Artificial Intelligence In Medicine 157 (2024) 102984

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

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

Dynamic functional connections analysis with spectral learning for brain disorder detection

Yanfang Xue [a,b], Hui Xue [a,b,*], Pengfei Fang [a,b], Shipeng Zhu [a,b], Lishan Qiao [c], Yuexuan An [a,b]

[a] School of Computer Science and Engineering, Southeast University, Nanjing, 210096, China
[b] Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Nanjing, 210096, China
[c] School of Mathematical Science, Liaocheng University, Liaocheng, 252000, China

ARTICLE INFO

Keywords:
Brain disorders detection
Dynamic functional connections
Fourier transform
Kernel methods
Spectral learning

ABSTRACT

Dynamic functional connections (dFCs), can reveal neural activities, which provides an insightful way of mining the temporal patterns within the human brain and further detecting brain disorders. However, most existing studies focus on the dFCs estimation to identify brain disorders by shallow temporal features and methods, which cannot capture the inherent temporal patterns of dFCs effectively. To address this problem, this study proposes a novel method, named **d**ynamic functional **c**onnections analysis with **s**pectral **l**earning (**dCSL**), to explore inherently temporal patterns of dFCs and further detect the brain disorders. Concretely, dCSL includes two components, dFCs estimation module and dFCs analysis module. In the former, dFCs are estimated via the sliding window technique. In the latter, the spectral kernel mapping is first constructed by combining the Fourier transform with the non-stationary kernel. Subsequently, the spectral kernel mapping is stacked into a deep kernel network to explore higher-order temporal patterns of dFCs through spectral learning. The proposed dCSL, sharing the benefits of deep architecture and non-stationary kernel, can not only calculate the long-range relationship but also explore the higher-order temporal patterns of dFCs. To evaluate the proposed method, a set of brain disorder classification tasks are conducted on several public datasets. As a result, the proposed dCSL achieves 5% accuracy improvement compared with the widely used approaches for analyzing sequence data, 1.3% accuracy improvement compared with the state-of-the-art methods for dFCs. In addition, the discriminative brain regions are explored in the ASD detection task. The findings in this study are consistent with the clinical performance in ASD.

1. Introduction

The human brain is the most complex dynamic system in the world, which is vulnerable to many neurological or psychiatric diseases such as Alzheimer's Disease (AD) [1,2]. Autism Spectrum Disorder (ASD) [3,4], and Major Depressive Disorder (MDD) [5,6]. Take AD for example, it, characterized by memory loss and cognitive decline in clinical practice, is a progressive neurodegenerative disorder that not only causes agony for patients but also brings heavy financial burdens to patients' families, as well as society [7]. Unfortunately, researchers have not yet found an effective way of treating AD completely. The detection and intervention at its early stage plays a key role in prolonging the onset or progression of AD. Nevertheless, the detection and identification at their early stage is still a challenging problem.

Resting-state functional magnetic resonance imaging (rs-fMRI), as a non-invasive and valuable way of exploring the human brain, characterizes brain activities and patterns by achieving blood-oxygen-level-dependent (BOLD) signals [8,9]. Functional connections (FCs), estimated from BOLD signals, can reflect the interaction between different brain regions of interest (ROIs), and it has become a potentially effective tool for poring abnormalities of the functional interaction and further identifying brain disorders at its early stage. Commonly, previous studies assume that FCs between different ROIs are constant. In practice, this assumption is inconsistent with the practical situation, due to that the status of FCs is dynamic even in the resting-state [10,11]. There is growing evidence that dynamic FCs (dFCs), represented by time-varying FCs between different ROIs, can help us better understand the pathology of brain disease than static FCs. Thus, it is meaningful to estimate dFCs and further investigate inherently temporal patterns over time in brain disorder detection.

Various methods have emerged in recent years to estimate dFCs and further analyze their temporal patterns. Among these methods, the sliding window (SW) approach has gained popularity as the primary method for estimating dFCs [12]. In the SW approach, a time window of fixed length is selected, and the BOLD signals within this window are utilized to estimate FCs. In addition to the vanilla SW, researchers

* Corresponding author at: School of Computer Science and Engineering, Southeast University, Nanjing, 210096, China.
E-mail address: hxue@seu.edu.cn (H. Xue).

https://doi.org/10.1016/j.artmed.2024.102984
Received 1 December 2023; Received in revised form 4 September 2024; Accepted 13 September 2024
Available online 17 September 2024

have developed other dFCs estimation and analysis methods with some priors, such as causal influence [13], topology structure of brain [14] and spatial information [15].

Despite their effectiveness in brain disease detection, most of them tend to emphasize FCs estimation and then identify brain disorders with some shallow temporal features. Li et al. propose to construct adaptive dFCs with a recursive least square algorithm [16]. In their study, the standard deviation (SD) and root-mean-square (RMS) of dFCs serve as temporal features for mild cognitive impairment (MCI) detection. Liang et al. propose to identify two types of brain states by using the K-means clustering method based on estimated dFCs and dynamic fractional amplitude of low-frequency fluctuations (dfALFF) [17]. Jie et al. propose to extract temporal features by calculating the correlation coefficient between dFCs, followed by the use of kernel SVM for identification [15]. Moreover, more temporal feature extraction methods are proposed to analyze dFCs [18–20]. Nevertheless, these studies exhibit three primary limitations. Firstly, the shallow temporal features cannot really reflect detailed dynamic patterns of dFCs. Secondly, correlation-based temporal feature extraction methods are input-independent, disregarding the statistical properties of the input data itself. Lastly, these methods extract temporal features with an implicit assumption that dFCs are stationary. That can lead to the over-stationary problem and make these approaches fail to capture eventful temporal dependencies of dFCs [21,22].

In order to address these problems, researchers have developed various approaches over the years. For the limitations associated with shallow temporal features, recurrent neural networks (RNN)-based models, as the most popular approach, are commonly used to explore higher-level temporal features of dFCs duo to the recurrent mechanism [23–25]. For instance, Lin et al. develop a convolutional recurrent neural network (CRNN) for automated brain disease classification using dFC networks [25]. Regarding the over-stationary problems, frequency domain-based methods have been applied to mine brain activities [26–28]. In which, Fourier transform-based methods are commonly utilized to compute signal frequencies [29]. Cao et al. calculate spectral features of dFCs using fast discrete Fourier transform to improve MCI detection [30]. However, it is worth noting that RNN-based methods may struggle to capture the temporal dependence between two widely spaced FCs effectively, while Fourier transform-based methods primarily capture the fixed spectral features of dFCs and do not inherently learn optimizable spectral features.

To tackle above mentioned issues, this paper proposes a novel analysis method, namely dFCs analysis with spectral learning (dCSL), by combining Fourier transform with kernel methods. Two components are included in dCSL, dFCs estimation and dFCs analysis. In the former, the SW method is used to generate dFCs, and Pearson's correlation (PC) is employed to estimate FCs within each window. In the letter, the Fourier transform is initially integrated with kernel methods to construct a non-stationary spectral kernel mapping. Subsequently, stacking the mapping into a deep architecture for poring time-varying spectrum and further analyzing higher-order/complex temporal patterns of dFCs. Benefiting from the deep architecture, dCSL can capture hierarchical representations of dFCs effectively. The non-stationary spectral kernel mapping enables dCSL to be input-dependent, meaning that the proposed dCSL can take the statistical properties of the input itself into account for correlation-based temporal feature extraction methods. Furthermore, the comprehensive theoretical foundation of Fourier transform and kernel methods ensures that dCSL can effectively explore the intrinsic temporal patterns of dFCs [31].

The contributions of this study are shown as follows:

- Constructing the spectral kernel mapping from the non-stationary kernel, which can capture the long-range correlation between the dFCs.
- Stacking the spectral kernel mapping into a deep architecture and sharing the benefits of the non-stationary kernel and deep hierarchical structure. That has a powerful representation capability, thereby exploring the temporal patterns of dFCs effectively.
- The proposed dCSL naturally has the periodic activation function (*i.e.,* the cosine function), which has higher-order derivatives and can be computed analytically. It means that dCSL can capture the higher-order temporal patterns of dFCs. Moreover, any derivative of the activation function is itself a composition of the activation functions, as the derivative of the cosine is a sine, *i.e.*, a phase-shifted cosine. Therefore, the derivatives of the activation function inherit its properties, enabling us to supervise any derivative with dFCs.
- Conducting a set of experiments on two public datasets. Experimental results demonstrate that the proposal is superior to baseline methods. In addition, the discriminative brain regions are explored in the ASD detection task. The findings in this study are consistent with the clinical performance in ASD.

2. Preliminary

2.1. Notation

To illustrate the proposed method better, some necessary notations are defined in this section. Formally, $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ denote n-dimensional Euclidean spaces and the space of $m\times n$ real-valued matrix. Throughout the paper, matrices, vectors, and scalars are denoted by bold capital letters (*e.g.*, $\boldsymbol{X}$), bold lower-case letters (*e.g.*, $\boldsymbol{x}$), and lower-case letters (*e.g.*, x), respectively.

