



IFC-GNN: Combining interactions of functional connectivity with multimodal graph neural networks for ASD brain disorder analysis

Xuan Wang^{a,*}, Xiaotong Zhang^a, Yang Chen^b, Xiaopeng Yang^c

^a University of Science and Technology Beijing, China

^b Dept. of Biochemistry and Molecular Biology Peking Union Medical College Beijing, China

^c China Petroleum Technology and Development Corporation, Beijing, China



ARTICLE INFO

Keywords:

Spatial and temporal characteristics
Interactions of functional connectivity
Deep feature selection
Multimodal data
Graph neural network
Brain disorder classification

ABSTRACT

Many studies now indicate that brain disorders are associated with functional connectivity between brain regions, but the impact of interactions among functional connections on diseases remains to be explored. Moreover, existing models for analyzing the temporal and spatial attributes of fMRI data exhibit limitations in their efficacy. Currently, there is an absence of a comprehensive model capable of simultaneously capturing disease-related interactions within functional connectivity, fully leveraging the temporal and spatial characteristics of fMRI, and achieving superior diagnostic accuracy. To address these gaps, we propose a model to classify brain disorders that employs a novel algorithm and multimodal graph convolutional networks to process the temporal and spatial features of fMRI data separately, along with the demographic information. In the proposed model, a deep feature selection algorithm is employed to discover disease-related biomarkers and evaluate their impact on brain disorders. Remarkably, this approach has led to a classification accuracy of 80.66%, which surpasses the performance of graph neural networks by approximately 10%. Notably, our investigations have uncovered novel biomarkers linked to autism, with our findings corroborated by relevant literature. These achievements underscore the model's potential in revealing critical insights pertinent to autism, thereby facilitating groundbreaking experimental research in the field.

1. Introduction

The rising number of cases, has brought increased awareness to mental health disorders. These conditions can significantly disrupt individuals' normal lives, with some cases posing life-threatening consequences. Autism Spectrum Disorder (ASD) is a notable example, impacting individuals' thinking (cognition), behavior, and emotions, significantly impairing daily functioning and potentially leading to disability. Furthermore, individuals with mental disorders necessitate lifelong treatment, and complications may arise if not addressed properly and promptly. The diagnosis and treatment of mental disorders continue to encounter numerous challenges, primarily attributed to the specific nature of mental disorders. One significant issue is the absence of clear quantitative measures for diagnosing the condition. Presently, the most prevalent methods for diagnosing mental disorders involve behavioral observation and questionnaire surveys[1]. However, both methods are susceptible to misdiagnosis due to the subjective judgment of the doctor or the patient's subjective responses. Therefore, the

development of standardized biometric indicators is imperative for accurately diagnosing psychiatric disorders.

As the current diagnostic techniques continue to advance, the diagnosis of mental disorders is progressively moving towards standardization. Many diagnostic methods that rely on the analysis of testing data are continually emerging. Numerous studies pertaining to the diagnosis of psychiatric disorders through the analysis of functional Magnetic Resonance Imaging (fMRI) data have been published. On the other hand, A growing body of research indicates that aberrations in Autism Spectrum Disorder (ASD) are frequently associated with functional connectivity between brain regions[2]. Similar studies have also been conducted on other mental disorders such as Schizophrenia and Attention Deficit Hyperactivity Disorder(ADHD)[3,4]. However, to our knowledge, the brain's functional network constitutes a complex network structure. Similar to the mutual influence observed between brain regions, it is highly probable that functional connectivity between these regions may also exhibit mutual influences, a notion that has rarely been addressed in research. Identifying new biomarkers related to

* Corresponding author.

E-mail address: ustb.wx@gmail.com (X. Wang).

diseases holds significant importance for diagnosis and treatment. Thus, analyzing the interactions between functional connectivity (IFCs) and their relationship with mental disorders may offer a novel perspective for discovering biomarkers.

Furthermore, given that fMRI data possess both temporal and spatial characteristics, many studies now analyze and process data based on their temporal and spatial features separately. For instance, one study introduced a Spatial-Temporal Attention Graph Convolutional Network (STAGCN) for the classification of functional connectivity[5]. In the spatial domain, it utilizes attention-enhanced graph convolutional networks to process the topological features of brain regions. In the temporal domain, it employs a multi-head self-attention method to capture the temporal relationships between different dynamic Functional Connectivity (FC). The model ultimately achieved classification accuracies of 83.38% and 66.96% in the diagnostic application for Major Depressive Disorder (MDD) and ASD, respectively. Although the study leverages the temporal and spatial characteristics of fMRI data, the accuracy of ASD diagnosis leaves something to be desired. This may be attributed to the feature extraction methods and classifiers not being well-suited, or the graph representation being inappropriate for the current data type. Therefore, there is an urgent need for research into more effective modeling approaches and classification models for fMRI data that adequately consider its temporal and spatial characteristics. Moreover, improved classification model could yield biomarkers that are more relevant to the disease.

Based on the aforementioned considerations, we propose an IFC-GNN model, which employs a temporal image processing-based approach[6] to model and analyze the temporal features of fMRI data. It uses graph convolutional networks to characterize the spatial features of fMRI data and integrates demographic information from the fMRI dataset. Subsequently, a deep feature selection algorithm(DFS) [7] is utilized for further dimensionality reduction and features extraction, to identify new disease-related biomarkers—IFCs. These IFCs are then used as inputs for a graph neural network to classify ASD, and the model's performance is evaluated on the publicly available ASD dataset, ABIDE. The proposed model is designed with dual objectives. The primary goal is to improve the classification accuracy of graph neural networks by effectively leveraging the temporal and spatial features present in functional magnetic resonance imaging (fMRI) data. The second objective is to investigate the potential associations between the mutual relationships of functional connectivity and various diseases, aiming to discover new biomarkers.

The highlights of this work are set out below:

- 1) Proposes a novel representation form of fMRI temporal characteristics;
- 2) Provides a novel multimodal GNN model that can integrate fMRI temporal and spatial features and demographic information;
- 3) Improves accuracy of GNN in brain disorder classification applications, particularly through applications that consider temporal and spatial features;
- 4) Reveals new biomarkers (IFCs) associated with ASD potentially.

The paper is organized as follows: Section II describes the existing work related to the study. Section III outlines the fMRI dataset employed in the research, and an in-depth exposition of the IFC-GNN model. Section IV shows the results obtained from our experiments, offering insights into the performance and efficacy of this model. Lastly, Section V concludes the research and engage in a discourse on potential directions for future advancements in this domain.

1.1. Related work

Currently, the application research of traditional machine learning algorithms in the diagnosis of mental disorders is well-established. Representative methodologies encompass Support Vector Machines

(SVM), Random Forests (RF), and diverse enhanced algorithms derived from these foundational models. In straightforward applications of SVM, the classifier exhibited a peak classification accuracy of 84% when functional connectivity was utilized as the feature input[8]. Typically, SVM is employed primarily as a classifier, often in conjunction with various feature extraction methods[9,10]. Regarding the spatial and temporal characteristics, several studies have employed SVM to analyze the spatial and temporal characteristics of fMRI. For instance, in [11], researchers extracted intensity and correlation-based features that encompass spatial, temporal, or spatiotemporal information associated with brain activity from a voxel's time course in the fMRI time series. This approach diverges from using voxel intensity solely in spatial and/or temporal dimensions as the input feature. However, the innovation lies in the approach to feature extraction instead of the SVM model. Similarly, Dynamic Time Warping (DTW) distance, a similarity measure between BOLD signals of brain regions, was used as features instead of correlation coefficient[12]. As research has shown, at the single-subject level, the inclusion of temporal and spatial features can enhance the classification accuracy of methods such as SVM[13]. Nevertheless, while existing methods employ feature extraction techniques based on temporal and spatial characteristics, they fail to integrate these features seamlessly into the model-building process.

With the escalation in both data volume and dimensionality, conventional machine learning methods prove to be increasingly inadequate. Consequently, deep learning is emerging as a pivotal player in addressing these challenges. Deep learning methods exhibit significant potential in fMRI data analysis. The most widely employed architectures in fMRI analysis include CNNs[14,15], Recurrent Neural Networks (RNNs)[16], Generative Adversarial Networks (GANs)[17,18], and Long Short-Term Memory networks (LSTMs)[19]. Most applications of CNNs in diagnosing ASD through fMRI data have evolved, progressing beyond rudimentary implementations to innovations in feature extraction, 3D convolutional neural networks[20,21], and other facets. Concerning the temporal aspects of fMRI data, RNNs and LSTMs are endowed with temporal attributes and have demonstrated notable efficacy in the classification of brain disorders using fMRI data. While these models excel in capturing the temporal characteristics of fMRI data, they may lack the incorporation of specific brain structures that are implicated in fMRI signals. To capture both spatial and temporal characteristics, a modified 3D CNN was employed for feature extraction and Alzheimer's disease (AD) classification, as highlighted in [22]. Additionally, an adapted convolutional neural network (CNN) referred to as ST-CNN was introduced for this purpose. This model incorporates a spatial network and a temporal network, interconnected through a convolutional operator[23]. Another modified CNN, known as the 4-D CNN based on granular computing, was proposed. This model was trained based on derivative changes in entropy and has the ability to calculate granularity at a coarse level by stacking layers. It achieved an accuracy of 71.3% on the ADHD-200 public dataset[24]. In the study by [25], they introduced a deep learning framework called spatial-temporal Transformer (ST-Transformer) for distinguishing ASD subjects from typical controls using fMRI data. This model featured a linear spatial-temporal multi-headed attention unit designed to capture the spatial and temporal representation of fMRI data, achieving robust accuracies of 71.0% and 70.6% on the ABIDE I and ABIDE II datasets, respectively.

