Contents lists available at ScienceDirect

Medical Image Analysis

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

Spatio-temporal directed acyclic graph learning with attention mechanisms on brain functional time series and connectivity

Shih-Gu Huang[a,1], Jing Xia[a,1], Liyuan Xu[c], Anqi Qiu[a,b,c,d,e,f,1,*]

[a] Department of Biomedical Engineering, National University of Singapore, Singapore
[b] NUS (Suzhou) Research Institute, Suzhou, China
[c] School of Computer Engineering and Science, Shanghai University, China
[d] The N.1 Institute for Health, National University of Singapore, Singapore
[e] Institute of Data Science, National University of Singapore, Singapore
[f] Department of Biomedical Engineering, The Johns Hopkins University, USA

ARTICLE INFO

Article history:
Received 3 August 2021
Revised 8 January 2022
Accepted 11 January 2022
Available online 30 January 2022

Keywords:
Brain functional network
Directed acyclic graph
Graph neural network
Attention mechanism
Graph pooling
Multi-scale analysis

ABSTRACT

We develop a deep learning framework, spatio-temporal directed acyclic graph with attention mechanisms (ST-DAG-Att), to predict cognition and disease using functional magnetic resonance imaging (fMRI). This ST-DAG-Att framework comprises of two neural networks, (1) spatio-temporal graph convolutional network (ST-graph-conv) to learn the spatial and temporal information of functional time series at multiple temporal and spatial graph scales, where the graph is represented by the brain functional network, the spatial convolution is over the space of this graph, and the temporal convolution is over the time dimension; (2) functional connectivity convolutional network (FC-conv) to learn functional connectivity features, where the functional connectivity is derived from embedded multi-scale fMRI time series and the convolutional operation is applied along both edge and node dimensions of the brain functional network. This framework also consists of an attention component, i.e., functional connectivity-based spatial attention (FC-SAtt), that generates a spatial attention map through learning the local dependency among high-level features of functional connectivity and emphasizing meaningful brain regions. Moreover, both the ST-graph-conv and FC-conv networks are designed as feed-forward models structured as directed acyclic graphs (DAGs). Our experiments employ two large-scale datasets, Adolescent Brain Cognitive Development (ABCD, $n = 7693$) and Open Access Series of Imaging Study-3 (OASIS-3, $n = 1786$). Our results show that the ST-DAG-Att model is generalizable from cognition prediction to age prediction. It is robust to independent samples obtained from different sites of the ABCD study. It outperforms the existing machine learning techniques, including support vector regression (SVR), elastic net's mixture with random forest, spatio-temporal graph convolution, and BrainNetCNN.

1. Introduction

Functional magnetic resonance imaging (fMRI) is one of non-invasive techniques to image temporal dynamics of blood oxygen level dependency (BOLD) response Glover (2011). It has been widely exploited to understand brain functional activities that is characterized by the fluctuation of fMRI time series Huettel et al. (2004) and functional organization that is often characterized by the synchronization of fMRI time series among brain regions Fornito et al. (2015). With machine learning technique, fMRI has shown to well predict disease diagnosis, individual demographic information, (i.e., age and gender), and cognitive ability Khosla et al. (2019); Pervaiz et al. (2020); Sui et al. (2020).

With the growth of deep learning technique, substantial research applies recurrent neural network (RNN), such that long short-term memory (LSTM) and gated recurrent unit (GRU), to learn temporal dependency of fMRI time series Dvornek et al. (2017); Li and Fan (2018). The integration of convolutional neural network (CNN) and RNN is also explored to learn spatial and temporal patterns of fMRI time series for the diagnosis of attention-deficit/hyperactivity disorder (ADHD)

[*] Corresponding author at: National University of Singapore, 4 Engineering Drive 3, Block E4 04-08 117583, Singapore.

E-mail address: bieqa@nus.edu.sg (A. Qiu).

[1] Co-first authors.

https://doi.org/10.1016/j.media.2022.102370
1361-8415/

Mao et al. (2019), schizophrenia Yan et al. (2019), and the prediction of sex and age Gadgil et al. (2020). Likewise, CNN or CNN-based deep neural networks are applied to brain functional connectivity for age prediction Li et al. (2018) and schizophrenia diagnosis Qiao et al. (2020). BrainNetCNN Kawahara et al. (2017) is specifically designed to leverage the topological locality of functional connectivity via three types of convolutional layers, including edge-to-edge layers, edge-to-node layers, and node-to-graph layer. BrainNetCNN has been recognized as one of state-of-the-art deep learning techniques for brain functional networks since it is comparable or outperforms to many machine learning methods, such as random forest, multi-layer perceptron (MLP), CNN, and neural network with dictionary learning D'Souza et al. (2019); Vaswani et al. (2017); Li and Duncan (2020); Lee et al. (2021); He et al. (2020).