In this paper, $\boldsymbol{D}$ denotes the data matrix of pre-processed BOLD signals, and $\boldsymbol{d}_i$ denotes the signal of ith ROI. $\boldsymbol{D}^t$ denotes the data matrix of tth window, and $\boldsymbol{A}^t$ denotes the symmetric adjacency matrix of the estimated FCs, which corresponds to the tth window. $\boldsymbol{X}$ is the input matrix of dCSL, and each row $\boldsymbol{x}_{i,j} = [\boldsymbol{A}^1_{i,j}, \boldsymbol{A}^2_{i,j}, \dots, \boldsymbol{A}^T_{i,j}], (i, j = 1, 2, \dots, N)$ of this matrix denotes the dFCs between ith ROI and jth ROI. $\boldsymbol{\Omega} = [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \dots, \boldsymbol{\omega}_M]$, $\boldsymbol{\Omega}' = [\boldsymbol{\omega}'_1, \boldsymbol{\omega}'_2, \dots, \boldsymbol{\omega}'_M]$ are the frequency matrix. $\langle\cdot\rangle$ denotes the inner product of two vector.

2.2. Kernel methods with Fourier transform

Kernel methods are a class of powerful statistical learning approaches. The kernel function $k(\boldsymbol{x}, \boldsymbol{x}') = \langle \boldsymbol{\Phi}(\boldsymbol{x}), \boldsymbol{\Phi}(\boldsymbol{x}')\rangle$ constructs the relationship between the input space and the feature space via the nonlinear mapping $\boldsymbol{\Phi} : \mathcal{X} \to \mathcal{H}$, aiming to make the indivisible data in the input space $\mathcal{X}$ tend to be sparse and further as divisible as possible in the feature space $\mathcal{H}$. In recent decades, kernel methods have been used in many applications, such as multi-label learning [32] and natural language processing [33].

Fourier transform is a fundamental tool of signal processing, which consists of decomposing a signal into a sum of elementary signals that have the property of being easy to implement and observe [34]. These elementary signals are periodic and complex, in order to allow an amplitude and phase study of the systems [35].

Yaglom's theorem [36] constructed the relationship between kernel methods and Fourier transform, where a general kernel is related to a spectral density in accordance with the following Fourier duals.

$$\begin{aligned} k(\boldsymbol{x}, \boldsymbol{x}') &= \int_{\mathbb{R}^T\times\mathbb{R}^T} e^{i(\boldsymbol{\omega}^\top \boldsymbol{x} - \boldsymbol{\omega}'^\top \boldsymbol{x}')} s(\boldsymbol{\omega}, \boldsymbol{\omega}') d\boldsymbol{\omega} d\boldsymbol{\omega}' \\ s(\boldsymbol{\omega}, \boldsymbol{\omega}') &= \int_{\mathbb{R}^T\times\mathbb{R}^T} e^{-i(\boldsymbol{\omega}^\top \boldsymbol{x} - \boldsymbol{\omega}'^\top \boldsymbol{x}')} k(\boldsymbol{x}, \boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}' \end{aligned} \tag{1}$$

where the kernel $k(\boldsymbol{x}, \boldsymbol{x}')$ is positive semi-definite if and only if $s(\boldsymbol{\omega}, \boldsymbol{\omega}')$ is the positive demi-definite bounded variation spectral density of a Lebesgue–Stieltjes measure.

3. Method

As discussed in both the abstract and introduction sections, a novel method dCSL is proposed to analyze the dFCs with spectral learning.

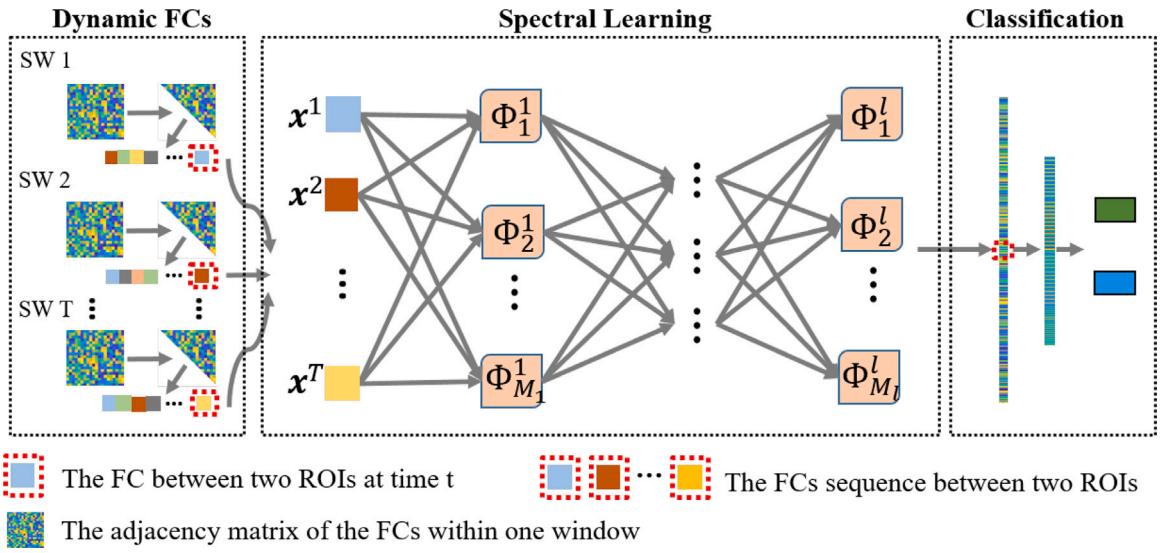


Fig. 1. The pipeline of the proposed dFCs analysis with spectral learning (dCSL) for brain disorders detection. In this pipeline, three steps, including dFCs estimation, spectral learning, and classification with fully connected neural networks. In the dFCs estimation step, the FCs sequence between different ROIs is obtained to characterize the brain activities via the change of FCs. In the spectral learning step, the potential dynamic features of the dFCs are obtained with a stacked non-stationary non-monotonic spectral kernel mapping. In the classification step, a classifier is obtained by optimizing the frequencies to identify the subject with brain disease from NCs.

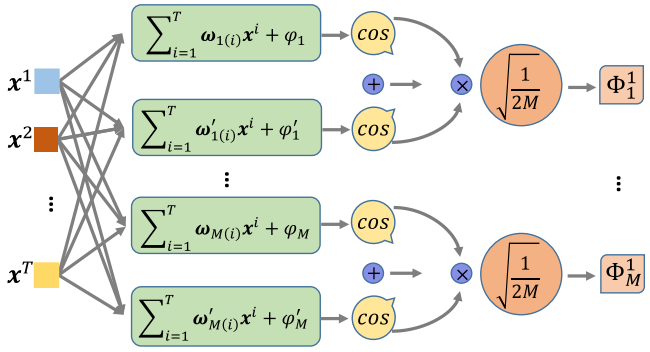


Fig. 2. The architecture of a single layer in spectral learning.

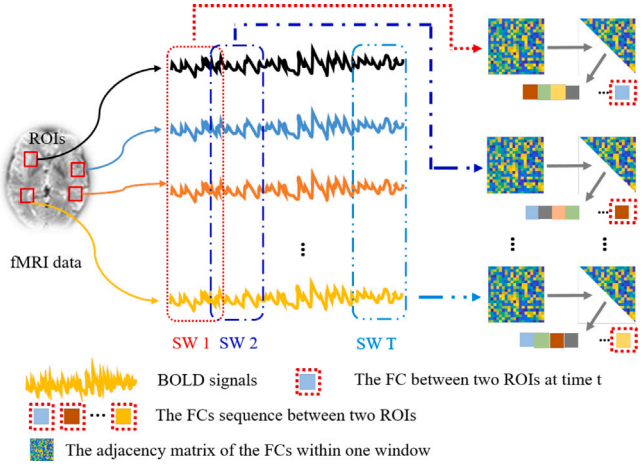


Fig. 3. The process of dFCs estimation.

dCSL mainly includes two components. The first component is used to estimate dFCs with the SW method. The second component is to directly analyze the dFCs via spectral learning and further extract higher-order temporal features to identify brain disorders from NCs. The overall architecture is shown in Fig. 1.

3.1. dFCs estimation with sliding window

SW technique is commonly used for examining the change of FCs over time due to its simplicity [12,37,38]. In this approach, a time window of fixed length is selected, and the BOLD signals within this window are used to estimate FCs. Then, this window is shifted in time by a fixed number of data points that define the amount of overlap between successive windows. This process can result in a set of FC sequences that are considered as dFCs.

In this paper, the pre-processed BOLD signals for each subject is $\boldsymbol{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N]^\top \in \mathbb{R}^{N\times P}$, and $\boldsymbol{d}_i \in \mathbb{R}^P, (i = 1, 2, \ldots, N)$ is the BOLD signal of ith ROI with P time points. Based on SWs, all the signals are segmented into T overlapping windows for each subject, resulting in a set of time series segmentation $\boldsymbol{S} = \{\boldsymbol{D}^t \in R^{N\times P^t}\}_{t=1}^T$, where $P^1 = P^2 = \cdots = P^T$ are the number of time points contained in each segmentation. Then, PC, as the most popular and simplest FCs estimation method, is used for estimating the FCs in each window. As results, obtaining T adjacency matrices $\boldsymbol{A} = \{\boldsymbol{A}^t \in \mathbb{R}^{N\times N}\}_{t=1}^T$ for each subject. The process is shown in Fig. 3.