All the previously discussed approaches for handling the spatio-temporal characteristics of fMRI data involve augmenting existing models with either temporal or spatial features. In contrast, the graph neural network stands out due to its intrinsic network structure, which inherently captures the spatial organization of the brain. Fundamental models of Graph Neural Networks (GNNs) include Graph Convolutional Network (GCN)[26], Graph Attention Network (GAT)[27], and GraphSAGE[28]. Studies in GNNs can be divided into various types based on their applications, such as node classification[29], graph classification[30], link prediction[31,32], and others. In the application of fMRI data analysis, the fundamental GNN models already have

Table 1

Demographic information of ABIDE I dataset.

DataSources	Sample Type	Age	Sex	FIQ	VIQ	PIQ
CALTECH	19/19	28.2 ± 10.6	30/8	111.3 ± 11.4	111.1 ± 14.2	109.2 ± 10.9
CMU	14/13	26.6 ± 5.7	21/6	114.6 ± 10.5	112.0 ± 11.8	113.6 ± 10.1
KKI	22/33	10.1 ± 1.3	42/13	106.7 ± 15.3	/	/
LEUVEN	29/35	18.0 ± 5.0	56/8	112.2 ± 13.0	108.3 ± 17.8	105.9 ± 14.9
MAX_MUN	24/33	26.2 ± 12.1	50/7	110.8 ± 11.3	/	111.0 ± 13.0
NYU	79/105	15.3 ± 6.5	147/37	110.8 ± 14.9	109.9 ± 14.8	109.5 ± 15.1
OHSU	13/15	10.8 ± 1.9	28/ 0	111.6 ± 16.9	/	/
OLIN	20/16	16.8 ± 3.5	31/5	113.9 ± 17.0	/	/
PITT	30/27	18.9 ± 6.9	49/8	110.1 ± 12.2	107.5 ± 12.3	109.9 ± 12.1
SBL	15/15	34.4 ± 8.6	30/0	109.2 ± 13.6	110.4 ± 12.4	114.2 ± 12.2
SDSU	14/22	14.4 ± 1.8	29/7	109.4 ± 13.8	108.0 ± 13.9	108.6 ± 13.8
STANFORD	20/20	10.0 ± 1.6	32/8	112.3 ± 16.4	110.2 ± 19.8	111.8 ± 15.6
TRINITY	24/25	17.2 ± 3.6	49/ 0	110.1 ± 13.5	109.1 ± 13.9	109.0 ± 13.0
UCLA	62/47	13.0 ± 2.2	96/13	102.2 ± 13.6	102.9 ± 14.1	100.9 ± 13.5
UM	68/77	14.0 ± 3.2	117/28	106.9 ± 14.0	111.3 ± 17.0	102.7 ± 16.2
USM	58/43	22.1 ± 7.7	101/0	106.6 ± 16.7	103.2 ± 18.3	108.7 ± 15.9
YALE	28/28	12.7 ± 2.9	40/16	99.8 ± 20.1	101.6 ± 20.4	96.8 ± 18.3

promising applications. GCNs, for instance, classified ASD using resting-state fMRI data, achieving a classification accuracy of around 60% [33]. In addition, the GCN model has demonstrated efficacy in unveiling functional connectivity and classifying schizophrenia with 85.8% accuracy [34]. As evident from the former literature, merely employing GCN for feature convolution and classification does not yield satisfactory results. Due to variations in locations, functions, and sizes among different brain regions, the GAT excels in ASD classifications by incorporating a consideration of the varied importance of nodes within the brain network [35]. For instance, in a specific study [36], the GAT achieved a classification accuracy of 72.4%, surpassing the 60% accuracy observed in the application of GCN [33]. In subsequent research on the application of GNN in fMRI data analysis, certain studies focused on optimizing the structure of GNNs [37], while others targeted optimization of the graph representation and other aspects [38]. For instance, in [37,39], the pooling layers were adapted to regularized pooling layers, encouraging sensible Region of Interest (ROI) selection and offering flexibility to preserve either individual- or group-level patterns. When in a specific graph representation method, all the mentioned graph neural networks inherently leverage the structural characteristics of the brain. For example, when considering brain regions as nodes and functional connectivity as edges in a graph, the spatial characteristics of the brain become integral to the graph neural network modeling process. Nevertheless, research on GNN in fMRI data analysis remains relatively scarce and the integration of temporal character has not been thoroughly explored in existing literature. In [40], a novel feature extraction method was proposed, incorporating dynamic temporal information. The extracted features were then fed into a GCN for the classification of recurrent Major Depressive Disorder (MDD). The model demonstrated a classification accuracy of 75.8% for recurrent MDD on the multi-site dataset. Some research based on graph neural networks has involved utilizing the temporal and spatial features of fMRI data [41,42]. Although these models achieve high accuracy, they have not explored the deeper feature, interactions of functional connectivity. Additionally, some studies focus on different types of diseases compared to autism investigated in this paper. While there are some related studies, both their number and methods are very limited.

2. Materials and methods

2.1. fMRI dataset

The resting-state fMRI data can be obtained from Magnetic Resonance Imaging Facilities. Additionally, there are various public fMRI datasets, such as the ABIDE (Autism Brain Imaging Data Exchange) dataset [43] used in this study. The ABIDE dataset comprises data

contributed by 16 major neural research institutions, including NYU, UCLA, Stanford, and others. It encompasses 1112 subjects, consisting of 539 individuals diagnosed with autism spectrum disorder (ASD) and 573 typical controls. This dataset openly shares neuroimaging data, including structural and resting-state functional MRI data, along with corresponding phenotypic information, as detailed in Table 1.

Notably, the neuroimaging data utilized in this paper are sourced from ABIDE-Preprocessed, where the data has undergone preprocessing through the C-PAC (Configurable Pipeline for the Analysis of Connectomes) pipeline. The C-PAC pipeline builds upon a robust set of existing software packages, including AFNI, ANTs, and FSL, incorporating custom Python code for implementation. Following preprocessing, there are 871 valid subjects remaining in ABIDE-Preprocessed, comprising 403 ASD samples and 468 typical controls, accounting for factors such as sample quality and other considerations.

2.2. Model

2.2.1. Interactions of functional connectivity (IFC) network

In the realm of time series network research, the focus lies on understanding the change process over time and identifying the factors contributing to these temporal changes. This is typically investigated by analyzing the interactions among vertices and assigning weights to these interactions to unveil change patterns and their underlying causes. This approach is often referred to as the "two layers model" in this paper. However, in many situations, this method may be insufficient in uncovering all relevant factors. With the advancements in deep learning and big data, there is an increased capacity to process a larger number of features. Consequently, there is a growing interest in exploring deeper and more intricate factors that may play a critical role, factors that might have gone unnoticed in previous studies. Thus, a deeper investigation is warranted to identify additional potentially relevant factors, making the simulation model more precise and closer to approximating real-world conditions.

To address this need, a simulation model named IFC Network has been proposed, introducing what we refer to as a "three layers model," as depicted in Fig. 1-(a). In the first layer, time series networks are utilized, expressing the connections between vertices as a diagonal matrix in which each element represents the weight of the connection between two vertices, as illustrated in the second layer. The algorithm employed to compute the functional connectome network is referred to as GFC (Generate Functional Connectome Algorithm) in this paper. In the third layer, an algorithm named GIFC (Generate Interaction of Functional Connectome Algorithm) is introduced, based on a concept from image processing algorithms applied to contrast two images and enable image recognition.