Both functional time series and connectivity have demonstrated their discriminative power for age and cognition prediction, as well as disease diagnosis. Nevertheless, there is a lack of deep learning methods that take the advantage of functional time series and connectivity and incorporate both of them. The functional connections among brain regions are often neglected when learning functional time series in existing methods Dvornek et al. (2017); Li and Fan (2018). Similarly, brain regions communicate with each other at multi-temporal scales, which are not explicitly modeled in BrainNetCNN Kawahara et al. (2017) or other deep learning methods for functional connectivity data D'Souza et al. (2019); Vaswani et al. (2017); Li and Duncan (2020); Lee et al. (2021); He et al. (2020). Moreover, sparse studies provide interpretable deep learning methods on functional time series or connectivity data that learn the contribution of individual brain regions to prediction or disease classification Mahmood et al. (2021); Huang et al. (2020a).

In this study, we propose a new deep learning framework, spatio-temporal directed acyclic graph with attention mechanisms (ST-DAG-Att), (1) to incorporate multiple spatio-temporal scales of functional time series and connectivity data; (2) to conduct spatial convolution of functional time series based on the functional connections of brain regions characterized by the brain functional network; (3) to design functional connectivity based attention mechanism for understanding the discriminative power of each brain region. To achieve these, this ST-DAG-Att framework comprises of two neural networks, (1) spatio-temporal graph convolutional network (ST-graph-conv) to learn the spatial and temporal information of functional time series at multiple temporal and spatial graph scales, where the graph is represented by the brain functional network, the spatial convolution is over the space of this graph, and the temporal convolution is over the time dimension. (2) functional connectivity convolutional network (FC-conv) to learn high-level relevant functional connectivity features, where the functional connectivity is derived from embedded multi-scale functional time series and the convolutional operation is applied along both edge and node dimensions of the brain functional network. We design the overall architecture of the ST-DAG-Att based on the concept of directed acyclic graph (DAG; Yang and Ramanan, 2015) to integrate embedded functional time series and connectivity at multiple scales via the skip connections among the layers of the ST-graph-conv and FC-conv networks. Both the ST-graph-conv and FC-conv networks are designed as feed-forward models structured as DAGs. Moreover, this framework consists of a key attention component, i.e., functional connectivity-based spatial attention (FC-SAtt), that generates a spatial attention map through learning the local dependency among high-level features of functional connectivity and emphasizing brain regions whose functional organization is interpretable in relation to the predictive outcome. This FC-SAtt additionally plays a role in spatial pooling of the brain functional network to reduce the dimensionality of the spatial domain and to facilitate the multi-scale analysis of the brain functional network and functional time signals.

We perform experiments for understanding the importance of the two networks in the proposed ST-DAG-Att framework based on the Adolescent Brain Cognitive Development (ABCD) dataset ($n = 7693$) for fluid intelligence prediction. Moreover, we examine the generalizability of the ST-DAG-Att framework based on both the ABCD dataset for fluid intelligence prediction and the Open Access Series of Imaging Study-3 (OASIS-3) dataset ($n = 1786$) for age prediction. Moreover, we compare the proposed ST-DAG-Att framework with a representative machine learning model, support vector regression (SVR), elastic net's mixture with random forest Pornpattananangkul et al. (2021b), and the state-of-the-art deep learning models, such as spatio-temporal graph convolution Gadgil et al. (2020) and BrainNetCNN Kawahara et al. (2017). Our results show that the ST-DAG-Att framework outperforms all for the prediction of fluid intelligence in the ABCD dataset.

This study contributes to

- A novel graph convolutional neural network combining signal processing and network processing in the space of the brain functional network.
- Multi-scale integration of spatial, temporal and functional connectivity information via the architecture of the direct acyclic graph.
- Spatial attention map based on high-level functional connectivity features.
- Spatial attention pooling of the brain functional network based on the FC-based spatial attention map.
- The multi-scale analysis of the brain functional signals and functional network via the spatial attention map and pooling.