After getting the dFCs $\boldsymbol{X} = [\boldsymbol{A}_{i,j}^1, \boldsymbol{A}_{i,j}^2, \ldots, \boldsymbol{A}_{i,j}^T]_{i,j=1}^N \in \mathbb{R}^{\frac{N\times(N-1)}{2}\times T}$ based on SWs for each subject, which will be treated as the input data matrix of the dFCs analysis module, the proposed method is used for capturing the higher-order temporal features of dFCs and further identifying the subject with brain disease from the NCs.

3.2. dFCs analysis with spectral learning

Once getting the FC sequence $\boldsymbol{x}_{i,j} = [\boldsymbol{A}_{i,j}^1; \boldsymbol{A}_{i,j}^2; \cdots; \boldsymbol{A}_{i,j}^T]_{i,j=1}^N \in \mathbb{R}^T$, spectral learning is used for analyzing its temporal patterns and extracting higher-order temporal features. Moreover, the extracted higher-order temporal features, which can effectively characterize the dynamics of the dFCs, are used for identifying subjects with brain disease from NCs. The details are shown below.

Table 1
The information of databases.

Database	Category (label)	Subjects (samples)	Age	ROIs	Volumes (time points)
ABIDE	ASD (+1)	125	17.59 ± 7.84	116	170
	NC (−1)	135	16.49 ± 7.68	116	170
ADNI	MCI (+1)	165	72.03 ± 7.71	116	137
	NC (−1)	154	75.36 ± 6.16	116	137

Firstly, the Fourier transform is combined with non-stationary kernel methods to generate the spectral kernel mapping, which is defined as follows:

$$\begin{aligned}\boldsymbol{\Phi}(\boldsymbol{x}_{i,j}) &= \sqrt{\frac{1}{2M}}\begin{bmatrix}\cos(\boldsymbol{\omega}_1^T\boldsymbol{x}_{i,j}+\varphi_1)+\cos(\boldsymbol{\omega}_1'^T\boldsymbol{x}_{i,j}+\varphi_1)\\ \cdots\\ \cos(\boldsymbol{\omega}_M^T\boldsymbol{x}_{i,j}+\varphi_M)+\cos(\boldsymbol{\omega}_M'^T\boldsymbol{x}_{i,j}+\varphi_M)\end{bmatrix}\\ &= \sqrt{\frac{1}{2M}}\left[\cos(\boldsymbol{\Omega}^T\boldsymbol{x}_{i,j}+\varphi)+\cos(\boldsymbol{\Omega}'^T\boldsymbol{x}_{i,j}+\varphi)\right],\end{aligned} \tag{2}$$

where $\boldsymbol{x}_{i,j}$ is the dFCs between any two ROIs. $\boldsymbol{\Omega} = [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_M]$ and $\boldsymbol{\Omega}' = [\boldsymbol{\omega}_1', \boldsymbol{\omega}_2', \ldots, \boldsymbol{\omega}_M']$ are the frequency matrices, which can be interpreted as angular frequencies. The frequency pairs $\{(\boldsymbol{\omega}_i, \boldsymbol{\omega}_i')\}_{i=1}^M$ are sampled from the spectral density $s(\boldsymbol{\omega}, \boldsymbol{\omega}')$. M is the sampling number. While φ is the bias, which can be interpreted as the phase offsets. The detailed architecture is shown in Fig. 2 and the detailed derivation can be found in Appendix.

Subsequently, spectral learning is constructed by stacking the spectral kernel mapping $\boldsymbol{\Phi}$ into a deep architecture to explore the temporal feature of dFCs. The feedforward form and the spectral learning are designed as:

$$\begin{aligned}\boldsymbol{\Phi}^l(\boldsymbol{x}_{i,j}) = \sqrt{\frac{1}{2M^l}}\Big[&\cos(\boldsymbol{\Omega}_{l-1}^T\boldsymbol{\Phi}^{l-1}(\boldsymbol{x}_{i,j})+\varphi_{l-1})\\ &+\cos(\boldsymbol{\Omega}_{l-1}'^T\boldsymbol{\Phi}^{l-1}(\boldsymbol{x}_{i,j})+\varphi_{l-1})\Big],\end{aligned} \tag{3}$$

$$f^{(l)}(\boldsymbol{x}_{i,j}) = \boldsymbol{\Phi}^l(\boldsymbol{\Phi}^{l-1}(\cdots\boldsymbol{\Phi}^1(\boldsymbol{x}_{i,j}))), \tag{4}$$

where $f^{(l)}(\boldsymbol{x}_{i,j})$ denotes the l-layer spectral kernel network, $\boldsymbol{\Phi}^l$ denotes the lth layer spectral kernel mapping.

Note that, in such a deep spectral kernel mapping, the activation function naturally is a cosine function, which offers two significant advantages. On one hand, the cosine is periodic and non-local, thereby capturing the inherently periodic nonlinearity patterns and the temporal dependency of dFCs over time. On the other hand, any derivative of a cosine function is itself a composition of cosine functions, as the derivative of the cosine function is a sine function, *i.e.,* a phase-shifted cosine function. Therefore, the derivatives of dCSL inherit the properties of dCSL, enabling us to explore the higher-order temporal features and supervise any derivative of dCSL with the complex dFCs.

After learning the higher-order temporal features of dFCs, the subsequent step is to perform the detection task with the following empirical loss:

$$\mathcal{L} = \frac{1}{R}\sum_{r=1}^{R}\ell(F(f^l(\boldsymbol{X}_r)), y_r) \tag{5}$$

where $\boldsymbol{X}_r$ denotes the explored higher-order temporal features by dCSL and y_r denotes the label. R is the number of subjects. F denotes a neural network with 4 fully connected layers. ℓ is the loss function, and it is selected as the cross entropy in this paper.

4. Experiment

4.1. Data acquisition and pre-processing

In this section, two public datasets, including Alzheimer's Disease Neuroimaging Initiative (ADNI)[1] and Autism Brain Imaging Data Exchange (ABIDE)[2] are involved in the experiment to systematically evaluate the proposed dCSL. The detailed information of these datasets is shown in Table 1.

For the ADNI dataset, 319 subjects/samples, including 165 MCIs with labeled "+1" and 154 NCs with labeled "− 1", are used for testing the proposed method under seven performance metrics. All subjects were scanned by Philips 3.0T scanner with the following imaging parameters: flip angle = 80, TR/TE = 3000 ms/300 ms, voxel thickness = 3.3 mm, and image matrix = 64 × 64. The scanning lasted 7 min, which generated 140 volumes for each subject.

For each subject in ADNI, SPM8[3] toolbox and DPABI [39] are used to preprocess their rs-fMRI data. The main steps include: (1) removing the first 3 volumes to keep the signal stabilization; (2) correcting head motion and slice timing for the remaining volumes; (3) regressing out the nuisance signals based on the Friston 24-parameters of head motion; (4) registering the corrected rs-fMRI images to Montreal Neurological Institute (MNI) standard space [40] and spatially smoothing the images by the full-width-half-maximum (FWHM) of Gaussian kernel, in which setting the FWHM to be 6 mm × 6 mm × 6 mm; (5) filtering the signals using band-pass frequencies between 0.01 and 0.1. Lastly, the BOLD signals from the ROIs are extracted according to an atlas and are put into the data matrix $\boldsymbol{D}$.

For the ABIDE dataset, 260 subjects/samples are randomly selected from the NYU site, including 125 ASDs with labeled "+1" and 135 NCs with labeled "−1". All rs-fMRI data were acquired based on a standard echo-planar imaging sequence on a clinical routine 3.0T Allegra scanner with the following imaging parameters: TR/TE is 2000 ms/15 ms with 180 volumes, the number of slices is 33, and the slice thickness is 4.0 mm.

For the subjects in the ABIDE dataset, the data pre-processing pipeline is similar to the ADNI dataset, and the details can be found in [41].

4.2. Experimental setting and implementation details

All the experiments are implemented with PyTorch [42] on a workstation with NVIDIA RTX 3090 GPU, AMD R7-5700X 3.40 GHz 8-core CPU, and 32 GB memory. For each dataset, 90% are used for training and 10% are used for testing. The unified fully connected architecture $\frac{N\times(N-1)}{2} \times 512 \times 32 \times 2$ are used to identify the disorders from the NCs based on the higher-order features, and the higher-order temporal features are extracted by the spectral learning with three layers $T\times T\times 1$. All the layers followed by rectified linear unit (ReLU) activation and 0.5 dropout. Adam algorithm and cross-entropy loss are selected as the optimization algorithm and loss function, respectively. In addition, the hyper-parameters, including the width and step size of SW, are involved in the dFCs estimation module, and the candidate values range in $[30, 35, 40, 45, 50]$ and $[1, 2, 3]$, respectively.

4.2.1. Compared methods

To verify the effectiveness of the proposed method, compare the proposed dCSL with some deep models that are commonly used to analyze dFCs, including:

RFF: Random Fourier feature is a method of approximating a kernel function. It also combines the Fourier transform with the kernel methods. By contrast, RFF just takes the transition invariant kernels, and the result is only dependent on the distance between samples.