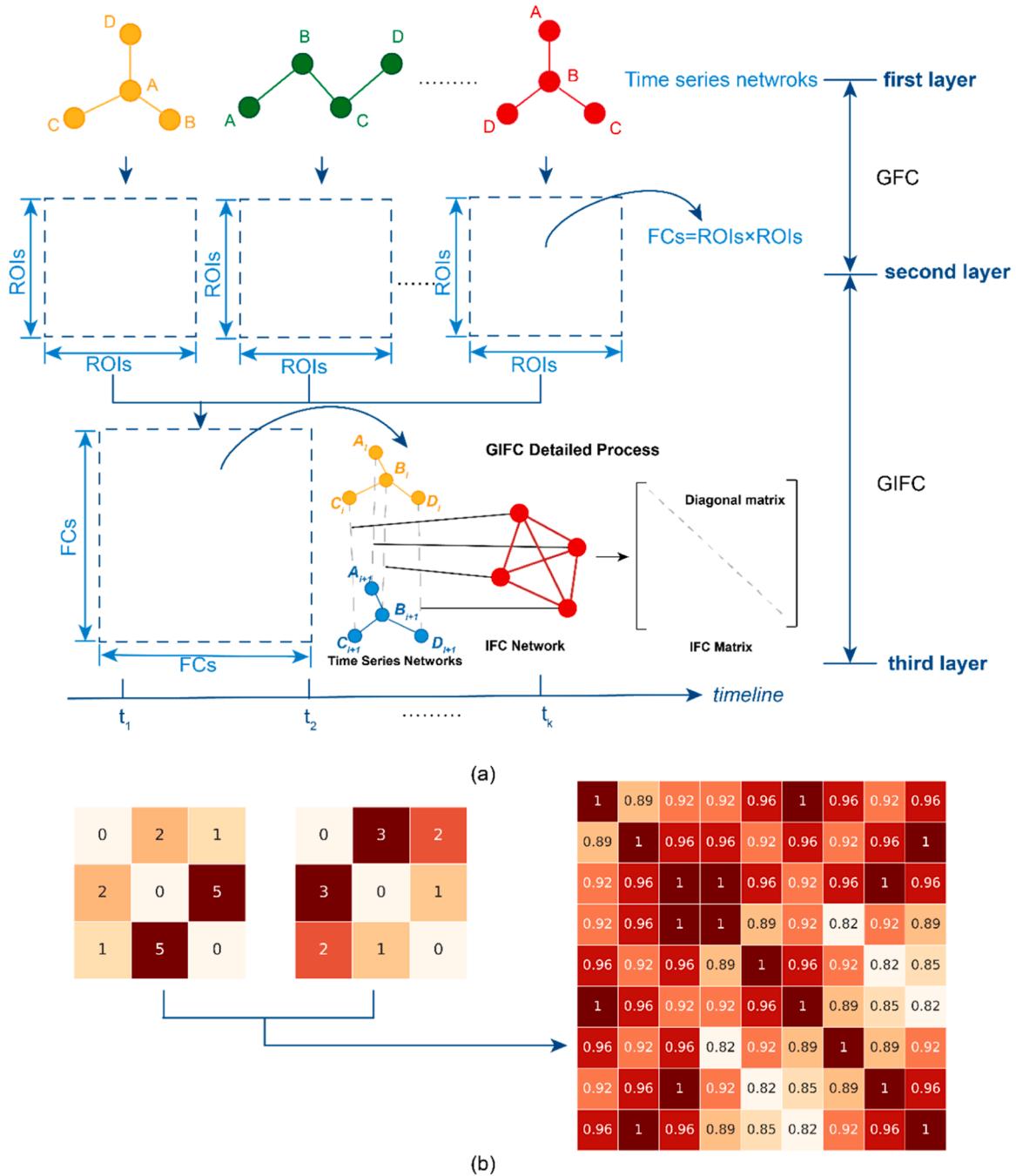


Fig. 1. The logical architecture of IFC and a detailed example. (a) The calculation process of IFC. (b) A detail example to understand the IFC.

In the IFC Network, if there is a network that evolves over time, a series of networks can be obtained chronologically after a specific time period. If $[0, t_t]$ denotes the time span of the time series data, this time span can be divided into $k = \lceil t_t / t_w \rceil$ non-overlapping windows, each characterized by a fixed-size time window t_w . The determination of the time window size will be elaborated on in the subsequent text.

For the Time Series Network (TSN), let n_i represent the current network during time window t_i , where i ranges from 1 to k . Thus, the sequence of TSN can be denoted as $\text{TSN} = \{n_1, n_2, \dots, n_k\}$. The algorithm to extract the time-series functional connectome network collection (FCN) can be expressed as a function $\text{FCN} = \{r_1, r_2, \dots, r_k\} = f_1(\text{TSN}) = f_1(n_1, n_2, \dots, n_k)$, hereinafter referred to as GFC. In various application scenarios, GFC may vary based on the specific application requirements. GFC relies on the reciprocal relationship between features of TSN. The

detailed expression of FCN can be described as follows:

$$\begin{aligned} \text{FCN} &= \{r_1, r_2, \dots, r_k\} = f_1(\text{TSN}) = f_1(n_1, n_2, \dots, n_k) \\ A(r_1), A(r_2), \dots, A(r_k) &= f_1(n_1, n_2, \dots, n_k) \\ A(r_i) &= \left\{ f_1(n_i) | i \in [1, k], |n_i| = m, f_1 : R^{k \cdot m} \rightarrow R^{k \cdot m^2} \right\} \end{aligned} \quad (1)$$

Here, GFC is represented as f_1 , $A(r_i)$ is the functional connectivity matrix of corresponding network n_i , k is the number of time slices, and m is the number of nodes in any time series network n_i . Let's assume that the size of n_i (denoted as $|n_i|$) corresponds to the number of nodes in n_i , so $|n_i| = m$. The mapping function f_1 allows us to map k time-sequence adjacent networks, each with a size of m , to k adjacent matrices, each with a size of m^2 . The mapping function f_1 can vary based on the

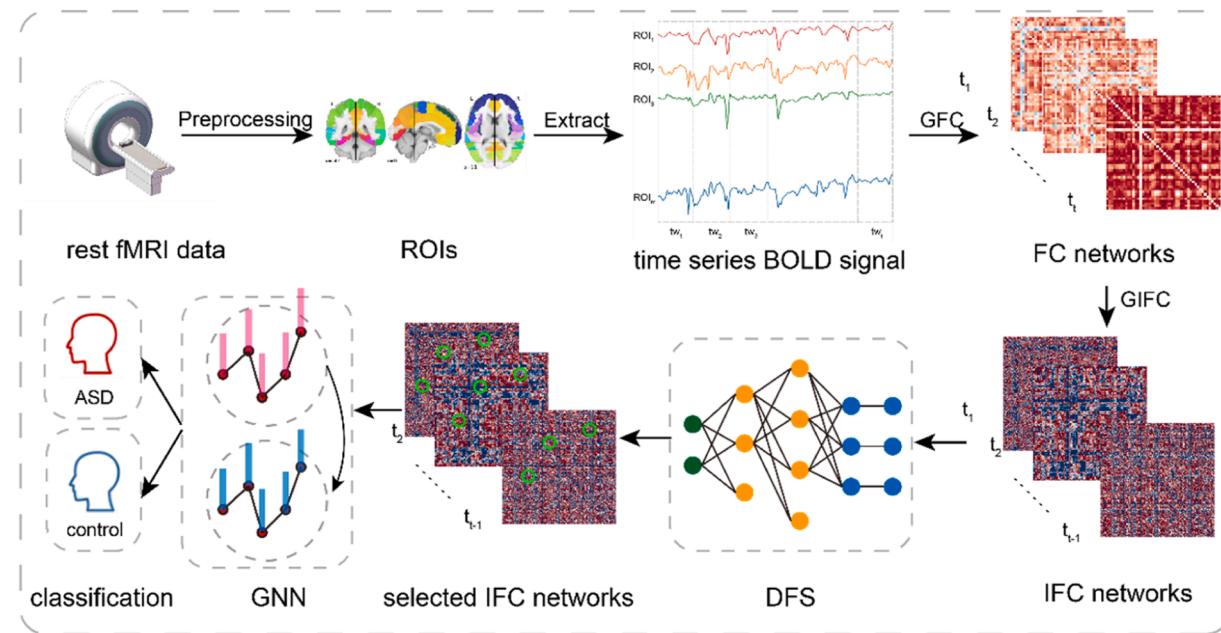


Fig. 2. The workflow of applying the IFC model to the diagnosis of brain disorders using fMRI data.

application. For instance, in this paper, we adopt the mapping function f_1 as the computational method to derive the functional connectivity matrix from fMRI data.

In this paper, GIFC was proposed to calculate the IFC matrix, where each element of the IFC matrix represents the relation between two connectomes of nodes at different times. To calculate the IFC matrix, the IFC network must first be generated, as illustrated in Figure1-(a). In the other words, the correspondence pairs must first be obtained as shown in Figure1-(a). The corresponding relation between any two nodes in two adjacent time series networks is considered as a node in the IFC network. As shown in this part in Figure1-(a), node A_i can be linked to nodes A_{i+1} , B_{i+1} , C_{i+1} , D_{i+1} , and the same goes for B_i , C_i , and D_i . Therefore, nodes A_i , B_i , C_i , and D_i in network n_i have $4^2 = 16$ possible links to nodes A_{i+1} , B_{i+1} , C_{i+1} and D_{i+1} in network n_{i+1} . It can be inferred from this example that if the size of the correlation matrix is m^2 , then the size of the corresponding matrix must be $m^2 \cdot m^2$. The mathematical expression is shown as follows:

$$IFCN = \{c_1, c_2, \dots, c_{k-1}\} = f_2(FCN) = f_2(r_1, r_2, \dots, r_k)$$

$$A(c_1), A(c_2), \dots, A(c_{k-1}) = f_2(r_1, r_2)f_2(r_2, r_3)\dots f_2(r_{k-1}, r_k) \quad (2)$$

$$A(c_i) = \left\{ f_2(A(r_i), A(r_{i+1})) | i \in [1, k-1], f_2 : R^{k \cdot m^2} \rightarrow R^{(k-1) \cdot m^4} \right\}$$

where GIFC was represented as f_2 , $A(r_i)$ was the FC matrix of FC network r_i , $A(c_i)$ was the IFC matrix of IFC network c_i , m was the number of nodes in n_i , and k was the number of time slices separated by a fixed-size time window.

For more details in GIFC, at times t_i and t_{i+1} , the node set of r_i is V_{r_i} and the node set of r_{i+1} is $V_{r_{i+1}}$. Two nodes, p_i , q_i , are selected randomly from V_{r_i} , and corresponding nodes are p_{i+1} and q_{i+1} in $V_{r_{i+1}}$. Since the IFC model describes one network that changes over time, at different times, the node set is same. This means that if there are two nodes p_i , $q_i \in V_{r_i}$, there must be $p_{i+1}, q_{i+1} \in V_{r_{i+1}}$.

For nodes p , q in V_{r_i} and $V_{r_{i+1}}$, two correspondence pairs denoted by $C_{p_ip_{i+1}}$ and $C_{q_iq_{i+1}}$ are considered as the change at features p and q from t_i to t_{i+1} . Suppose $c_i = (C_{p_ip_{i+1}}, C_{q_iq_{i+1}})$, so all the possible correspondence pairs set between r_i and r_{i+1} can be expressed as $C_i = \{c | c \in V_{r_i} \times V_{r_{i+1}}\}$. As defined above, the time series network $TSN = \{n_1, n_2, \dots, n_k\}$, the FC

network is $FCN = \{r_1, r_2, \dots, r_k\}$, and obviously the IFC network will be $IFCN = \{c_1, c_2, \dots, c_{k-1}\}$. The IFC matrix of IFC network c_i was expressed as $A(c_i)$, where c_i is the correspondences set between r_i and r_{i+1} . Suppose that the distance between p and q , namely the weight expressed relation between p and q which can be found in FC matrix, is l_{pq} . The weight of $c_i = (C_{p_ip_{i+1}}, C_{q_iq_{i+1}})$ is $w(c_i) = \exp\left(-\frac{|l_{p_ip_{i+1}} - s \cdot l_{q_iq_{i+1}}|}{\sigma^2}\right)$, where s is scale factor that reflects the variation degree, and σ is a manually set threshold.