2. Related work

2.1. Deep learning on functional time series

A recurrent neural network (RNN) is a class of neural networks that allow outputs from previous time points to be used as inputs. It is suitable to analyze time series data. RNN, such as long short-term memory (LSTM) and gated recurrent unit (GRU), is used to learn temporal dependency between time points without considering the spatial dependency between brain regions to predict Autism Dvornek et al. (2017) and to decode brain states Li and Fan (2018). To consider fMRI spatial information, 3D CNN was applied to fMRI volumes to diagnose Alzheimer's disease (AD), where fMRI volumes are treated as inputs of multiple 3D CNNs Parmar et al. (2020). Compared to this 3D CNN model, Wang et al. (2019) take the fMRI time series of one brain region into a convolutional RNN, where a traditional operator in RNN, Hadamard product, is replaced by 1D convolution, for identifying individuals. Yan et al. (2019) employ multi-scale temporal convolutions and a weighted sum across all brain regions followed by a stacked GRU module to discriminate schizophrenia. Mao et al. (2019) propose two deep learning approaches, a 3D residual neural network (ResNet) with LSTM and a 4D ResNet for the diagnosis of attention-deficit/hyperactivity disorder (ADHD). Even though RNN is a popular neural network for time series data, recent research on fMRI suggests that temporal convolution can achieve better classification or prediction than RNN Mao et al. (2019); Gadgil et al. (2020). Moreover, RNN with CNN is computationally intensive and may require large memory for fMRI.

Beyond RNN, graph convolutional networks (GCNs) also apply to fMRI, where the fMRI time series are considered as graph-structured data and the graph is constructed via the brain functional network Gadgil et al. (2020); Azevedo et al. (2020). The spatial convolution in the GCN is across brain regions that are functionally connected, which is superior to traditional 3D volume convolution. The GCN is typically designed as a combination of 1D temporal convolutions, spatial graph convolutions Kipf and Welling (2016), and pooling. It has been shown that the GCN can well predict age and gender Gadgil et al. (2020); Azevedo et al. (2020). This study takes the advantage of the spatial convolution in the GCN but designs it at multiple temporal and spatial scales. Moreover, this study integrates the FC-based attention map with the GCN for the spatial pooling of the brain functional network to achieve multiple spatial scales.

2.2. Deep learning on functional connectivity

As for the brain functional network, several studies employ CNN and deep neural networks (DNNs), where brain functional connectivity maps are treated as images, to predict age Li et al. (2018) and schizophrenia diagnosis Qiao et al. (2020). In contrast to CNN and DNN, Kawahara et al. (2017) designs BrainNetCNN for the brain functional network that learns local relationship among edges and nodes through three kinds of convolutional layers, including edge-to-edge layers,edge-to-node layers, and node-to-graph layers. Therefore, BrainNetCNN emphasizes the topological locality of the brain functional network. Compared to the above mentioned methods, such as CNN with self-attention mechanism Vaswani et al. (2017), graph convolutional network (GCN) Parisot et al. (2018), graph neural network (GNN) Li et al. (2019), and neural networks with dictionary learning D'Souza et al. (2019), BrainNetCNN performs equally or better in classification and prediction problems Kawahara et al. (2017).

The GCN has also been applied to brain functional networks Parisot et al. (2017, 2018); Qu et al. (2020). The graph in the GCN is constructed based on brain functional networks and each node is characterized by connectivity features to evaluate the similarity among individuals. Each node in the graph represents one subject and an edge weight represents the similarity between two subjects based on their functional networks. Then, embedding approaches are developed to map individuals to a low dimensional space for disease classification Parisot et al. (2017, 2018), cognitive prediction Qu et al. (2020), or individual recognition Jiang et al. (2020).

This study takes the advantage of GCN and BrainNetCNN. However, this study designs the graph in the GCN based on the functional connections among brain regions rather than the similarity among individuals. We leverage the local and global dependency among edges and nodes via edge convolution and node convolution and create a spatial attention map to focus brain regions that most contribute to outputs.

3. Methods

The aim of this study is to design the architecture of the ST-DAG-Att framework so that it can extract multi-scale features from brain functional signals and functional networks and integrate them for classification or prediction. Since functional signals are communicated based on the underlying functional connections among brain regions, it is necessary to spatially process brain functional signals in the space of brain functional networks. Therefore, our ST-DAG-Att framework incorporates a ST-graph-conv network where the graph is represented by the brain functional network, each node represents one brain region and is associated with its functional signals, each edge represents the functional connection between two regions. Spatial signal processing of brain functional signals is performed in this brain functional network space, while temporal signal processing of brain functional signals is examined in the temporal domain. We obtain spatial multi-scale features of brain functional signals by introducing spatial attentional mechanisms (FC-SAtt module) and spatial graph pooling to remove brain regions and their connections that have less discriminative power for classification or prediction. Moreover, the ST-DAG-Att framework incorporates the FC-conv network that allows processing the functional connectivities among brain regions and the functional connectivity strength of each brain region. Again, the multi-scale features of the brain functional networks are achieved in the aid of the spatial attentional mechanisms and spatial graph pooling. We employ the idea of the directed acyclic graph architecture to integrate the multi-scale features of functional signals and functional network for classification or prediction.