[1] https://adni.loni.usc.edu/.
[2] https://fcon_1000.projects.nitrc.org/indi/abide/.
[3] http://www.fil.ion.ucl.ac.uk/spm/.

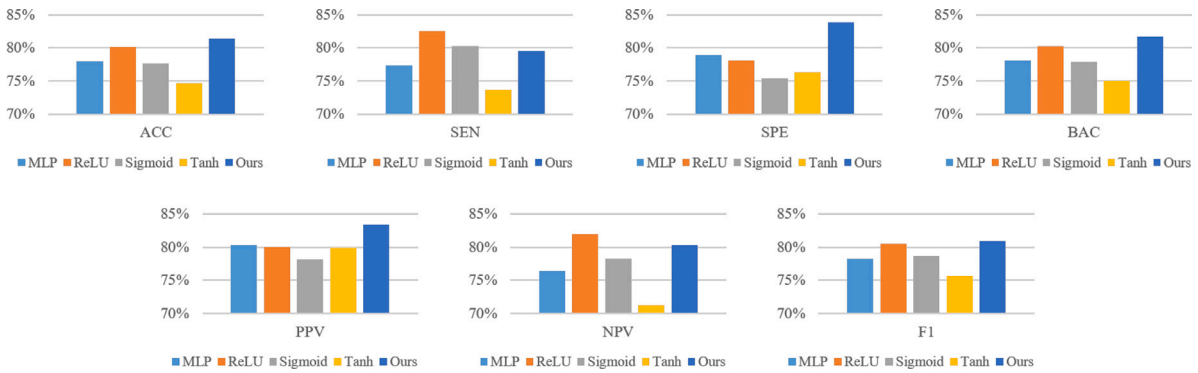


Fig. 4. The results on the MCI detection under different cases, including the traditional MLP model with ReLU activation functions, as well as the neural networks with the same architecture as dCSL but with different activation functions (*i.e.,* ReLU, Sigmoid, and Tanh).

Table 2
Definitions for different measurements. TP is the number of positive subjects that are predicted correctly. FN is the number of negative subjects that are predicted incorrectly. TN and FP are the numbers of their corresponding subjects, respectively.

Quantitative measurements	Definition
Accuracy (ACC)	$\frac{TP+TN}{TP+FP+TN+FN}$
Sensitivity (SEN)	$\frac{TP}{TP+FN}$
Specificity (SPE)	$\frac{TN}{FP+TN}$
Balanced Accuracy (BAC)	$\frac{SEN+SPE}{2}$
Positive Predictive Value (PPV)	$\frac{TP}{TP+FP}$
Negative Predictive Value (NPV)	$\frac{TN}{TN+FN}$
F1	$\frac{TP+TN}{TP+TP+FP+FN}$

LSTM: LSTM network is the most popular method to analyze the time sequence. Compared with the RNN method, it can remove the useless information by the forget gate to capture the long-range dependence. In this method, two steps are also adopted. The first step (i.e., FCs estimation) is the same as the proposal. Then LSTM is used to analyze the dFCs. Specifically, a sequence with T time points is input to the LSTM network, and each time point includes $\frac{N\times(N-1)}{2}$ features. The output of the last moment is considered as the extracted features to perform the classification task.

GRU: The GRU method is the variant of LSTM network. It also can capture the long-range dependence of the time sequence. Similar to the framework of LSTM method, the dFCs sequence with T time points is analyzed by the GRU network, and the output of the last moment is also considered as the extracted features to identify the subjects with the disorders from the NCs.

CRNN: CRNN is a variant of CNN. In this method, CNN is used to extract the spatial information, and then LSTM is used to extract the temporal information with the dFC networks.

SA-CRN [43]: Self-attention (SA) based convolutional recurrent network (SA-CRN). Pooling strengths of CNN, LSTM, and self-attention to extract higher-order sequence information of dFCs.

BolT [44]: Blood-oxygen-level-dependent transformer (BolT). BolT leverages a cascade of transformer encoders equipped with a novel fused window attention mechanism. Encoding is performed on temporally-overlapped windows within the time series to capture local representations. Cross-window attention is computed between base tokens in each window and fringe tokens from neighboring windows to integrate the temporal information.

MDGL [45]: MDGL capture multi-scale spatiotemporal dynamic representations of rs-fMRI data for automated brain disorder diagnosis. The MDGL framework consists of three major components: (1) multi-scale dynamic FCN construction using multiple brain atlases to model multi-scale topological information, (2) multi-scale dynamic graph representation learning to capture spatiotemporal information conveyed in fMRI data, and (3) multi-scale feature fusion and classification.

BrainTGL [46]: BrainTGL model exploits the temporal characteristics in rs-fMRI data through the proposed attention-based graph pooling for removing noisy edges and dual temporal graph learning LSTM for learning temporal characteristics in fMRI data from two aspects.

BrainGB [47]: BrainGB standardizes the process by (1) summarizing brain network construction pipelines for both functional and structural neuroimaging modalities and (2) modularizing the implementation of Graph Neural Networks designs.

BrainGNN [48]: BrainGNN includes (1) novel ROI-aware graph convolutional layers that efficiently assign each ROI a unique kernel that reflects ROI community patterns, and (2) novel regularization terms (unit loss, group-level consistency loss and topK pooling loss) for pooling operations to encourage reasonable ROI-selection and provide flexibility to encourage either fully individual- or patterns that agree with group-level data.

MVS-GCN [49]: MVS-GCN collaborates the graph structure learning and multi-task graph embedding learning to improve the classification performance and identify the potential functional subnetworks.

SHeC [50]: SHeC adaptively assigns heterogeneous connections to different pairs of brain regions, thereby effectively encoding the complex interaction patterns in the brain.

DART [51]: DART uses the static brain network as a baseline, integrating dynamic brain networks to enhance performance against traditional methods, which innovatively employ attention mechanisms, enhancing model explainability and exploiting the dynamic brain network's temporal variations.

4.2.2. Performance evaluate

For evaluating the performance of the proposal, a set of quantitative measurements are calculated, including classification accuracy (ACC), sensitivity (SEN), specificity (SPE), balanced accuracy (BAC), positive predictive value (PPV), negative predictive value (NPV), and F1. Their definitions are given in Table 2, where TP is the number of positive subjects that are predicted correctly in the classification task, FN is the number of negative subjects that are predicted incorrectly, and similarly, TN and FP are the numbers of their corresponding subjects, respectively.

4.3. Results and analysis

4.3.1. Results of ASD detection

In this experiment, the hyper-parameters for all the methods are set as the width and step size of SWs are 30 and 2, respectively, learning rate = 0.01, epoch = 200. Moreover, the detection results are reported in Table 3.

For the ASD detection task, the proposed method achieves the average results of ACC = 71.54%, SEN = 68.35%, SPE = 74.52%, BAC = 71.44%, PPV = 70.47%, NPV = 73.20%, and F1 = 68.69%. From Table 3, it can be observed that the proposed dCSL performs better

Table 3
Detection results. / denotes that the metric has not been calculated. Best results are highlighted in **bold** and suboptimal results are highlighted in underline.

Dataset	Method	ACC	SEN	SPE	BAC	PPV	NPV	F1
ABIDE	GRU	60.77%	60.59%	60.53%	60.56%	57.90%	63.36%	58.86%
	LSTM	63.08%	61.26%	65.23%	63.25%	59.18%	67.86%	59.07%
	RFF	65.58%	65.47%	65.97%	65.72%	61.58%	69.90%	62.90%
	CRNN	68.08%	65.99%	69.90%	67.95%	66.45%	69.53%	65.47%
	SA-CRN	62.12%	60.73%	63.22%	61.98%	60.22%	63.85%	59.93%
	BolT	70.76%	64.10%	/	70.59%	**72.35%**	/	66.44%
	MDGL	60.88%	60.03%	/	60.76%	58.29%	/	57.83%
	MVS-GCN	50.00%	36.00%	/	/	47.37%	/	48.79%
	DART	65.38%	65.38%	65.38%	/	/	/	**72.93%**
	BrainTGL	57.96%	72.10%	54.50%	66.30%	32.96%	**85.55%**	43.26%
	SHeC	56.54%	55.20%	57.78%	56.49%	54.76%	58.21%	54.98%
	BrainGB	58.46%	54.40%	/	58.31%	57.78%	/	55.30%
	BrainGNN	55.77%	**78.95%**	/	60.69%	44.12%	/	56.60%
	dCSL (Ours)	**71.54%**	68.35%	**74.52%**	**71.44%**	70.47%	73.20%	68.69%
ADNI	GRU	73.28%	73.00%	74.69%	73.84%	76.18%	72.77%	73.46%
	LSTM	71.25%	70.76%	72.67%	71.71%	77.36%	65.74%	73.38%
	RFF	70.94%	70.19%	72.42%	71.30%	73.73%	69.37%	71.17%
	CRNN	77.81%	79.69%	76.13%	77.91%	81.18%	75.28%	79.65%
	SA-CRN	63.75%	67.08%	61.28%	64.18%	64.74%	63.81%	65.09%
	BolT	80.12%	81.80%	/	80.25%	80.74%	/	81.10%
	MDGL	78.89%	**89.78%**	/	78.90%	75.68%	/	81.83%
	MVS-GCN	58.73%	50.00%	/	/	57.69%	/	58.21%
	DART	76.92%	67.64%	**87.10%**	/	/	/	**86.43%**
	BrainTGL	67.03%	65.03%	68.20%	66.61%	67.20%	65.75%	65.29%
	SHeC	56.93%	52.94%	60.87%	56.91%	57.14%	56.76%	54.96%
	BrainGB	72.86%	77.74%	/	72.69%	69.89%	/	73.09%
	BrainGNN	73.08%	47.06%	/	66.39%	61.54%	/	53.33%
	dCSL (Ours)	**81.41%**	79.55%	83.86%	**81.71%**	**83.33%**	**80.25%**	80.95%

Table 4
Detection results on different atlases. / denotes that the metric has not been calculated.