Regarding the scale factor s , in the computer vision area, s represents the size ratio of two pictures used to uniform dimension. Here we set the parameter s as 1 because the IFC networks were in the same dimension, as we know. The parameter σ would be changing in its wide range of application scenarios, which we will discuss it later in its brain image classification application.

Overall, the IFC matrix $A(c_i)$ can be defined as

$$A(c_i) = f_2(A(r_i), A(r_{i+1})) = \exp\left(-\frac{|w(P_i, Q_i) - s \cdot w(P_{i+1}, Q_{i+1})|}{\sigma^2}\right) \quad (3)$$

Where $w(P_i, Q_i)$ is the element of $A(r_i)$ that represents the correspondence between any two nodes p and q in r_i , also is the weight between p and q in $A(r_i)$, and $w(P_{i+1}, Q_{i+1})$ is the element of $A(r_{i+1})$ respectively. GIFC actually is the function f_2 here. More details of GIFC were illustrated through an example in Figure1-(b) to enhance comprehension,

considering two adjacent FC matrices $A(r_1) = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 5 \\ 1 & 5 & 0 \end{bmatrix}$,

$A(r_2) = \begin{bmatrix} 0 & 3 & 2 \\ 3 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}$, obviously there are three nodes in each time series

network. For convenience, we have omitted a portion of the IFC matrix as shown below because the matrix scale is too large to display entirely. The abridged form is presented to focus on the calculative process. We selected columns 1–5 and rows 1–5 of $A(c_1)$, denoted as $A'(c_1) = A(c_1)_{5,5}$. With σ is set to 5,

$$A'(c_1) = \begin{bmatrix} 1 & 0.89 & 0.92 & 0.92 & 0.96 \\ 0.89 & 1 & 0.96 & 0.96 & 0.92 \\ 0.92 & 0.96 & 1 & 1 & 0.96 \\ 0.92 & 0.96 & 1 & 1 & 0.89 \\ 0.96 & 0.92 & 0.96 & 0.89 & 1 \end{bmatrix} \quad (4)$$

The entire process of the IFC model has been illustrated above for one sample with m features each time and k time points, which includes two crucial algorithms: GFC and GIFC.

For practical applications, the vertices of the model's network can be considered as chromosomes of DNA or regions of the human brain. Similarly, the edges in the network can be construed as the relationships among different chromosomes or encephalic regions, and the weight of each edge is a parameter that can represent the degree of relevance. For another example, the vertices of the model's network can be considered as atoms, and the edges can stand for the interactions between atoms. Therefore, this algorithm can be used in new drug discovery and the discovery of new compounds. The model can be applied in various fields to discover more corresponding relationships.

2.2.2. Deep feature selection (DFS)

In most brain research, the interactions between different brain regions have been studied from various aspects, while the relationships among these interactions have often been overlooked. It is known that some mental disorders cannot be detected or treated solely based on the interactions between brain regions. In response to this gap, the IFC model has been developed to explore interactions among brain regions when analyzing brain disorders. This approach aims to discover more information and improve the potential for effective treatments.

The model employed in the analysis of brain fMRI data follows the processing procedures outlined in Fig. 2. These steps encompass the preprocessing of resting-state functional magnetic resonance (rs_fMRI) data, the acquisition of time-series FC networks, the calculation of IFC networks using GIFC, deep feature selection, and the classification of brain disorder using graph neural network algorithms on the selected data. Subsequent sections provide more detailed mathematical modeling and process descriptions for further clarification.

In this section, we employ the Deep Feature Selection (DFS) method [20] to extract critical IFC features from functional connectome in autism spectrum disorder (ASD). The DFS network optimizes the Multilayer Perceptron (MLP) by incorporating a bias cell in middle layers and introducing a weighted input layer between the input layer and the first hidden layer.

The DFS network serves multiple purposes:

- Identification of a subset of relevant IFC features specific to ASD for classification;
- Recognition of informative features with nonlinear behaviors through a deep nonlinear structure.

This method addresses the limitations of linear approaches, providing a more straightforward feature selection process. Additionally, the intuitive and straightforward feature selection procedure eliminates the need for complex feature fusion and decomposition. Leveraging a deep nonlinear structure, the DFS network automatically extracts nonlinear features for the classification task, outperforming low-level linear methods. Comparing it with the popular MLP framework, the DFS network introduces a new layer to the deep neural network. Specifically, it incorporates a sparse one-to-one linear layer between the input layer and the first hidden layer of an MLP. This layer differs from a fully connected layer in that each neuron has only one connection to a specific input feature, aiding in the identification of a subset of relevant input features (variables) in the dataset.

Consider a DFS network with L layers, where the weight vectors of

the network are denoted as $\vec{W} = \{w, W^{(1)}, \dots, W^{(L-1)}\}$. Here, w represents the weight vector of the weighted input layer, and $W^{(k)}$ is the weight vector between the k -th layer and the $(k+1)$ -th layer with a size of $W^{(k)} \in R^{n_{k-1} \times n_k}$. In the forward propagation process, the weight vector w of the weighted input layer serves the purpose of selecting only non-zero features. Let the input feature vector be $x = [x_1, x_2, \dots, x_n]^T$, where n is the feature dimension. Each x_i is connected to the i -th neuron in the weighted input layer with a linear activation function. To ensure proper training and backward propagation for each layer in DFS, it is crucial to define a specific loss function. For the weighted input layer, a regularization term $l_r(w)$ is introduced to penalize the loss, as defined below:

$$l_{ri}(w) = \mu_1 \left(\frac{1 - \mu_2}{2} \|w\|_2^2 + \mu_2 \|w\|_1 \right) \quad (5)$$

The $\|\cdot\|_1$ and $\|\cdot\|_2$ in this expression represent the 1-norm and 2-norm, respectively. Here, $\mu_2 \in [0, 1]$ serves as a control parameter governing the trade-off between the smoothness and sparsity of w , while μ_1 is a hyper-parameter.

For the hidden layers, each layer includes a bias neuron that is added and fully connected with one another through a nonlinear activation function, such as sigmoid, tangent, or rectified linear unit (ReLU). The structural layout of DFS is illustrated in Fig. 2. Similar to the weighted input layer, each hidden layer is equipped with a regularization loss to guide the parameter training process. This regularization loss helps in achieving the desired behavior during the training of DFS.

$$l_{rh}(\vec{W}) = \delta_1 \left(\frac{1 - \delta_2}{2} \sum_{k=1}^{L-1} \|W^{(k)}\|_2^2 + \delta_2 \sum_{k=1}^{L-1} \|W^{(k)}\|_1 \right) \quad (6)$$

To obtain the optimal parameters for maximizing classification accuracy, we formulate the entire loss function as follows:

$$\begin{aligned} l(\vec{W}) &= l_{ce}(\vec{W}) + l_{ri}(w) + l_{rh}(\vec{W}) \\ &= l_{ce}(\vec{W}) + \mu_1 \left(\frac{1 - \mu_2}{2} \|w\|_2^2 + \mu_2 \|w\|_1 \right) \\ &\quad + \delta_1 \left(\frac{1 - \delta_2}{2} \sum_{k=1}^{L-1} \|W^{(k)}\|_2^2 + \delta_2 \sum_{k=1}^{L-1} \|W^{(k)}\|_1 \right) \end{aligned} \quad (7)$$

The term $l_{ce}(\vec{W})$ represents the SoftMax cross-entropy loss, and it is employed because SoftMax is used in the output layer of the model. In the DFS model, the Adam optimizer is utilized as it is a stochastic gradient descent method known for effectively searching optimal parameters during network training, and it has demonstrated good performance in various studies. Furthermore, given that the input features are expected to be sparse after the weighted input layer, the DFS architecture is designed to be shallow, meaning that the number of hidden layers should be kept small.

To identify the relevant IFCs of ASD, the DFS network needs to be trained, and the ranked features result is obtained. To enhance the credibility of the results, we adopt K-fold Cross-Validation to measure the mean weight of different IFCs. As the weight of an IFC reflects its contribution to the disease, the weights should be sorted from large to small. Assuming the weight of a certain IFC, IFC_j , in the k -th fold is w_j^k , the overall average weight of IFC_j can be expressed as:

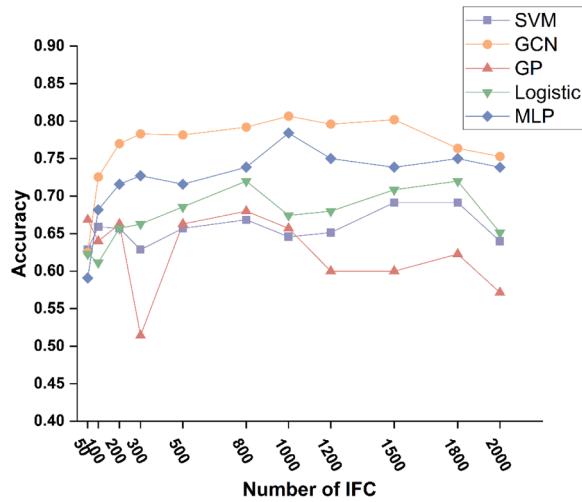
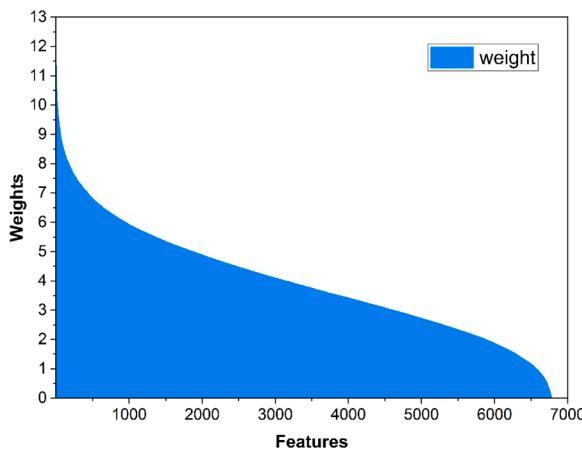
$$w_j = \sum_{k=1}^K |w_j^k| / K \quad (8)$$

Where K is the CV folds number. We will select the top-ranked IFCs, but the appropriate number of features selected is still explored through experiments.