In the following sections, we will first describe the ST-graph-conv (blue in Fig. 1) and FC-conv networks (green in Fig. 1), and then FC-SAtt as well as spatial pooling modules in the ST-DAG-Att framework.

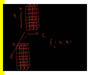
3.1. Spatio-temporal graph convolutional network (ST-graph-conv)

Denote a brain functional network (or graph) as $\mathcal{G} = \{V, E\}$ with a set of nodes, V, and a set of edges, E, where a node represents one brain region and an edge represents the functional connection between two brain regions. We represent the functional time series as $f(x, t)$, at node, x, and time, t, where $t = 1, \ldots, T$.

As illustrated in Fig. 2, we design the ST-graph-conv network to characterize brain functional signals and their communication at multiple spatial and temporal scales. We achieve this goal via (1) temporal convolution; (2) spatial graph convolution of functional signals across brain regions that are functionally connected; (3) spatial and temporal aggregation; (4) spatial graph pooling based on the functional connectivity to reduce the dimensionality of the brain network. We describe the details of these components and discuss how this ST-graph-conv network learns spatio-temporal features of functional time series and embeds them into a low-dimensional space.

3.1.1. Temporal convolution

The temporal convolution network is a computational unit that can map input functional signals, $\mathbf{f} = \{f_i(x,t)\}_{i=1,2,\ldots,C} \in \mathcal{R}^{n\times T\times C}$ to $\mathbf{f}' = \{f'_i(x,t)\}_{i=1,2,\ldots,C'} \in \mathcal{R}^{n\times \frac{T-w+1}{p_t}\times C'}$, where n is the number of brain regions, T is the number of functional time points, C, C' are the numbers of filter channels for the input and output signals, w is the temporal filter size, and p_t is a temporal pooling stride. This network involves sequential operators, including convolution, activation, and temporal average pooling ($Tpool$), that is,



$$f'_j\left(x, \frac{t-w+1}{p_t}\right) = Tpool\left(\sigma\left(\sum_{i=1}^{C} h_j(t,i) * f_i(x,t)\right)\right) \tag{1}$$

for $j = 1, 2, \ldots, C'$, where C' is the number of temporal filters. p_t is the pooling rate in the temporal domain. h_j is the jth filter with a size of $1 \times w$. The parameters for these filters are found via the back propagation optimization described below. We aggregate filtered signals through all C channels so that channel dependencies are implicitly embedded in functional signals, but are entangled with the local temporal correlation captured by filter, h_j. The channel relationships modelled by convolution are inherently *local*. We expect the learning of convolutional features to be enhanced by explicitly modelling channel interdependencies, so that the network is able to increase its sensitivity to informative features that can be exploited by subsequent transformations discussed below. We then employ leaky rectified linear unit (ReLU) as an activation function, σ, since negative time series in the ST-graph-conv network is considered to be meaningful in this study. The average pooling with a stride of p_t is used in the temporal domain.

3.1.2. Spatial graph convolution

In the ST-graph-conv network, we explore the interaction of functional signals across multiple brain regions via spatial con-