Methods	Atlas	ACC	SEN	SPE	BAC	PPV	NPV	F1
dCSL	AAL	88.21%	86.21%	89.85%	88.03%	89.57%	88.85%	86.47%
	fcMVPA	89.29%	88.59%	89.13%	88.86%	91.93%	89.10%	89.42%
	Power's 264	92.14%	95.36%	87.11%	91.23%	91.02%	95.33%	92.74%
	Schaefer's 400	90.71%	87.92%	93.30%	90.61%	93.76%	89.59%	89.86%
SRFF	AAL	86.43%	84.58%	87.53%	86.06%	86.78%	88.36%	84.74%
	fcMVPA	83.93%	82.56%	85.85%	84.21%	87.11%	82.91%	83.33%
	Power's 264	85.00%	80.62%	88.51%	84.57%	89.07%	85.20%	82.84%
	Schaefer's 400	84.64%	86.10%	84.39%	85.25%	85.82%	86.10%	84.82%
BolT	AAL	80.16%	77.38%	/	79.88%	82.64%	/	78.66%
	fcMVPA	86.26%	86.90%	/	86.31%	88.64%	/	86.20%
	Power's 264	89.12%	90.05%	/	89.04%	88.48%	/	89.25%
	Schaefer's 400	88.35%	86.90%	/	88.45%	90.47%	/	88.06%

than the baseline methods on all metrics. Specifically, the proposal is significantly better on the metrics of SEN and SPE, which means that the proposal can not only improve the classification accuracy but is also sensitive to both disorders and NCs.

4.3.2. Results of MCI detection

In this experiment, the hyper-parameters are set as the width and step size of SWs are 30 and 2, respectively, learning rate = 0.01, epoch = 200. Moreover, the detection results are reported in Table 3.

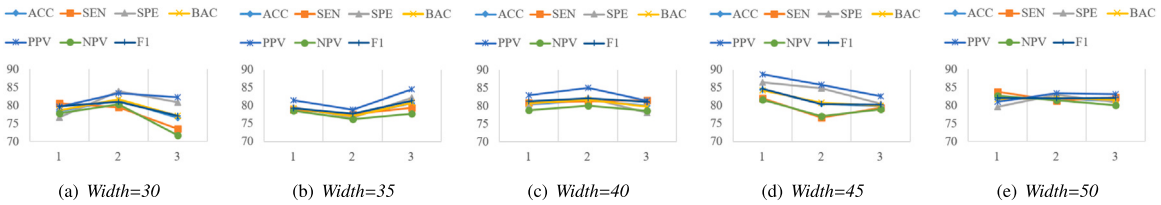
For this detection task, the proposed method achieves the detection results of ACC = 81.41%, SEN = 79.55%, SPE = 83.86%, BAC = 81.71%, PPV = 83.33%, NPV = 80.25%, and F1 = 80.95%. The results, reported in Table 3, show that dCSL obtains a more competitive performance than the baseline methods on most of the evaluation metrics, meaning that the spectral kernel network, as the temporal features extractor, can capture the dynamics of dFCs.

4.4. Ablation study

As shown in Fig. 1, spectral learning is the core of the proposed dCSL, which is constructed by stacking the non-stationary spectral mapping. Spectral learning shares the advantages of deep hierarchical structure and the non-stationary kernel, enabling us to explore the long-range correlation and higher-order temporal patterns of dFCs. An ablation study is conducted to evaluate the methods with different cases to demonstrate that performance improvement comes from the powerful representation brought by spectral learning. The case includes the traditional MLP model with ReLU activation functions, as well as the neural networks with the same architecture as spectral learning in dCSL but with different activation functions. All the results, reported in Fig. 4, show that the spectral learning in dCSL generally performs better than other cases. Specifically, spectral learning outperforms the traditional MLP model, which may be due to the increasing of parameters. Furthermore, spectral learning is commonly superior to neural networks with the same architecture as spectral learning (the same parameter numbers), which contributes to the capability of capturing the long-range correlation and exploring the higher-order temporal patterns of spectral learning.

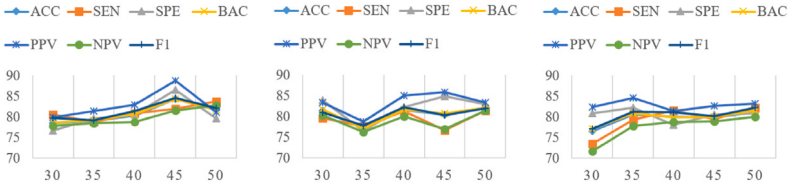
5. Discussion

In this section, this study focuses on the ABIDE dataset to analyze the influence of the width and step size of the SW method on the results and explore which brain regions are associated with ASD detection



(a) *Width=30* (b) *Width=35* (c) *Width=40* (d) *Width=45* (e) *Width=50*

Fig. 5. The classification results with fixed width of SWs.



(a) *Step size=1* (b) *Step size=2* (c) *Step size=3*

Fig. 6. The classification results with fixed step size of SWs.

tasks. In addition, the limitations of dCSL and the possible future research are discussed.

5.1. Parameter analysis

In addition to the hyper-parameters related to the deep structure, two additional parameters, including the width and step size of the SW method, are to be tuned in the proposed method. In this section, the influence of these two parameters on the final detection results achieved by dCSL will be discussed. As discussed in the experimental setting section, parameters of the SW method are selected from the candidate values range in $[30, 35, 40, 45, 50]$ for the width and $[1, 2, 3]$ for the step size, respectively. Note that the other parameters, involved in the deep structure (i.e., the batch size, epoch, and learning rate), are the same as the above experimental settings. The detection results under different hyper-parameters are presented in Figs. 5 and 6.

In Fig. 5, maintaining a fixed width and investigating the influence of the step size on the detection performance. Simultaneously, investigating the influence of the width while maintaining a fixed step size in Fig. 6. From Figs. 5 and 6, it can be observed that both the width and step size have notable effects on the results. Notably, the influence of the step size tends to decrease as the width increases. This trend could be attributed to the fact that wider windows contain more intricate patterns. Compared with the step size, the width seems to exert more influence on the detection results. As shown in Fig. 6, the best width tends to decrease with the step size increasing, suggesting a dynamic relationship between these parameters that requires careful consideration for optimal performance.

5.2. Parcellation and important brain regions

Besides the hyper-parameter selection, such as the width and step size of the SW method in the dFCs estimation component, the schemes of brain regions parcellation and BOLD signals extraction in the data pre-processing step also play an important role in brain disease detection. Currently, researchers have proposed many schemes to parcel the human brain, which can be roughly divided into atlas-based and data-driven methods. Atlas-based parcellation includes Automated Anatomical Labeling (AAL) [52], anatomical Harvard Oxford (HO) [53], while functional-based atlas includes functional connectivity multivariate pattern analysis (fcMVPA) [54], and Power's 264 functional ROIs [55]. Primarily, AAL atlas and fcMVPA atlas are commonly used structure-atlas and functional-atlas, respectively.

As an alternative to atlas-based methods, data-driven methods directly work on the present dataset. Clustering and group independent component analysis (GICA) are the commonly used methods for defining ROIs. The clustering method, groups voxels together such that all the voxels in one parcel are more similar to each other than those in other parcels. For GICA, it aims to decompose the present data into a set of independent components, which are described by spatial maps, and each map corresponds to an ROI.

In order to investigate the influence of the atlas selection on the final results, four atlases, including AAL, fcMVPA, Power's 264, Schaefer's 400 [56] are used for performing the AD detection task. The results, reported in Table 4 demonstrate that these four methods with the functional-based atlas perform better than with the structure-based atlas, and there is only a slight difference between two functional-based atlas (i.e., fcMVPA atlas and Power's 264 atlas). Therefore, it can be concluded that the functional-based atlas may be more appropriate for the functional connections analysis. It also can be observed from Table 4, Schaefer's 400, as the functional-based atlases, does not perform as well as the other two functional-based atlas. One of the possible reasons is overfitting. There are $\frac{400\times(400-1)}{2}$ features for each subject, and the used dataset only included 137 subjects. All in all, the performance of different approaches (or atlases) may vary under different conditions. No atlas is always optimal for ROI definition and brain disease detection tasks. It is worth noting that the proposed dCSL is consistently superior to other methods for different atlases.