Table 2

The best parameters in each classifier.

classifiers	parameters
SVM	$C = 0.01, \text{kernel} = \text{Linear}$
GP	$\text{max_iteration} = 83, \text{kernel} = \text{RBF}$
Logistic	$\text{tolerance for stopping criteria} = 1e - 4, C = 0.1$
MLP	$\text{hidden_size} = (16, 16), \alpha = 0.001, \text{learning_rate} = 0.001, \text{activation} = \text{relu}$
GCN	$\text{hidden layer number} = 1, \text{learning_rate} = 0.005, \text{drop_rate} = 0.5, \alpha = 0.0005$

**Fig. 3.** The classification accuracy of different methods in various numbers of IFCs.**Fig. 4.** Distribution of feature weights obtained by DFS.

2.2.3. IFC-GNN Model

In order to take full advantage of the spatial-temporal properties of fMRI, we propose the IFC-GNN model, where the IFC modeling process is able to incorporate the temporal nature of fMRI, and in terms of classification prediction, we use the GCN, which on the one hand is able to take into account spatial attributes, and on the other hand is able to incorporate the demographic information of the samples into the modeling process. Here, we adopt the GCN[26] as a classifier. With DFS used before, the Chebyshev Polynomials will be applied as a filter during the GCN modeling.

In Graph Convolutional Networks (GCNs), the application of spectral theory significantly enhances the adjacency matrix through the utilization of Fourier transforms and Taylor expansion, thereby achieving an effective filtering impact. The convolution operation within the spectral

domain on graphs is articulated as the interaction between a signal x and a filter $g_\theta = \text{diag}(\theta)$. This operation can be mathematically represented by:

$$g_\theta * x = U g_\theta(\Lambda) U^T * x = \sum_{k=0}^K \theta_k T_k(\tilde{L}) x \quad (9)$$

Here the matrix U , representing the eigenvectors, is derived through the equation $U = (D^{-\frac{1}{2}}AD^{-\frac{1}{2}} - I_N)^T$, where A is the adjacency matrix, D is the diagonal degree matrix, I_N is the identity matrix of size N , and L is the normalized Laplacian matrix expressed as $L = I_N - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. The Chebyshev polynomial approximation effectively captures the filter $g_\theta(\Lambda)$ in terms of a K-order truncated expansion. Here, Λ denotes the diagonal matrix of eigenvalues, and $g_\theta(\Lambda)$ is approximated using Chebyshev polynomials of the first kind T_k , with θ_k representing the vector of Chebyshev coefficients. The scaling factor $\lambda_{\max}/2$ is used to normalize the eigenvalues, where λ_{\max} is the largest eigenvalue of L .

Consider a graph $G = \{V, E\}$ with N nodes($|V| = N$) and E edges, we can get the degree matrix D and the adjacency matrix A that represents the graph's connections. $A_{ij} = 1$ if there is an edge between node i and j , and $A_{ij} = 0$ otherwise. In GCN theory, the normalized Laplacian matrix can be defined as $L = I - DAD$. In the Chebyshev polynomials definition, $T_0(x) = 1$, $T_1(x) = x$, $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ for $k \geq 2$. Then, the graph convolutional operation using Chebyshev polynomials as $f_\theta(L) = \sum_{k=0}^{K-1} \theta_k T_k(L)$, where θ_k are learnable parameters. The filter $f_\theta(L)$ is applied to node features X as $\hat{X} = f_\theta(L)X$. Now, using Chebyshev polynomials, the layer-wise propagation rule can be approximated as:

$$H^{(l+1)} = \sigma \left(\sum_{k=0}^{K-1} T_k(L) H^{(l)} \Theta_k^{(l)} \right) \quad (10)$$

where $T_k(L)$ is the Chebyshev polynomial of order k applied to the Laplacian matrix L , $\Theta_k^{(l)}$ are the weight matrices associated with the Chebyshev polynomial terms. The Chebyshev polynomials $\Theta_k(L)$ are defined on the interval $[-1, 1]$ and are computed recursively. The K -th order Chebyshev polynomial $\Theta_k(L)$ can be obtained by recursively applying the Chebyshev recurrence relation to the normalized terms.

The use of Chebyshev polynomials allows the approximation of the spectral filter as a polynomial expansion of the Laplacian. This provides a computationally efficient way to capture localized and global information. As to the graph neural network structure, stacking multiple GCN layers allows the model to capture hierarchical features and more complex relationships in the graph. There are several advantages by combining the GCN and Chebyshev Polynomials: Chebyshev polynomials enable the capturing of localized information from a node's neighborhood. The polynomial expansion provides an efficient way to approximate spectral filters. GCNs can handle large graphs efficiently.

During the GCN modeling, we considered the IFC features of each sample as a node and the similarities between the samples as an edge in our application. To obtain the edge information of the graph representation, we adopt a measurement method[44]. Here, we encode the similarity of two individuals based on phenotypic characteristics, including sex, data sources, and handedness. Assuming that there are P kinds of phenotypic features used here, the similarity can be defined by the conception of distance using both distance correlation and a Kronecker delta function as shown below:

$$W_{ij} = \text{sim}(n_i, n_j) = d_c(A_i, A_j) \cdot \sum_{p=1}^P d_k(H_p(i), H_p(j)) \quad (11)$$

where W_{ij} is the weight of the edge between node i and node j , $H_p(i)$ and $H_p(j)$ are the p -th phenotypic data of node i and node j . The two functions in Eq. 10 are d_c and d_k . The function d_k is the Kronecker delta function as shown below.

$$d_k(x, y) = \begin{cases} 0 & x \neq y \\ 1 & x = y \end{cases} \quad (12)$$

The aim of this function is to ensure that the phenotypic feature of two nodes is the same, and the function returns 1; otherwise, it returns 0. The former is a distance correlation function that can be expressed as:

$$d_c(A_i, A_j) = \exp\left(-\frac{|dCor(A_i, A_j)|^2}{2\theta^2}\right) \quad (13)$$

where $dCor$ is the distance correlation, A_i and A_j are the selected feature vectors of node i and node j after DFS, and θ represents the kernel size. The entire program flow can be represented as the following pseudo-code.

The Algorithm of IFC-GNN model

Input: fMRI data, Demographic Information D , Sample Labels Y

```

1: For each pair of ROIs do
2:   calculate the functional connectivity  $FC$  with fMRI data;
3: For each pair  $(FC_i, FC_j)$  do
4:   compute the IFC with  $IFC = GIFC(FC_i, FC_j)$ ;
5: Extract and normalize the feature vector  $X$  from  $IFCs$ ;
6: Initialize the DFS network with random weight vectors;
7: For each iteration do
8:   Train the DFS network with the whole dataset;
9: Choose the DFS network with the lowest loss;
10: For each fold do
11:   Get the weight vectors  $W$  from the trained DFS in each fold;
12: Compute the magnitude vector  $S$  of each feature using  $W$  according to Eq. (8);
13: Sorting  $S$  and select the top  $K$  features to reorganize the feature vectors as  $\tilde{X}$ ;
14: Compute the graph edge matrix by Demographic Information  $D$ ;
15: For each iteration do
16:   Train the IFC-GNN model with the  $\tilde{X}$  and  $Y$  of samples in training dataset;
17:   let  $C$  denote the change in loss between each iteration;
18:   if  $C < threshold$  then
19:     save the model and parameters; return
20:   else
21:     continue the training;
22:   endif
23: Use the trained model to test the model with the  $\tilde{X}$  and  $Y$  of samples in testing dataset, get
the classification accuracy  $Acc$ .
```

Output: The classification accuracy Acc and the selected features \tilde{X}

3. Experimental results and discussions

We adopted a novel algorithm to model and analyze the temporal features of fMRI data as IFCs, utilizing graph convolutional networks to describe the spatial features of fMRI data and integrating demographic information from the fMRI dataset. We further employed a DFS to reduce dimensionality and extract features to identify new disease-related biomarkers, termed IFCs. Subsequently, these IFCs were utilized as inputs for a graph neural network to classify ASD, and the model's performance was evaluated on the publicly available dataset ABIDE. We conducted comparative experiments on the performance of different models using the same set of IFC features and validated them from two perspectives: brain regions and functional connections extracted from the IFCs, based on existing research related to ASD.

Table 3
The magnitude of top 10 IFCs.