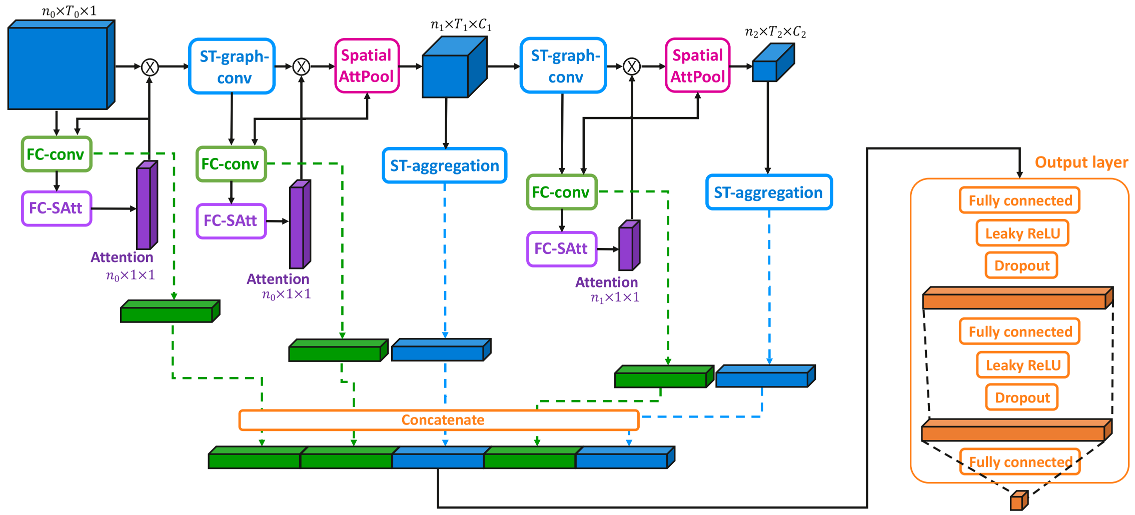


Fig. 1. Schematic architecture of spatio-temporal directed acyclic graph learning framework with attention mechanisms (ST-DAG-Att). Abbreviations: ST-graph-conv, spatio-temporal graph convolution; ST-aggregation, spatio-temporal aggregation; FC-conv, functional connectivity convolution; FC-SAtt, functional connectivity based spatial attention; AttPool, attention pooling

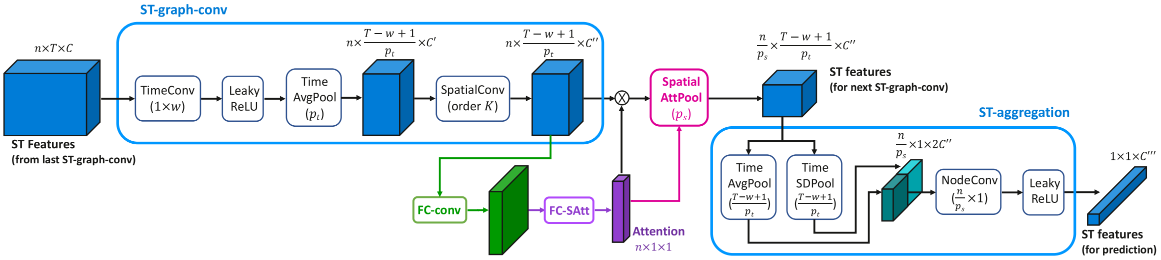


Fig. 2. Architecture of a spatio-temporal graph convolutional network (ST-graph-conv). n is the number of brain regions, T is the number of functional time points, C, C', C'', C''' represent the number of filter channels, w is the kernel size of filters, and p_t and p_s are the temporal and spatial pooling strides. Abbreviations: Conv, convolution; ReLU, rectified linear unit; Avg, average; SD, standard deviation; SAtt, spatial attention; AttPool, attention pooling.

volution on the brain functional network (or graph) in which brain functional connections are well characterized. Nevertheless, applying convolution on a graph is not straightforward. In this study, we realize the graph convolution through learning spectral filters in the graph Fourier domain, which has commonly used in spectral graph CNN and kernel smoothing on manifolds Bruna et al. (2013); Defferrard et al. (2016); Kipf and Welling (2016); Yi et al. (2017); Huang et al. (2020b, 2021).

We now define spectral filters. Let Δ denote the graph Laplacian on the brain functional network, $\mathcal{G}$. We take an advantage of the fast implementation of spectral filters based on the recursive relation of Chebyshev polynomials. We design a filter g in the spectral domain based on the Chebyshev polynomials of order K as

$$g(\lambda) = \sum_{k=0}^{K-1} \theta_k T_k(\lambda) \tag{2}$$

where λ is the eigenvalue of the graph Laplacian Δ, and θ_k are the parameters that determine the shape of g. T_k is the Chebyshev polynomial of the form $T_k(\lambda) = \cos(k\cos^{-1}\lambda)$. The order K of Chebyshev polynomials is related to the spatial localization property of g, which is the key parameter representing the kernel size of the spatial graph convolution Bruna et al. (2013); Defferrard et al. (2016). Moreover, this study employs the normalized form of the graph Laplacian, $\Delta = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$, where I is an identity matrix and D and A are the degree and adjacency matrices of $\mathcal{G}$. The normalized form of the graph Laplacian allows the generalization of g from one brain functional network to others since their eigenvalues are in the same range.

We apply g to functional signals at time t, $\mathbf{f}' = \{f_i'(x,t)\} \in \mathcal{R}^{n\times\frac{T-w+1}{p_t}\times C'}$ obtained from the above temporal convolution, that is,

$$f_j''(x,t) = \sum_{i=1}^{C'} \sum_{k=0}^{K-1} \theta_k^{ij} T_k(\Delta) f_i'(x,t), \tag{3}$$

for $j = 1, 2, \ldots, C''$, where C'' is the number of spectral filters. This spectral graph convolution is applied to each time point t, and all time points share the same filters. The filtering operation, $\sum_{k=0}^{K-1} \theta_k^{ij} T_k(\Delta)$, is implemented by first taking $f_i'(x,t)$ to the Fourier domain, filtering it based on the shape of Chebyshev polynomials in the Fourier domain, and transforms it back to the time domain. Finally, we aggregate the signals along the dimension related to the former temporal filters to explicitly model *local* spatial and temporal interdependencies. The obtained functional signals, $\mathbf{f}'' \in \mathcal{R}^{n\times\frac{T-w+1}{p_t}\times C''}$, from this spatial graph convolution will be used to learn functional connectivity features and spatial attention map that will be described later.