Except for the detection performance, researchers also want to know the critical features that are associated with the task in the medical field [57–62]. Therefore, to determine which brain regions contribute more to the ASD detection task, the deep neural network with a single layer is employed as a classifier at the bottom of dCSL. Specifically, the top 50 FCs are selected, whose temporal features contribute more to the ASD classification task. These connections are associated with 15 discriminative brain regions, including the middle frontal gyrus, inferior frontal gyrus, middle cingulate gyrus, posterior cingulate gyrus, parahippocampal, amygdala, calcarine cortex, cuneus, superior parietal gyrus, heschl gyrus, temporal pole, orbitofrontal cortex, and two cerebellar regions. These brain regions are visualized in Fig. 7.

Most of the discriminative brain regions are related to the default mode network (DMN) [63], executive control network (ECN) [64], and visual network (VN) [65]. Previous studies have found that the abnormal activation and connections in these networks are most likely to cause ASD [66,67]. Remarkably, the found discriminative brain regions, including the inferior frontal gyrus, amygdala, parahippocampal, middle cingulate gyrus, and posterior cingulate gyrus, are related to memory, social cognition, and emotion processing [68–70]. These findings in this study are consistent with the previous studies and the clinical performance in ASD.

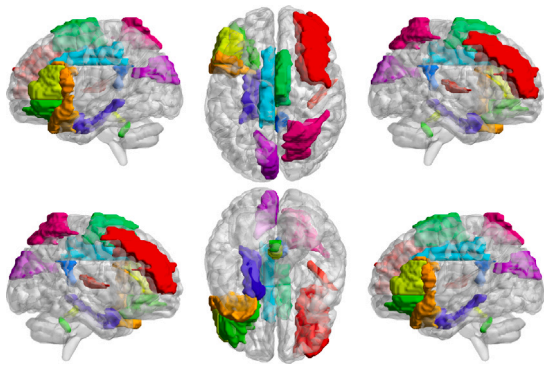


Fig. 7. The important brain regions that are associated with the ASD classification task.

5.3. Limitations and future work

As the detection results show, the proposed method outperforms existing deep learning models in most cases. However, there are several technical issues that warrant attention in the future.

Firstly, in the proposed dCSL, the use of the cosine function as the activation function presents an optimization challenge (it is easy to fall into a local optimum) that is caused by the periodicity of the cosine function. Therefore, a non-stationary monotone activation function will be studied in future work.

Secondly, despite the integration of dynamic feature extraction and classifier training into a unified framework through the proposed approach, dFCs estimation remains independent of this unified framework. In the future, an end-to-end framework will be studied for joint dFCs estimation, temporal features extraction, and classifier training with the aim that dFCs estimation is optimized and directly contributes to the effectiveness of the classifier training process.

In addition, although rs-fMRI, by achieving BOLD signals, provides a non-invasive and valuable way of exploring the human brain [8], BOLD signals are arbitrarily scaled and have no unit, resulting in the difficulty of comparing them directly across different subjects [71]. As discussed in 1, dCSL can effectively reveal the input-dependent characteristics and the potential temporal dependence by compactly representing higher-order temporal features. Therefore, an attempt will be made to directly capture the potential dynamics of the BOLD signals based on the spectral kernel network, and further identify the disorders from NCs.

6. Conclusion

This study proposes a novel analysis method called dCSL to explore the temporal patterns of dFCs. In the proposal, dFCs are first estimated through the SW method. Then, spectral learning is constructed via stacking the non-stationary spectral kernel mapping into a deep neural network to reveal the dynamic patterns of dFCs. To investigate the effectiveness of dCSL, a set of experiments is performed on two public datasets. The results show that the proposed method is better than the baseline methods commonly used for dFCs analysis. Therefore, it is empirically concluded that the deep neural network with non-stationary spectral kernel mapping can mine more dynamic characteristics of dFCs.

CRediT authorship contribution statement

Yanfang Xue: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Formal analysis, Data curation. **Hui Xue:** Funding acquisition. **Pengfei Fang:** Conceptualization. **Shipeng Zhu:** Writing – review & editing. **Lishan Qiao:** Writing – review & editing, Methodology. **Yuexuan An:** Writing – review & editing.

Declaration of competing interest

None

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62476056, 62076062, and 62306070) and the Social Development Science and Technology Project of Jiangsu Province (No. BE2022811). Furthermore, the work was also supported by the Big Data Computing Center of Southeast University.

Appendix

In this section, we show more details about the non-stationary spectral kernel mapping in (2).

Note that the definition in Eq. (1) can result in a complex-value kernel due to the exponential component $e^{i(\boldsymbol{\omega}^\top \boldsymbol{x}-\boldsymbol{\omega}'^\top \boldsymbol{x}')}$. In order to derive the real-value kernel k, we use the following $\zeta_{\boldsymbol{\omega},\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{x}')$ to replace the exponential component, and $\zeta_{\boldsymbol{\omega},\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{x}')$ is defined as:

$$\begin{aligned}\zeta_{\boldsymbol{\omega},\boldsymbol{\omega}'}(x,x') = \frac{1}{8}\Big[&e^{i(\boldsymbol{\omega}^T\boldsymbol{x}-\boldsymbol{\omega}'^T\boldsymbol{x}')} + e^{i(-\boldsymbol{\omega}^T\boldsymbol{x}+\boldsymbol{\omega}'^T\boldsymbol{x}')}\\ &+e^{i(\boldsymbol{\omega}'^T\boldsymbol{x}-\boldsymbol{\omega}^T\boldsymbol{x}')} + e^{i(-\boldsymbol{\omega}'^T\boldsymbol{x}+\boldsymbol{\omega}^T\boldsymbol{x}')}\\ &+e^{i(\boldsymbol{\omega}^T\boldsymbol{x}-\boldsymbol{\omega}^T\boldsymbol{x}')} + e^{i(-\boldsymbol{\omega}^T\boldsymbol{x}+\boldsymbol{\omega}^T\boldsymbol{x}')}\\ &+e^{i(\boldsymbol{\omega}'^T\boldsymbol{x}-\boldsymbol{\omega}'^T\boldsymbol{x}')} + e^{i(-\boldsymbol{\omega}'^T\boldsymbol{x}+\boldsymbol{\omega}'^T\boldsymbol{x}')}\Big].\end{aligned} \tag{6}$$

Similarity, the spectral density $s(\boldsymbol{\omega},\boldsymbol{\omega}')$ is replaced with the following $p(\boldsymbol{\omega},\boldsymbol{\omega}')$:

$$\begin{aligned}p(\boldsymbol{\omega},\boldsymbol{\omega}') = \frac{1}{8}\Big[&s(\boldsymbol{\omega},\boldsymbol{\omega}') + s(-\boldsymbol{\omega},-\boldsymbol{\omega}')\\ &+s(\boldsymbol{\omega}',\boldsymbol{\omega}) + s(-\boldsymbol{\omega}',-\boldsymbol{\omega})\\ &+\int_{\mathbb{R}^T} s(\boldsymbol{\omega},\boldsymbol{\omega}')d\boldsymbol{\omega} + \int_{\mathbb{R}^T} s(-\boldsymbol{\omega},-\boldsymbol{\omega}')d\boldsymbol{\omega}\\ &+\int_{\mathbb{R}^T} s(\boldsymbol{\omega},\boldsymbol{\omega}')d\boldsymbol{\omega}' + \int_{\mathbb{R}^T} s(-\boldsymbol{\omega},-\boldsymbol{\omega}')d\boldsymbol{\omega}'\Big],\end{aligned} \tag{7}$$

where the spectral density $p(\boldsymbol{\omega},\boldsymbol{\omega}')$ can be seen as a probability density function.

Considering the following equations,

$$\begin{aligned}&e^{i(\boldsymbol{\omega}^T\boldsymbol{x}-\boldsymbol{\omega}'^T\boldsymbol{x}')} + e^{i(-\boldsymbol{\omega}^T\boldsymbol{x}+\boldsymbol{\omega}'^T\boldsymbol{x}')}\\ =&2\cos(\boldsymbol{\omega}^T\boldsymbol{x}-\boldsymbol{\omega}'^T\boldsymbol{x}')\\ =&4\mathbb{E}_{\varphi\sim[-\pi,\pi]}\Big[\cos(\boldsymbol{\omega}^T\boldsymbol{x}+\varphi)\cos(\boldsymbol{\omega}'^T\boldsymbol{x}'+\varphi)\Big].\end{aligned} \tag{8}$$

$$\begin{aligned}k(\boldsymbol{x},\boldsymbol{x}') &= \int_{\mathbb{R}^T\times\mathbb{R}^T} \zeta_{\boldsymbol{\omega},\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{x}')p(\boldsymbol{\omega},\boldsymbol{\omega}')d\boldsymbol{\omega}d\boldsymbol{\omega}'\\ &= \int_{\mathbb{R}^T\times\mathbb{R}^T} \frac{1}{8}\Big[2\cos(\boldsymbol{\omega}^T\boldsymbol{x}-\boldsymbol{\omega}'^T\boldsymbol{x}')\\ &+ 2\cos(\boldsymbol{\omega}'^T\boldsymbol{x}-\boldsymbol{\omega}^T\boldsymbol{x}')\\ &+ 2\cos(\boldsymbol{\omega}^T\boldsymbol{x}-\boldsymbol{\omega}^T\boldsymbol{x}')\\ &+ 2\cos(\boldsymbol{\omega}'^T\boldsymbol{x}-\boldsymbol{\omega}'^T\boldsymbol{x}')\Big]p(\boldsymbol{\omega},\boldsymbol{\omega}')d\boldsymbol{\omega}d\boldsymbol{\omega}'\\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} \Psi_{\boldsymbol{\omega},\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{x}')p(\boldsymbol{\omega},\boldsymbol{\omega}')d\boldsymbol{\omega}d\boldsymbol{\omega}'\\ &= \mathbb{E}_{(\boldsymbol{\omega},\boldsymbol{\omega}')\sim P}\Big[\Psi_{\boldsymbol{\omega},\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{x}')\Big]\end{aligned} \tag{9}$$