No.	IFC – FC ₁	IFC – FC ₂	r _{ASD}	r _{ID}
1	Rectus gyrus(R)	Superior frontal gyrus, orbital part(L)	0.15	0.21
	Cerebellum10(R)	Parietal lobe, inferior division(L)		
2	Paracentral Lobule(L)	Rolandic operculum(R)	0.11	0.15
	Occipital Cortex, middle division(R)	Postcentral gyrus(L)		
3	Precuneus(L)	Cuneus gyrus(R)	0.13	0.22
	Cerebellum7b(L)	Occipital Cortex, middle division(L)		
4	Rectus gyrus(R)	Superior frontal gyrus, orbital part(L)	0.17	0.25
	Cerebellum10(R)	Parietal lobe, superior division(R)		
5	Rectus gyrus(R)	Superior frontal gyrus, orbital part(L)	0.14	0.26
	Parietal lobe, inferior division(L)	Cerebellum10(R)		
6	Precuneus(L)	Cuneus gyrus(R)	0.11	0.20
	Occipital Cortex, middle division(L)	Cerebellum7b(L)		
7	Middle temporal gyrus(L)	ParaHippocampal(R)	0.12	0.25
	Heschl's Gyrus(L)	Cuneus gyrus(L)		
8	Rectus gyrus(R)	Superior frontal gyrus, orbital part(L)	0.14	0.22
	Parietal lobe, superior division(R)	Cerebellum10(R)		
9	Middle temporal gyrus(L)	ParaHippocampal(R)	0.10	0.08
	Cuneus gyrus(L)	Heschl's Gyrus(L)		
10	Paracentral lobule(L)	Rolandic operculum(R)	0.12	0.15
	Postcentral gyrus(L)	Occipital Cortex, middle division(R)		

3.1. The impact of model parameters on the classification accuracy

We conducted a comprehensive comparative analysis of the classification outcomes achieved by the IFC-GNN model against several established baseline methods. To assess the impact of varying numbers of IFC selected on classification results, we employed the same feature selection method. We obtained feature scores using the definition provided in Eq. (8) under 10-fold cross-validation from the DFS network.

Subsequently, these feature scores were input into different models for performance comparison. The comparison involved representative machine learning algorithms, our IFC-GNN model, and MLP. To enhance results, we pre-trained the selected methods and utilized grid search on the validation set to identify optimal parameters for each model. The best parameters for each model are summarized in Table 2.

By applying these parameters in our application, we obtained the classification accuracy of different models for varying numbers of selected IFCs, as illustrated in Fig. 3. All models were trained on ten possible sets of IFCs, and the accuracy values represent the mean results. The Gaussian Process (GP) model tends to converge after just 10 training epochs, making it the fastest to converge. The Logistic Regression model approaches convergence after 200 epochs, while both the Support Vector Machine (SVM) and Multilayer Perceptron (MLP) models tend to converge around 100 epochs. The Random Forest (RF) model converges after 83 epochs, and the IFC-GNN model approaches convergence after approximately 80 epochs. Examining the accuracy curves, we observe that GCN consistently outperforms MLP across all selected IFC numbers. Notably, the GICF-GNN model achieves the highest accuracy of 80.66% with 1000 selected IFCs. Additionally, all classifiers demonstrate effective performance. For instance, GCN attains a mean classification accuracy of 76.35% underscoring the significance of IFC in disease classification and affirming the validity and relevance of our study. Previous studies directly applied GCN to disease classification using fMRI data, achieving an average accuracy of 64.27% and a maximum accuracy of 75.9% on the ABIDE dataset[45]. These accuracies are lower than the classification accuracy achieved by our proposed IFC-GNN model. One reason for the superior performance of the IFC-GNN model could be its integration of demographic features related to the disease, thus overcoming the information barrier of single-modal data. Additionally, it is possible that our model's graph representation method is better suited for this application.

In general, the two deep learning methods (MLP, GCN) outperform traditional machine learning methods (SVM, GP, Logistic), as depicted in Fig. 3. However, when the number of selected IFCs is less than 100, deep learning methods exhibit inferior performance compared to machine learning methods. This phenomenon is likely attributed to the limited size of training samples, with machine learning being more suitable for analyzing and processing small samples than deep learning. Among the machine learning methods, they exhibit similar performance

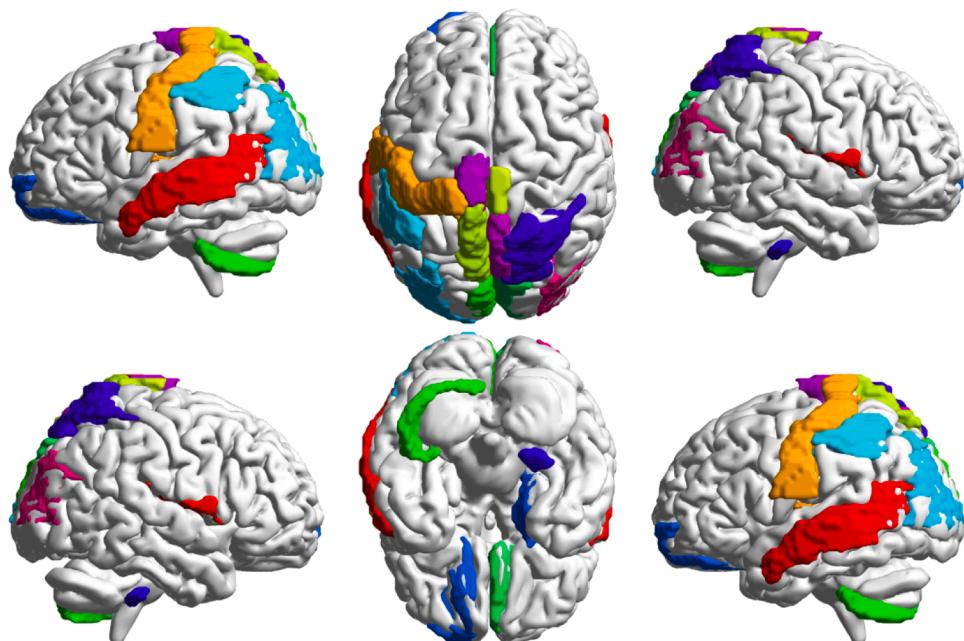


Fig. 5. The related ROIs in Top 10 IFCs.

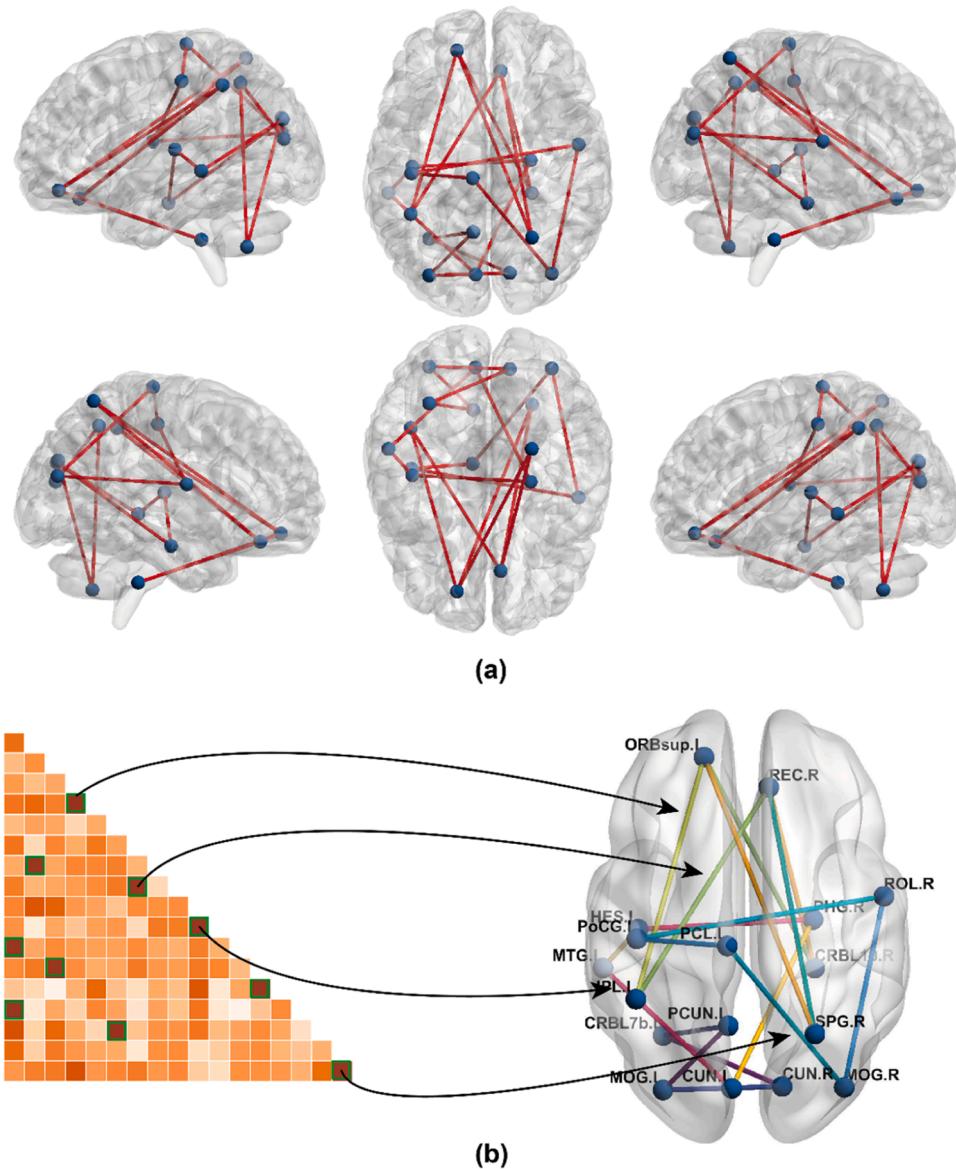


Fig. 6. (a) All the related FCs in Top 10 IFCs. (b) The magnitude matrix and the corresponding IFCs in Brain.

when fewer than 200 IFCs are selected. Although they perform better with fewer than 100 selected IFCs, as the number of selected IFCs increases, the accuracy becomes less stable compared to deep learning methods. The Logistic model achieved its highest accuracy of 72.00% when 800 IFCs were selected, while SVM attained its best accuracy of 69.14% with 1500 IFCs selected. GP, on the other hand, achieved the best accuracy of 66.86% with only 50 IFCs selected. These results suggest that GP is more suited for smaller datasets compared to SVM and Logistic. Notably, GP experiences significant performance fluctuations with parameter changes, while Logistic outperforms SVM and GP.