For now, assume $\mathbf{s} \in \mathcal{R}^{n\times1\times1}$ to be the spatial attention map. Spatial pooling is defined as

$$\mathbf{f}''' = Spool\left(\mathbf{f}'' \otimes \mathbf{s}\right), \tag{4}$$

where $\otimes$ denotes the element-wise multiplication over the spatial domain. *Spool* is a computational unit that selects brain regions based on spatial attention map, $\mathbf{s}$, and a spatial stride, p_s. Overall, the spatial graph network transforms the functional signal, $\mathbf{f}' = \{f_i'(x,t)\} \in \mathcal{R}^{n\times\frac{T-w+1}{p_t}\times C'}$ to $\mathbf{f}''' \in \mathcal{R}^{\frac{n}{p_s}\times\frac{T-w+1}{p_t}\times C''}$.

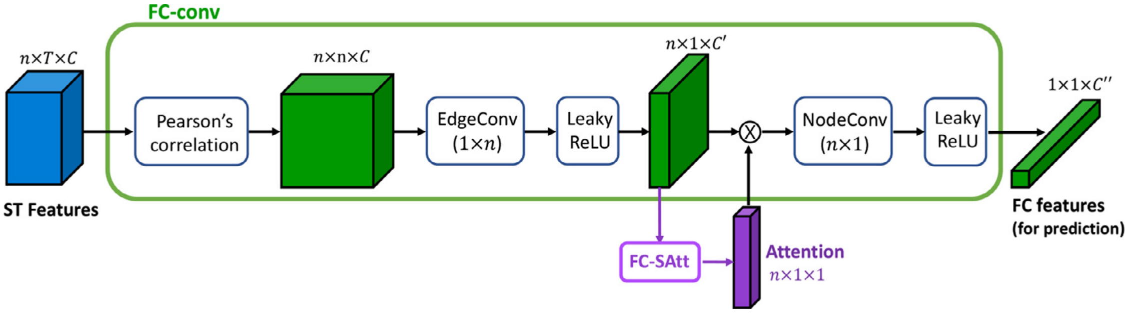


Fig. 3. Architecture of a functional connectivity convolutional (FC-conv) network. n is the number of brain regions, C, C', C'' represent the numbers of filter channels.

3.1.3. *Spatio-temporal aggregation*

In order to tackle the issue of exploiting spatial and temporal dependencies, we propose a spatio-temporal aggregation module. Each of the spatial and temporal learned filters operates with a local kernel and consequently each unit of the transformation output is unable to exploit contextual information outside of the kernel coverage. To mitigate this problem, we propose to squeeze *global* temporal and spatial information into a descriptor. As illustrated in Fig. 2, this is achieved by first using global average ($Tavg$) and standard deviation (Tsd) pooling to generate channel-wise statistics of $\mathbf{f}'''$ over the temporal dimension. We then employ spatial filters with kernel size of $\frac{n}{p_S}\times 1$ to aggregate the spatial dimension. Overall, the spatio-temporal aggregation unit maps $\mathbf{f}'''$ to $\mathbf{y}$, that is,

$$\mathbf{y}=\sigma\left(h_s*\begin{bmatrix}Tavg(\mathbf{f}''')\\ Tsd(\mathbf{f}''')\end{bmatrix}\right) \tag{5}$$

where σ is leaky ReLU and $\mathbf{y}\in\mathcal{R}^{1\times 1\times C'''}$. $\mathbf{y}$ can be treated as features that are related to specific spatial and temporal scales and fed into an output layer for classification or prediction (see Fig. 1).

3.2. *Functional connectivity convolutional (FC-conv) network*

As illustrated in Fig. 1, the second major network in the ST-DAG-Att framework is a functional connectivity convolutional (FC-conv) network to map functional connectivity into a low dimensional space given embedded functional time series at each spatial and temporal scale.