where

$$\begin{aligned}\Psi_{\boldsymbol{\omega},\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{x}') = \frac{1}{2}\mathbb{E}_{\varphi\sim[-\pi,\pi]}\Big[&\cos(\boldsymbol{\omega}^T\boldsymbol{x}+\varphi)\cos(\boldsymbol{\omega}'^T\boldsymbol{x}'+\varphi)\\ &+\cos(\boldsymbol{\omega}'^T\boldsymbol{x}+\varphi)\cos(\boldsymbol{\omega}^T\boldsymbol{x}'+\varphi)\\ &+\cos(\boldsymbol{\omega}^T\boldsymbol{x}+\varphi)\cos(\boldsymbol{\omega}^T\boldsymbol{x}'+\varphi)\\ &+\cos(\boldsymbol{\omega}'^T\boldsymbol{x}+\varphi)\cos(\boldsymbol{\omega}'^T\boldsymbol{x}'+\varphi)\Big].\end{aligned} \tag{10}$$

Discretizing $k(\boldsymbol{x}, \boldsymbol{x}')$ in Eq. (9) by Monte Carlo integral:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}_{\boldsymbol{\omega}, \boldsymbol{\omega}' \sim P}\left[\Psi_{\boldsymbol{\omega}, \boldsymbol{\omega}'}(\boldsymbol{x}, \boldsymbol{x}')\right] \approx \langle \boldsymbol{\Phi}(\boldsymbol{x}), \boldsymbol{\Phi}(\boldsymbol{x}') \rangle \tag{11}$$

where the non-stationary spectral kernel mapping $\boldsymbol{\Phi}(\boldsymbol{x})$ is defined as follows:

$$\begin{aligned} \boldsymbol{\Phi}(\boldsymbol{x}) &= \sqrt{\frac{1}{2M}} \begin{bmatrix} \cos(\boldsymbol{\omega}_1^T \boldsymbol{x} + \varphi_1) + \cos(\boldsymbol{\omega}_1'^T \boldsymbol{x} + \varphi_1) \\ \cdots \\ \cos(\boldsymbol{\omega}_M^T \boldsymbol{x} + \varphi_M) + \cos(\boldsymbol{\omega}_M'^T \boldsymbol{x} + \varphi_M) \end{bmatrix} \\ &= \sqrt{\frac{1}{2M}} \left[cos(\boldsymbol{\Omega}^T \boldsymbol{x} + \varphi) + cos(\boldsymbol{\Omega}'^T \boldsymbol{x} + \varphi) \right] \end{aligned} \tag{12}$$

where $\boldsymbol{\Omega} = [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_M]$ and $\boldsymbol{\Omega}' = [\boldsymbol{\omega}_1', \boldsymbol{\omega}_2', \ldots, \boldsymbol{\omega}_M']$ are the frequency matrices. The frequency pairs $\{(\boldsymbol{\omega}_i, \boldsymbol{\omega}_i')\}_{i=1}^M$ are sampled from the spectral density $s(\boldsymbol{\omega}, \boldsymbol{\omega}')$. M is the sampling number.

References

[1] Tarawneh R, Holtzman DM. The clinical problem of symptomatic Alzheimer Disease and mild cognitive impairment. Cold Spring Harbor Perspect Med 2012;2(5):a006148.

[2] Wang J, Zuo X, Dai Z, Xia M, Zhao Z, Zhao X, Jia J, Han Y, He Y. Disrupted functional brain connectome in individuals at risk for Alzheimer's Disease. Biol Psychiatry 2013;73(5):472–81.

[3] Anagnostou E, Taylor MJ. Review of neuroimaging in autism spectrum disorders: what have we learned and where we go from here. Mol Autism 2011;2(1):4.

[4] Wang M, Zhang D, Huang J, Yap PT, Liu M. Identifying autism spectrum disorder with multi-site fMRI via low-rank domain adaptation. IEEE Trans Med Imaging 2019;PP(99):1.

[5] Yan CG, Chen X, Li L, Castellanos FX, Zang YF. Reduced default mode network functional connectivity in patients with recurrent major depressive disorder. Proc Natl Acad Sci 2019;116(18):201900390.

[6] Schmaal L, Hibar D, Sämann PG, Hall G, Baune B, Jahanshad N, Cheung J, van Erp TG, Bos D, Ikram MA, et al. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. Mol Psychiatry 2016;22(6):900–9.

[7] Martinez B, Peplow PV. MicroRNAs as diagnostic and therapeutic tools for Alzheimer's disease: advances and limitations. Neural Regen Res 2019;14(2):242.

[8] Lee JH, Durand R, Gradinaru V, Zhang F, Goshen I, Kim DS, Fenno LE, Ramakrishnan C, Deisseroth K. Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. Nature 2010;465(7299):788–92.

[9] Brunetti M, Belardinelli P, Del Gratta C, Pizzella V, Penna SD, Ferretti A, Caulo M, Cianflone F, Olivetti Belardinelli M, Romani GL. Human brain activation elicited by the localization of sounds delivering at attended or unattended positions: an fMRI/MEG study. Cogn Process 2006;7(1):116–7.

[10] Monti RP, Hellyer P, Sharp D, Leech R, Anagnostopoulos C, Montana G. Estimating time-varying brain connectivity networks from functional MRI time series. Neuroimage 2014;103:427–43.

[11] Damaraju E, Allen EA, Belger A, Ford JM, McEwen S, Mathalon D, Mueller B, Pearlson G, Potkin S, Preda A, et al. Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. NeuroImage: Clin 2014;5:298–308.

[12] Chang C, Glover GH. Time–frequency dynamics of resting-state brain connectivity measured with fMRI. Neuroimage 2010;50(1):81–98.

[13] Li Y, Liu J, Jiang Y, Liu Y, Lei B. Virtual adversarial training-based deep feature aggregation network from dynamic effective connectivity for MCI identification. IEEE Trans Med Imaging 2021;41(1):237–51.

[14] Zhu Q, Xu R, Wang R, Xu X, Zhang Z, Zhang D. Stacked topological preserving dynamic brain networks representation and classification. IEEE Trans Med Imaging 2022;41(11):3473–84.

[15] Jie B, Liu M, Shen D. Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease. Med Image Anal 2018;47:81–94.

[16] Li Y, Liu J, Tang Z, Lei B. Deep spatial-temporal feature fusion from adaptive dynamic functional connectivity for MCI identification. IEEE Trans Med Imaging 2020;39(9):2818–30.

[17] Liang L, Yuan Y, Wei Y, Yu B, Mai W, Duan G, Nong X, Li C, Su J, Zhao L, et al. Recurrent and concurrent patterns of regional BOLD dynamics and functional connectivity dynamics in cognitive decline. Alzheimer's Res Ther 2021;13:1–12.

[18] Fiorenzato E, Strafella AP, Kim J, Schifano R, Weis L, Antonini A, Biundo R. Dynamic functional connectivity changes associated with dementia in Parkinson's disease. Brain 2019;142(9):2860–72.

[19] Liu F, Wang Y, Li M, Wang W, Li R, Zhang Z, Lu G, Chen H. Dynamic functional network connectivity in idiopathic generalized epilepsy with generalized tonic–clonic seizure. Hum Brain Mapp 2017;38(2):957–73.

[20] Jie B, Liu M, Lian C, Shi F, Shen D. Developing novel weighted correlation kernels for convolutional neural networks to extract hierarchical functional connectivities from fMRI for disease diagnosis. In: Machine learning in medical imaging: 9th international workshop, MLMI 2018, held in conjunction with MICCAI 2018, granada, Spain, September 16, 2018, proceedings 9. Springer; 2018, p. 1–9.

[21] Davey CE, Grayden DB, Johnston LA. Correcting for non-stationarity in BOLD-fMRI connectivity analyses. Front Neurosci 2021;15:574979.

[22] Li M, Gao Y, Anderson AW, Ding Z, Gore JC. Dynamic variations of resting-state BOLD signal spectra in white matter. Neuroimage 2022;250:118972.

[23] Yan C, Zhao S, Han W, Xie L, Liu T. Identifying brain networks at multiple time scales via deep recurrent neural network. IEEE J Biomed Health Inf 2018;PP(99):1.

[24] Wang M, Lian C, Yao D, Zhang D, Shen D. Spatial-temporal dependency modeling and network hub detection for functional MRI analysis via convolutional-recurrent network. IEEE Trans Bio-Med Eng 2019;PP(99).

[25] Lin K, Jie B, Dong P, Ding X, Bian W, Liu M. Convolutional recurrent neural network for dynamic functional mri analysis and brain disease identification. Front Neurosci 2022;16:933660.