3.2. The important IFCs and its validation of existing literature

In the DFS network, feature importance is represented by the vector w . As detailed in the DFS section, features are ranked using $w_j = \sum_{k=1}^K |w_j^k|/K$. A higher corresponding weight indicates a greater impact of the feature on the classification results. Utilizing 10-fold cross-validation, DFS is trained to obtain a converged model after approximately 2000 iterations. In the feature ranking results, it is observed that the weights corresponding to lower-ranked features approach after top

6779 features selected, as illustrated in Fig. 4. The maximum weight for the entire set of features is 12.86. Notably, when the selected number of IFCs is 1000, the weights of these features drop by half compared to the highest weight. Hence, when selecting the number of IFCs, it may not be necessary to include all non-zero features. This explains why we investigate the impact of the number of IFCs in the last section by selecting 2000 IFCs at most.

The number of IFCs is the square of the number of FCs. For ease of presentation, only the top 10 IFCs are analyzed and displayed when examining the sorting results. We calculate the magnitude of the top 10 features related to brain disorder classification, and the results are presented in Table 3. Previous studies on features related to brain disorder diagnosis have primarily focused on brain regions and the functional connectivity between them. Therefore, we analyzed the results primarily by examining the FCs and the associated brain regions involved in IFCs, aiming to validate the experimental results.

In existing research, individuals with ASD often demonstrate heightened activity in the default network, which includes the medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC)/cuneus (Cu), inferior parietal lobe, lateral temporal cortex, and hippocampus, even during rest periods [44,46]. Table 3 and Fig. 5 illustrate that all these

brain regions are encompassed within the top 10 IFCs. Among the brain regions involved in functional connectivity, four FCs were observed in the right hemisphere, six in the left hemisphere, and the remaining ten between the right and left hemispheres. This indicates that abnormal functional connectivity between the left and right hemispheres accounts for a significant portion of the issue.

As previously elucidated, the magnitude matrix can be computed using Eq. (8), with a representative example illustrated in Fig. 6-(b). It is discernible from the figure that the magnitude matrix assumes a diagonal form, wherein the horizontal and vertical coordinates denote the involved FCs, and the value at the corresponding position signifies the IFC between the two FCs. Following feature ranking, a selection of prominent Interactions of Feature Connectivity (IFCs) can be identified, exemplified by the top 10 IFCs depicted in Fig. 6-(b). Each node within the matrix corresponds to a distinct IFC, wherein each IFC involves two FCs. To visually depict these relationships, we employ a consistent color scheme to represent the FCs associated with each IFC in the figure. Notably, in instances where an FC is implicated in multiple IFCs, only one color is represented during the visualization process. Fig. 6-(a) shows the top 10 IFCs in detail, and all the FCs they involve are shown in Fig. 6-(b). Furthermore, it is evident that the ASD-related IFCs predominantly involve under-connectivity ($r_{ASD} < r_{TD}$) FCs, consistent with findings in some literature [47]. For instance, compared to the control group, individuals with ASD tend to exhibit weaker functional connectivity between the frontal gyrus and the parietal lobe [48], a phenomenon reflected in IFCs No.1 and No.4 as shown in Table 3. This result, to a certain extent, validates that the experimental outcomes of the current model encompass common functional connectivity abnormalities observed in ASD patients, providing a biological foundation for ASD diagnosis.

4. Conclusion

The model introduces an innovative algorithm designed to extract IFCs from the temporal characteristics of fMRI data. It utilizes the DFS algorithm to identify potential biomarkers and employs a multimodal GNN for the classification of autism. This GNN architecture integrates spatial features and demographic information derived from fMRI data, providing a comprehensive analysis that leverages diverse data dimensions. By fusing these different modalities, the model enhances its capability to accurately diagnose autism, showcasing the potential of advanced computational methods in interpreting complex biomedical data for neurological research. The experimental results surpass those of traditional methods, underscoring the efficacy of this approach. Moreover, the model identifies IFCs potentially linked to diseases, with results validated against existing research findings, offering a fresh perspective for further exploration into the biological characteristics and diagnosis of mental abnormalities. Currently, the model has only been validated on one disease dataset, and the performance of the model may be coupled with the type of disease. This aspect needs further exploration when the model is extended to more applications in the future.

Moving forward, in the field of neuroimaging research, our aim is to extend the application of this model to encompass a wider array of diseases and data types, with the goal of uncovering additional unexplored disease features to expedite diagnosis. Furthermore, our experiments underscored the importance of jointly analyzing multimodal data. In future iterations, we envision integrating feature fusion techniques for multimodal data into the model, enabling comprehensive analysis of disease data and yielding more robust and credible experimental outcomes. The IFC-GNN model can be applied to explore deep networks or deep features. For instance, the vertices of the model's network can be regarded as atoms, while the edges represent the interaction forces between atoms. The IFC-GNN model can uncover whether there exists a relationship of mutual influence between interaction forces among atoms. Therefore, this algorithm can be inferred for applications in new drug development and exploration of new chemical substances. It can

also be applied to gene interaction modeling to explore more gene relationships, leading to a deeper understanding of the genetic basis of complex traits and the molecular mechanisms of diseases.

CRediT authorship contribution statement

xuan wang: Methodology, Software, Writing – original draft, Writing – review & editing. **Xiaotong Zhang:** Funding acquisition, Resources, Supervision. **Yang Chen:** Conceptualization, Investigation, Supervision. **Xiaopeng Yang:** Investigation, Visualization.

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.

Acknowledgements

This work was supported by the National Natural Science Foundation of China, China, 61971031; the Foshan Higher Education Foundation, China, BKBS202203; the National Key R&D Program of China, China, 2018YFA0801402; the CAMS Innovation Fund for Medical Sciences, China, 2021-RC310-007; the CAMS Innovation Fund for Medical Sciences, China, 2021-I2M-1-020.

References

- [1] V.N. Vahia, Diagnostic and statistical manual of mental disorders 5: a quick glance, Indian J. Psychiatry 55 (2013) 220, <https://doi.org/10.4103/0019-5545.117131>.
- [2] B. Huang, Chapter 12 - Diagnosis of Autism Spectrum Disorder by Causal Influence Strength Learned from Resting-State fMRI Data, in: A.S. El-Baz, J.S. Suri (Eds.), In Neural Engineering Techniques for Autism Spectrum Disorder, Academic Press, 2021, pp. 237–267. ISBN 978-0-12-822822-7.
- [3] O. Demirci, V.P. Clark, V.A. Magnotta, N.C. Andreasen, J. Lauriello, K.A. Kiehl, G. D. Pearlson, V.D. Calhoun, A review of challenges in the use of fMRI for disease Classification / Characterization and A projection pursuit application from A multisite fMRI schizophrenia Study, Brain Imaging Behav. 2 (2008) 207–226, <https://doi.org/10.1007/s11682-008-9028-1>.
- [4] L. Tian, T. Jiang, M. Liang, Y. Zang, Y. He, M. Sui, Y. Wang, Enhanced resting-state brain activities in ADHD patients: a fMRI study, Brain Dev. 30 (2008) 342–348, <https://doi.org/10.1016/j.braindev.2007.10.005>.
- [5] W. Wang, Y. Kong, Z. Hou, C. Yang, Y. Yuan, Spatio-Temporal Attention Graph Convolution Network for Functional Connectome Classification. In Proceedings of the ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2022, pp. 1486–1490.
- [6] H. Liu, S. Yan, Common visual pattern discovery via spatially coherent correspondences (June), Proc. 2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2010) 1609–1616.
- [7] Y. Li, C.-Y. Chen, W.W. Wasserman, Deep feature selection: theory and application to identify enhancers and promoters, J. Comput. Biol. 23 (2016) 322–336, <https://doi.org/10.1089/cmb.2015.0189>.
- [8] P. Orban, C. Dansevere, L. Desbois, V. Mongeau-Pérusse, C.-É. Giguère, H. Nguyen, A. Mendrek, E. Stip, P. Bellec, Multisite generalizability of schizophrenia diagnosis classification based on functional brain connectivity, Schizophr. Res. 192 (2018) 167–171, <https://doi.org/10.1016/j.schres.2017.05.027>.
- [9] S. Xie, R. Guo, N. Li, G. Wang, H. Zhao, Brain fMRI Processing and Classification Based on Combination of PCA and SVM. In Proceedings of the 2009 International Joint Conference on Neural Networks, IEEE, 2009, pp. 3384–3389.
- [10] R. Zafar, A.S. Malik, A.N. Shuaibu, M. Javaid ur Rehman, S.C. Dass, Classification of fMRI data using support vector machine and convolutional neural network (September), Proc. 2017 IEEE Int. Conf. Signal Image Process. Appl. (2017) 324–329.
- [11] X. Song, A.M. Wyrywicz, Unsupervised spatiotemporal fMRI data analysis using support vector machines, NeuroImage 47 (2009) 204–212, <https://doi.org/10.1016/j.neuroimage.2009.03.054>.
- [12] R. Meszlényi, L. Peska, V. Gál, Z. Vidnyánszky, K. Buza, Classification of fMRI data using dynamic time warping based functional connectivity analysis, Proc. 2016 24th Eur. Signal Process. Conf. (August 2016) 245–249.
- [13] J. Mourão-Miranda, E. Reynaud, F. McGlone, G. Calvert, M. Brammer, The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data, NeuroImage 33 (2006) 1055–1065, <https://doi.org/10.1016/j.neuroimage.2006.08.016>.
- [14] J. Zheng, X. Wei, J. Wang, H. Lin, H. Pan, Y. Shi, Diagnosis of schizophrenia based on deep learning using fMRI, Comput. Math. Methods Med. 2021 (2021) e8437260, <https://doi.org/10.1155/2021/8437260>.

