test results show that EC-RAGAN has a significant difference with the other four methods (p-value < 0.05), which indicates that the proposed method has an excellent ability to identify AD patients from normal controls.

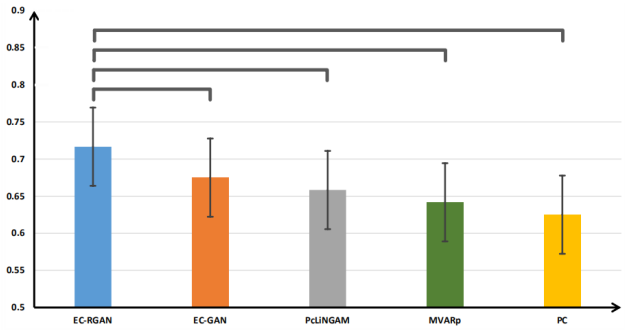


Fig. 13. The comparison of classification performance. The height of the column is the mean value of accuracy, and the error bar is the variance. The line above the histogram indicates that there is a statistical difference between the two algorithms (p-value < 0.05).

TABLE VI
COMPARISON OF THE CLASSIFICATION PERFORMANCE

Methods	Metrics			
	Precision	Recall	F1	BAcc
PC	0.81±0.09	0.75±0.05	0.78±0.06	0.49±0.10
MVARp	0.81±0.11	0.74±0.04	0.77±0.06	0.47±0.09
PcLiNGAm	0.77±0.08	0.77±0.02	0.77±0.05	0.55±0.04
EC-GAN	0.86±0.05	0.75±0.06	0.80±0.02	0.50±0.12
EC-RGAN	0.73±0.08	0.87±0.01	0.79±0.05	0.70±0.04

To analyze the classification performance more comprehensively, we also employ the precision, recall, F1 value, and balanced accuracy (BAcc) to evaluate the performance of different methods. Table VI shows the mean value and the standard deviation of the 10 runs for each method.

From Table VI we can find that EC-RGAN achieves the best performance on recall and BAcc. In particular, the BAcc

of EC-RGAN is significantly higher than all other methods. In a word, the comparison of classification further verified that EC-RGAN performs well on real-world data.

IV. Conclusion

In this paper, we proposed a new model to estimate brain effective connectivity network from fMRI time series data based on recurrent generative adversarial networks, called EC-RGAN. The proposed framework infers effective connectivity via a generator and a discriminator. In detail, the generator is composed of several effective connectivity generators that can generate the fMRI time series of each brain region based on effective connectivity, and the discriminator is employed to distinguish between the joint distributions of real and generated fMRI time series. Experimental results on both simulated and real-world data demonstrate the efficacy of our proposed framework.

Acknowledgment

The authors would like to thank the reviewers for their insightful comments and suggestions. They also thank the developers of the Tetrad IV toolbox whose software were referenced during their experiments.

References

[1] K. J. Friston, "Functional and effective connectivity in neuroimaging: A synthesis," *Hum. Brain Mapping*, vol. 2, nos. 1–2, pp. 56–78, 1994.

[2] J. Liu, J. Ji, X. Jia, and A. Zhang, "Learning brain effective connectivity network structure using ant colony optimization combining with voxel activation information," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 2028–2040, Jul. 2020.

[3] A. Hahn *et al.*, "Reconfiguration of functional brain networks and metabolic cost converge during task performance," *Elife*, vol. 9, 2020, Art. no. e52443.

[4] H. Cagnan, E. P. Duff, and P. Brown, "The relative phases of basal ganglia activities dynamically shape effective connectivity in Parkinson's disease," *Brain*, vol. 138, no. 6, pp. 1667–1678, 2015.

[5] L. Gao and T. Wu, "The study of brain functional connectivity in Parkinson's disease," *Transl. Neurodegener.*, vol. 5, no. 1, pp. 1–7, 2016.

[6] B. Li *et al.*, "Abnormal effective connectivity in the brain is involved in auditory verbal hallucinations in schizophrenia," *Neurosci. Bull.*, vol. 33, no. 3, pp. 281–291, Jun. 2017.

[7] P. Bühlmann, J. Peters, and J. Ernest, "CAM: Causal additive models, high-dimensional order search and penalized regression," *Ann. Statist.*, vol. 42, no. 6, pp. 2526–2556, Dec. 2014.

[8] D. Kalainathan, O. Goudet, I. Guyon, D. Lopez-Paz, and M. Sebag, "Structural agnostic modeling: Adversarial learning of causal graphs," 2018, *arXiv:1803.04929*. [Online]. Available: https://arxiv.org/abs/1803.04929

[9] X. Zheng, B. Aragam, P. Ravikumar, and E. P. Xing, "Dags with no tears: Continuous optimization for structure learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 9472–9483.

[10] Y. Yu, J. Chen, T. Gao, and M. Yu, "DAG-GNN: DAG structure learning with graph neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 7154–7163.