[26] Lin Y-P, Wang C-H, Jung T-P, Wu T-L, Jeng S-K, Duann J-R, Chen J-H. EEG-based emotion recognition in music listening. IEEE Trans Biomed Eng 2010;57(7):1798–806.

[27] Li Y, Zheng W, Wang L, Zong Y, Cui Z. From regional to global brain: A novel hierarchical spatial-temporal neural network model for EEG emotion recognition. IEEE Trans Affect Comput 2022;13(2):568–78. http://dx.doi.org/10.1109/TAFFC.2019.2922912.

[28] Ning L, Rathi Y. A dynamic regression approach for frequency-domain partial coherence and causality analysis of functional brain networks. IEEE Trans Med Imaging 2018;37(9):1957–69. http://dx.doi.org/10.1109/TMI.2017.2739740.

[29] Emoto T, Akutagawa M, Abeyratne U, Nagashino H, Kinouchi Y. A comparison of neural network and fast Fourier transform-based approach for the state analysis of brain. In: 2005 international conference on neural networks and brain. Vol. 1, 2005, p. 94–9. http://dx.doi.org/10.1109/ICNNB.2005.1614575.

[30] Cao T, Lin R, Zheng Y, Shen D, Xu L. A novel approach analysing the dynamic brain functional connectivity for improved MCI detection. IEEE Trans Biomed Eng 2023.

[31] Remes S, Heinonen M, Kaski S. Non-stationary spectral kernels. Adv Neural Inf Process Syst 2017;30.

[32] Pham AT, Raich R. Kernel-based instance annotation in multi-instance multi-label learning. IEEE; 2014.

[33] Filice S, Castellucci G, Croce D, Basili R. Kelp: a kernel-based learning platform for natural language processing. In: Proceedings of ACL-IJCNLP 2015 system demonstrations. 2015, p. 19–24.

[34] Chapeau-Blondeau F, Belin E. Fourier-transform quantum phase estimation with quantum phase noise. Signal Process 2020;170:107441.

[35] Olbrys J, Mursztyn M. Measuring stock market resiliency with Discrete Fourier Transform for high frequency data. Phys A 2019;513:248–56. http://dx.doi.org/10.1016/j.physa.2018.09.028.

[36] Yaglom AM. Correlation theory of stationary and related random functions, Volume I: Basic results, vol. 131, Springer; 1987.

[37] Savva AD, Matsopoulos GK, Mitsis GD. A wavelet-based approach for estimating time-varying connectivity in resting-state functional magnetic resonance imaging. Brain Connect 2022;12(3):285–98.

[38] Zhuang X, Yang Z, Mishra V, Sreenivasan K, Bernick C, Cordes D. Single-scale time-dependent window-sizes in sliding-window dynamic functional connectivity analysis: a validation study. NeuroImage 2020;220:117111.

[39] Yan C, Zang Y. DPARSF: a MATLAB toolbox for" pipeline" data analysis of resting-state fMRI. Front Syst Neurosci 2010;4:13.

[40] Brett M, Christoff K, Cusack R, Lancaster J, et al. Using the Talairach atlas with the MNI template. Neuroimage 2001;13(6):85.

[41] Xue Y, Zhang Y, Zhang L, Lee SW, Shen D. Learning brain functional networks with latent temporal dependency for MCI identification. IEEE Trans Bio-Med Eng 2021;PP(99).

[42] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, et al. Pytorch: An imperative style, high-performance deep learning library. In: In advances in neural information processing systems. Vol. 32, Vancouver, BC, Canada; 2019, p. 8024–35.

[43] Zhang Z, Jie B, Wang Z, Zhou J, Yang Y. Self-attention based high order sequence feature reconstruction of dynamic functional connectivity networks with rs-fMRI for brain disease classification. 2022, arXiv preprint arXiv:2211.11750.

[44] Bedel HA, Sivgin I, Dalmaz O, Dar SU, Çukur T. Bolt: Fused window transformers for fMRI time series analysis. Med Image Anal 2023;88:102841.

[45] Ma Y, Wang Q, Cao L, Li L, Zhang C, Qiao L, Liu M. Multi-scale dynamic graph learning for brain disorder detection with functional MRI. IEEE Trans Neural Syst Rehabil Eng 2023.

[46] Liu L, Wen G, Cao P, Hong T, Yang J, Zhang X, Zaiane OR. BrainTGL: A dynamic graph representation learning model for brain network analysis. Comput Biol Med 2023;153:106521.

[47] Cui H, Dai W, Zhu Y, Kan X, Gu AAC, Lukemire J, Zhan L, He L, Guo Y, Yang C. Braingb: a benchmark for brain network analysis with graph neural networks. IEEE Trans Med Imaging 2022;42(2):493–506.

[48] Li X, Zhou Y, Dvornek N, Zhang M, Gao S, Zhuang J, Scheinost D, Staib LH, Ventola P, Duncan JS. Braingnn: Interpretable brain graph neural network for fmri analysis. Med Image Anal 2021;74:102233.
[49] Wen G, Cao P, Bao H, Yang W, Zheng T, Zaiane O. MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis. Comput Biol Med 2022;142:105239.
[50] Zhang C, Ma Y, Qiao L, Zhang L, Liu M. Learning functional brain networks with heterogeneous connectivities for brain disease identification. Neural Netw 2024;106660.
[51] Kan X, Gu AAC, Cui H, Guo Y, Yang C. Dynamic brain transformer with multi-level attention for functional brain network analysis. In: 2023 IEEE eMBS international conference on biomedical and health informatics. BHI, IEEE; 2023, p. 1–4.
[52] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 2002;15(1):273–89.
[53] Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006;31(3):968–80.
[54] Dosenbach NU, Nardos B, Cohen AL, Fair DA, Power JD, Church JA, Nelson SM, Wig GS, Vogel AC, Lessov-Schlaggar CN, et al. Prediction of individual brain maturity using fMRI. Science 2010;329(5997):1358–61.
[55] Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, et al. Functional network organization of the human brain. Neuron 2011;72(4):665–78.
[56] Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo X-N, Holmes AJ, Eickhoff SB, Yeo BT. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. Cereb Cortex 2018;28(9):3095–114.
[57] Wang Y, Xu Y, Yang Z, Liu X, Dai Q. Using recursive feature selection with random forest to improve protein structural class prediction for low-similarity sequences. Comput Math Methods Med 2021;2021(1):5529389.
[58] Guo X, Guo Y, Chen H, Liu X, He P, Li W, Zhang MQ, Dai Q. Systematic comparison of genome information processing and boundary recognition tools used for genomic island detection. Comput Biol Med 2023;166:107550.
[59] Dai Q, Bao C, Hai Y, Ma S, Zhou T, Wang C, Wang Y, Huo W, Liu X, Yao Y, et al. MTGIpick allows robust identification of genomic islands from a single genome. Brief Bioinform 2018;19(3):361–73.
[60] Yang Z, Yi W, Tao J, Liu X, Zhang MQ, Chen G, Dai Q. HPVMD-C: a disease-based mutation database of human papillomavirus in China. Database 2022;2022:baac018.
[61] Yang S, Wang Y, Chen Y, Dai Q. MASQC: next generation sequencing assists third generation sequencing for quality control in N6-methyladenine DNA identification. Front Genet 2020;11:269.
[62] Kong R, Xu X, Liu X, He P, Zhang MQ, Dai Q. 2SigFinder: the combined use of small-scale and large-scale statistical testing for genomic island detection from a single genome. BMC Bioinform 2020;21:1–15.
[63] Raichle ME. The brain's default mode network. Annu Rev Neurosci 2015;38:433–47.
[64] Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD. Dissociable intrinsic connectivity networks for salience processing and executive control. J Neurosci 2007;27(9):2349–56.
[65] Wang K, Jiang T, Yu C, Tian L, Li J, Liu Y, Zhou Y, Xu L, Song M, Li K. Spontaneous activity associated with primary visual cortex: a resting-state FMRI study. Cereb Cortex 2008;18(3):697–704.
[66] Padmanabhan A, Lynch CJ, Schaer M, Menon V. The default mode network in autism. Biol Psychiatry: Cogn Neurosci Neuroimaging 2017;2(6):476–86.
[67] Solomon M, Ozonoff SJ, Ursu S, Ravizza S, Cummings N, Ly S, Carter CS. The neural substrates of cognitive control deficits in autism spectrum disorders. Neuropsychologia 2009;47(12):2515–26.
[68] Ha S, Sohn I-J, Kim N, Sim HJ, Cheon K-A. Characteristics of brains in autism spectrum disorder: structure, function and connectivity across the lifespan. Exp Neurobiol 2015;24(4):273.
[69] Geschwind DH, Levitt P. Autism spectrum disorders: developmental disconnection syndromes. Curr Opin Neurobiol 2007;17(1):103–11.
[70] Herbert MR, Ziegler DA, Deutsch C, O'Brien LM, Kennedy DN, Filipek P, Bakardjiev A, Hodgson J, Takeoka M, Makris N, et al. Brain asymmetries in autism and developmental language disorder: a nested whole-brain analysis. Brain 2005;128(1):213–26.
[71] Bijsterbosch J, Beckmann C. An introduction to resting state fMRI functional connectivity. Oxford University Press; 2017.