- [15] Y.-W. Niu, Q.-H. Lin, Y. Qiu, L.-D. Kuang, V.D. Calhoun, Sample augmentation for classification of schizophrenia patients and healthy controls using ICA of fMRI data and convolutional neural networks, *Proc. 2019 Tenth Int. Conf. Intell. Control Inf. Process.* (2019) 297–302.
- [16] W. Yan, V. Calhoun, M. Song, Y. Cui, H. Yan, S. Liu, L. Fan, N. Zuo, Z. Yang, K. Xu, et al., Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site fMRI data, *EBioMedicine* 47 (2019) 543–552, <https://doi.org/10.1016/j.ebiom.2019.08.023>.
- [17] Q. Yao, H. Lu, *Brain Functional Connectivity Augmentation Method for Mental Disease Classification with Generative Adversarial Network*, in: Z. Lin, L. Wang, J. Yang, G. Shi, T. Tan, N. Zheng, X. Chen, Y. Zhang (Eds.), In Proceedings of the Pattern Recognition and Computer Vision, Springer International Publishing, Cham, 2019, pp. 444–455.
- [18] J. Zhao, J. Huang, D. Zhi, W. Yan, X. Ma, X. Yang, X. Li, Q. Ke, T. Jiang, V. D. Calhoun, et al., Functional network connectivity (FNC)-based generative adversarial network (GAN) and its applications in classification of mental disorders, *J. Neurosci. Methods* 341 (2020) 108756, <https://doi.org/10.1016/j.jneumeth.2020.108756>.
- [19] R. A.S., C.M. Sujatha, *Identification of Schizophrenia Using LSTM recurrent neural network*, *Proc. 2021 Seven-Int. Conf. Bio Signals Images Instrum.* (March 2021) 1–6.
- [20] M.N.I. Qureshi, J. Oh, B. Lee, 3D-CNN based discrimination of schizophrenia using resting-state fMRI, *Artif. Intell. Med.* 98 (2019) 10–17, <https://doi.org/10.1016/j.artmed.2019.06.003>.
- [21] Q.-H. Lin, Y.-W. Niu, J. Sui, W.-D. Zhao, C. Zhuo, V.D. Calhoun, SSPNet: an interpretable 3D-CNN for classification of schizophrenia using phase maps of resting-state complex-valued fMRI data, *Med. Image Anal.* 79 (2022) 102430, <https://doi.org/10.1016/j.media.2022.102430>.
- [22] H. Parmar, B. Nutter, R. Long, S. Antani, S. Mitra, Spatiotemporal feature extraction and classification of Alzheimer's disease using deep learning 3D-CNN for fMRI data, *JMI* 7 (2020) 056001, <https://doi.org/10.11117/1.JMI.7.5.056001>.
- [23] Y. Zhao, X. Li, W. Zhang, S. Zhao, M. Makkie, M. Zhang, Q. Li, T. Liu, *Modeling 4d Fmri Data via Spatio-Temporal Convolutional Neural Networks (St-Cnn)*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 181–189.
- [24] Z. Mao, Y. Su, G. Xu, X. Wang, Y. Huang, W. Yue, L. Sun, N. Xiong, Spatio-temporal deep learning method for ADHD fMRI classification, *Inf. Sci.* 499 (2019) 1–11, <https://doi.org/10.1016/j.ins.2019.05.043>.
- [25] X. Deng, J. Zhang, R. Liu, K. Liu, Classifying ASD based on time-series fMRI using spatial-temporal transformer, *Comput. Biol. Med.* 151 (2022) 106320, <https://doi.org/10.1016/j.combiomed.2022.106320>.
- [26] Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv preprint arXiv:1609.02907* 2016.
- [27] Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks.; February 15 2018.
- [28] Hamilton, W.L.; Ying, R.; Leskovec, J. Inductive Representation Learning on Large Graphs. *arXiv preprint arXiv:1706.02216* 2017.
- [29] J. Sun, B. Wang, B. Wu, *Automated graph representation learning for node classification*, *Proc. 2021 Int. Jt. Conf. Neural Netw.* (July 2021) 1–7.
- [30] M.T. Do, N. Park, K. Shin, Two-stage training of graph neural networks for graph classification, *Neural Process Lett.* (2022), <https://doi.org/10.1007/s11063-022-10985-5>.
- [31] S. Yun, S. Kim, J. Lee, J. Kang, H.J. Kim, Neo-GNNs: Neighborhood Overlap-Aware Graph Neural Networks for Link Prediction, in: In Proceedings of the Advances in Neural Information Processing Systems, Vol. 34, Curran Associates, Inc, 2021, pp. 13683–13694.
- [32] M. Zhang, P. Li, Y. Xia, K. Wang, L. Jin, 2021, Revisiting Graph Neural Netw. Link. Predict.
- [33] M, K, S. Jaganathan, *Graph Convolutional Model to Diagnose Autism Spectrum Disorder Using Rs-fMRI Data*, May 24. In Proceedings of the 2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP), IEEE, Chennai, India, 2021, pp. 1–5. May 24.
- [34] D. Lei, K. Qin, W.H.L. Pinaya, J. Young, T. Van Amelsvoort, M. Marcelis, G. Donohoe, D.O. Mothersill, A. Corvin, S. Vieira, et al., Graph convolutional networks reveal network-level functional dysconnectivity in schizophrenia, *Schizophr. Bull.* 48 (2022) 881–892, <https://doi.org/10.1093/schbul/sbac047>.
- [35] W. Yin, L. Li, F.-X. Wu, *A Graph Attention Neural Network for Diagnosing ASD with fMRI Data*. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2021, pp. 1131–1136.
- [36] C. Yang, P. Wang, J. Tan, Q. Liu, X. Li, *Autism spectrum disorder diagnosis using graph attention network based on spatial-constrained sparse functional brain networks*, *Comput. Biol. Med.* 139 (2021) 104963.
- [37] X. Li, Y. Zhou, N. Dvornek, M. Zhang, S. Gao, J. Zhuang, D. Scheinost, L.H. Staib, P. Ventola, J.S. Duncan, *Braingnn: interpretable brain graph neural network for fmri analysis*, *Med. Image Anal.* 74 (2021) 102233.
- [38] Bessadok, A.; Mahjoub, M.A.; Rekik, I. *Graph Neural Networks in Network Neuroscience*. *arXiv preprint arXiv:2106.03535* 2021.
- [39] X. Li, Y. Zhou, N.C. Dvornek, M. Zhang, J. Zhuang, P. Ventola, J.S. Duncan, Pooling regularized graph neural network for fMRI biomarker analysis, *Med. Image Comput. Comput. Assist. Interv. MICCAI Int. Conf. Med. Image Comput. Comput. Assist. Interv.* 12267 (2020) 625–635, https://doi.org/10.1007/978-3-030-59728-3_61.
- [40] P. Dai, D. Lu, Y. Shi, Y. Zhou, T. Xiong, X. Zhou, Z. Chen, B. Zou, H. Tang, Z. Huang, et al., Classification of recurrent major depressive disorder using a new time series feature extraction method through multisite Rs-fMRI data, *J. Affect. Disord.* 339 (2023) 511–519, <https://doi.org/10.1016/j.jad.2023.07.077>.
- [41] K.-W. Park, S.-B. Cho, A residual graph convolutional network with spatio-temporal features for autism classification from fMRI brain images, *Appl. Soft Comput.* 142 (2023) 110363, <https://doi.org/10.1016/j.asoc.2023.110363>.
- [42] W. Zhou, W. Luo, L. Gong, J. Ou, B. Peng, *Spatial-Temporal Graph Convolutional Network for Insomnia Classification via Brain Functional Connectivity Imaging of Rs-fMRI*, in: Q. Liu, H. Wang, Z. Ma, W. Zheng, H. Zha, X. Chen, L. Wang, R. Ji (Eds.), In Proceedings of the Pattern Recognition and Computer Vision, Springer Nature, Singapore, 2024, pp. 110–121.
- [43] J.A. Nielsen, B.A. Zielinski, P.T. Fletcher, A.L. Alexander, N. Lange, E.D. Bigler, J. E. Lainhart, J.S. Anderson, Multisite functional connectivity MRI classification of autism: ABIDE results, *Front. Hum. Neurosci.* 7 (2013), <https://doi.org/10.3389/fnhum.2013.00599>.
- [44] C.S. Monk, S.J. Peltier, J.L. Wiggins, S.-J. Weng, M. Carrasco, S. Risi, C. Lord, Abnormalities of intrinsic functional connectivity in autism spectrum disorders, *NeuroImage* 47 (2009) 764–772, <https://doi.org/10.1016/j.neuroimage.2009.04.069>.
- [45] T. Yang, M.A. Al-Duailij, S. Bozdag, F. Saeed, *Classification of Autism Spectrum Disorder Using Rs-fMRI Data and Graph Convolutional Networks*. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), IEEE, 2022, pp. 3131–3138.
- [46] R.N. Spreng, W.D. Stevens, J.P. Chamberlain, A.W. Gilmore, D.L. Schacter, Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition, *NeuroImage* 53 (2010) 303–317, <https://doi.org/10.1016/j.neuroimage.2010.06.016>.
- [47] N. Yahata, J. Morimoto, R. Hashimoto, G. Lisi, K. Shibata, Y. Kawakubo, H. Kuwabara, M. Kuroda, T. Yamada, F. Megumi, A small number of abnormal brain connections predicts adult autism spectrum disorder, *Nat. Commun.* 7 (2016) 11254, <https://doi.org/10.1038/ncomms11254>.
- [48] V.L. Cherkassky, R.K. Kana, T.A. Keller, M.A. Just, Functional connectivity in a baseline resting-state network in autism, *NeuroReport* 17 (2006) 1687, <https://doi.org/10.1097/01.wnr.0000239956.45448.4c>.