[11] S. M. Smith *et al.*, "Network modelling methods for FMRI," *NeuroImage*, vol. 54, no. 2, pp. 875–891, Jan. 2011.

[12] N. Talebi, A. M. Nasrabadi, I. Mohammad-Rezazadeh, and R. Coben, "NCREANN: Nonlinear causal relationship estimation by artificial neural network; applied for autism connectivity study," *IEEE Trans. Med. Imag.*, vol. 38, no. 12, pp. 2883–2890, Dec. 2019.

[13] R. Sanchez-Romero, J. D. Ramsey, K. Zhang, M. R. K. Glymour, B. Huang, and C. Glymour, "Estimating feedforward and feedback effective connections from fMRI time series: Assessments of statistical methods," *Netw. Neurosci.*, vol. 3, no. 2, pp. 274–306, Jan. 2019.

[14] K. J. Friston, J. Kahan, B. Biswal, and A. Razi, "A DCM for resting state fMRI," *NeuroImage*, vol. 94, pp. 396–407, Jul. 2014.

[15] R. Schlösser *et al.*, "Altered effective connectivity during working memory performance in schizophrenia: A study with fMRI and structural equation modeling," *NeuroImage*, vol. 19, no. 3, pp. 751–763, Jul. 2003.

[16] K. M. Gates and P. C. M. Molenaar, "Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples," *NeuroImage*, vol. 63, no. 1, pp. 310–319, Oct. 2012.

[17] J. A. Mumford and J. D. Ramsey, "Bayesian networks for fMRI: A primer," *NeuroImage*, vol. 86, pp. 573–582, Feb. 2014.

[18] J. Liu, J. Ji, G. Xun, and A. Zhang, "Inferring effective connectivity networks from fMRI time series with a temporal entropy-score," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Apr. 22, 2021, doi: 10.1109/TNNLS.2021.3072149.

[19] J. Ji, J. Liu, P. Liang, and A. Zhang, "Learning effective connectivity network structure from fMRI data based on artificial immune algorithm," *PLoS ONE*, vol. 11, no. 4, Apr. 2016, Art. no. e0152600.

[20] J. D. Ramsey *et al.*, "TETRAD—A toolbox for causal discovery," in *Proc. 8th Int. Workshop Climate Informat.*, 2018, pp. 1–4.

[21] P. H. Kassani *et al.*, "Causality-based feature fusion for brain neurodevelopmental analysis," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3290–3299, Nov. 2020.

[22] M. Gilson, A. T. Campo, X. Chen, A. Thiele, and G. Deco, "Non-parametric test for connectivity detection in multivariate autoregressive networks and application to multiunit activity data," *Netw. Neurosci.*, vol. 1, no. 4, pp. 357–380, Dec. 2017.

[23] S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen, "A linear non-Gaussian acyclic model for causal discovery," *J. Mach. Learn. Res.*, vol. 7, pp. 2003–2030, Dec. 2006.

[24] T. Tashiro, S. Shimizu, A. Hyvärinen, and T. Washio, "ParceLiNGAM: A causal ordering method robust against latent confounders," *Neural Comput.*, vol. 26, no. 1, pp. 57–83, Jan. 2014.

[25] Y. Li, H. Yang, B. Lei, J. Liu, and C.-Y. Wee, "Novel effective connectivity inference using ultra-group constrained orthogonal forward regression and elastic multilayer perceptron classifier for MCI identification," *IEEE Trans. Med. Imag.*, vol. 38, no. 5, pp. 1227–1239, May 2019.

[26] L. Ambrogioni, M. Hinne, M. Van Gerven, and E. Maris, "GP CaKe: Effective brain connectivity with causal kernels," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 950–959.

[27] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 2672–2680.

[28] J. J. Liu, J. Ji, G. Xun, L. Yao, M. Huai, and A. Zhang, "EC-GAN: Inferring brain effective connectivity via generative adversarial networks," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, Apr. 2020, pp. 4852–4859.

[29] J. Yoon, D. Jarrett, and M. V. D. Schaar, "Time-series generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2019, pp. 5508–5518.

[30] S. Hayasaka and P. J. Laurienti, "Comparison of characteristics between region-and voxel-based network analyses in resting-state fMRI data," *NeuroImage*, vol. 50, no. 2, pp. 499–508, Apr. 2010.

[31] W. Liu, S. P. Awate, J. S. Anderson, and P. T. Fletcher, "A functional network estimation method of resting-state fMRI using a hierarchical Markov random field," *NeuroImage*, vol. 100, pp. 520–534, Oct. 2014.

[32] Y. Li, J. Liu, Z. Tang, and B. Lei, "Deep spatial-temporal feature fusion from adaptive dynamic functional connectivity for MCI identification," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2818–2830, Sep. 2020.