Fig. 3 shows the architecture of a FC-conv network. Denote an input of functional time series $\mathbf{f}\in\mathcal{R}^{n\times T\times C}$. We first construct the functional connectivity matrix via Pearson's correlation between any two brain regions over the time course for each channel. Let $\mathbf{F}\in\mathcal{R}^{n\times n\times C}$ represents the functional connectivity matrix. Subsequently, edge and node features are learned via edge and node convolution filters with a kernel size of n to aggregate *local and global* functional connectivity information. Hence, we respectively define the edge and node operations as follows:

$$\mathbf{Z}=\sigma(h_e*\mathbf{F}), \tag{6}$$

and

$$\mathbf{Z}'=\sigma(h_n*(\mathbf{Z}\otimes\mathbf{s})), \tag{7}$$

where h_e is the edge filter with a kernel size of $1\times n$ and h_n is the node filter with a kernel size of $n\times 1$. The edge filters apply the convolution over the edges of the functional networks, while the node filters apply the convolution over the nodes of the functional network. C', C'' are the numbers of the edge and node filters, respectively. Again, σ is leaky ReLU as an activation function and $\mathbf{s}$ is the spatial attention map that will be described below. The output of the functional connectivity convolution network, $\mathbf{Z}'\in\mathcal{R}^{1\times 1\times C''}$, encodes the local and global functional connectivity features that are explicitly learned through edge and node filters and their aggregation. The FC-conv network is applied to functional signals at all spatial and temporal scales.

3.3. *Functional connectivity based spatial attention (FC-SAtt)*

We design spatial attention mechanism based on the functional connectivity to identify brain regions that have the most discriminate power to classification or prediction. The functional connectivity, instead of functional time series, is used for attention mechanism mainly because the brain functional network has great statistical power to predict diagnosis and cognition, which can be shown in our results below. In this study, the functional connectivity based spatial attention (FC-SAtt) module not only facilitates the enhancement or suppression of brain regions' contribution to outcomes but also plays a role in spatial pooling of the brain functional network that will be discussed next.

The FC-SAtt mechanism is motivated by an existing squeeze-and-excitation network Hu et al. (2018) and convolutional block attention Woo et al. (2018). Fig. 4 shows the architecture of our proposed FC-SAtt module. Recall the functional connectivity features defined in Eq. (6), $\mathbf{Z}\in\mathcal{R}^{n\times 1\times C'}$. We first integrate the edge information over all edge filter channels through global average pooling

$$Cavg(\mathbf{Z})=\sum_{i=1}^{C'}\mathbf{Z}_i\in\mathcal{R}^{n\times 1\times 1}.$$

The global pooling operation is an efficient method to generate channel-wise statistics to highlight the informative nodes that are related to these edges Hu et al. (2018); Woo et al. (2018). Next, we employ a bottleneck-like multi-layer perceptron (MLP) with two fully connected layers, one ReLU and one sigmoid activation to identify brain regions that most contribute to the prediction of the outcome, that is,

$$\mathbf{s}=Sigmoid\Big[Fully_2\Big(ReLU\big(Fully_1(Cavg(\mathbf{Z}))\big)\Big)\Big].$$

The first fully connected layer reduces the spatial dimension from n nodes to $\frac{n}{r}$ hidden nodes, and the second fully connected layer recalibrates n nodes from $\frac{n}{r}$ hidden nodes, where r is the drop rate of the fully connected layers. The MLP with fully connected layers connecting all nodes can learn node dependencies.

3.4. *Spatial attention graph pooling*

The FC-SAtt module may generate the spatial attention map different from one sample to the other. This may increase the training difficulty in the following layers due to distinct graph configurations across training samples. Hence, we design a spatial attention pooling operator to generate a common spatial mask applicable to all samples, as shown in Fig. 5.

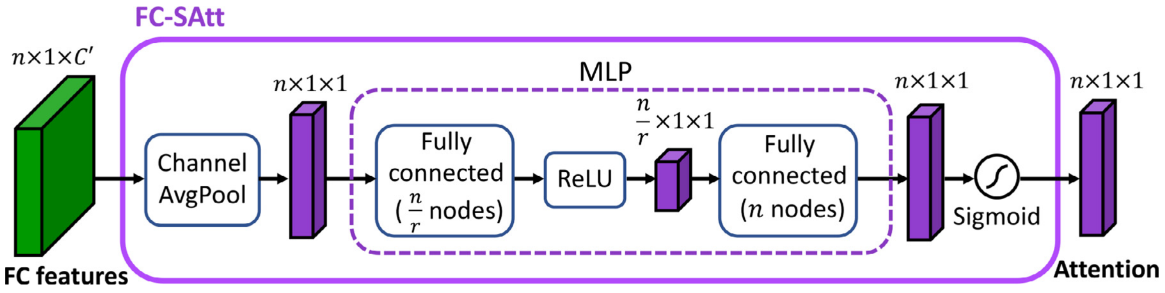


Fig. 4. Architecture of a functional connectivity based spatial attention (FC-SAtt) module.

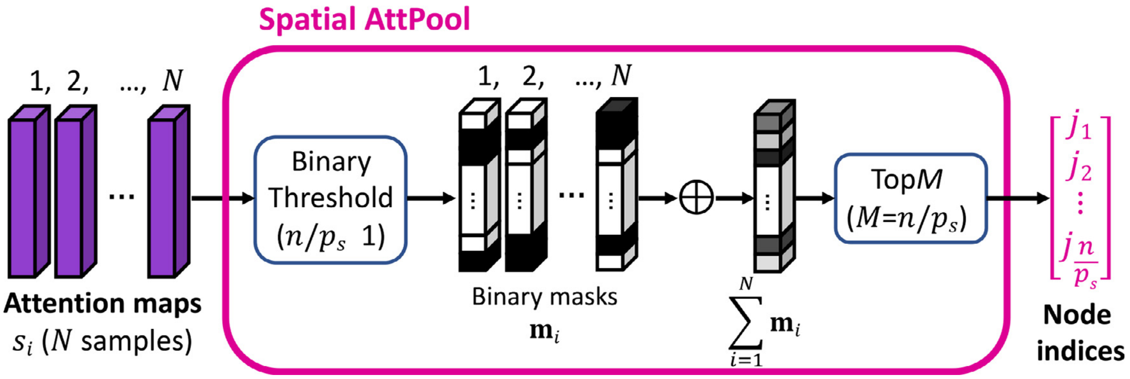


Fig. 5. The architecture of the spatial attention pooling operation to concatenate the attention maps of individual samples and to generate a common attentional map for all samples.

Let $\mathbf{s}_i \in \mathcal{R}^n$ be the spatial attention map of the ith sample. If the spatial pooling size is p_s, the attention map is first thresholded to generate a binary mask $\mathbf{m}_i \in \mathcal{R}^n$ where the top n/p_s nodes with the highest attention value are assigned 1 and the others are assigned 0. Given the set of nodes, V, we select top $M = n/p_s$ nodes from the summation of all the binary masks, and the new set of nodes is given by

$$V' = \left\{ x : TopM\left(\sum_{i=1}^{N} \mathbf{m}_i\right) \right\} \subset V,$$

where N is the number of samples in the training set. Therefore, the new set of nodes will be applied to all the samples.

This spatial attention pooling is integrated with the ST-graph-conv network to reduce the brain functional network graph. Only the functional signals on these selected nodes are retained in the subsequent analysis. The new brain functional network (i.e., a subgraph of the original graph $\mathcal{G} = \{V, E\}$) is defined by removing the edges connecting to these removed nodes and keeping the same weights for the remaining edges. Hence, the new brain functional network is given as $\mathcal{G}' = \{V', E'\}$ with $|V'| = n/p_s$. The new graph Laplacian of size $\frac{n}{p_s} \times \frac{n}{p_s}$ is updated accordingly. Hence, the spatial attention pooling facilitates the spatial multi-scale analysis and reduces the spatial dimensionality of the functional signals and functional network.

3.5. Directed acyclic graph for multi-scale analysis on functional signals and connectivity

Most existing deep learning neural networks extract features from a single output layer. This study takes an advantage of a directed acyclic graph (DAG) neural network Yang and Ramanan (2015). Our proposed ST-DAG-Att framework extracts functional signal features and functional connectivity features from multiple layers and simultaneously takes these high, middle, and low-level features during learning and prediction while skipping connections among the layers. As illustrated in Fig. 1, both the ST-graph-conv (blue) and FC-conv networks (green) are feed-forward models structured as directed acyclic graphs (DAGs). Hence, the ST-DAG-Att framework can learn a set of multi-scale functional time series and connectivity features that can be effectively fed into the output layer (see Fig. 1).

3.6. Implementation

To implement the ST-DAG-Att framework, we first construct the brain functional network, $\mathcal{G}$, based on the training samples. For each sample, the functional connectivity matrix is constructed via Pearson's correlation analysis between the functional time series of any two brain regions. It is then thresholded with a sparsity level of 20% such that the corresponding binary graph only retains edges with top 20% absolute correlation Achard and Bullmore (2007). The brain functional network (or binary graph) is built based on the connections that are common among all the training samples.

We implement the framework in Python 3.7 and TensorFlow 1.13.1 library. The architecture of the ST-graph-conv network is determined by the number of layers in the ST-graph-conv, FC-conv, and output networks, which is optimized via exhausted search. The number of filters in the convolutional layers is searched in the set of $\{8, 16, 32, 64, 128, 256\}$. Hence, the ST-DAG-Att framework consists of the ST-graph-conv network with 2 convolutional layers, the FC-conv network with 3 convolutional layers and FC-SAtt modules, and the output layer with 3 fully connected layers, as shown in Fig. 1. Each ST-graph-conv layer has 8 filters in the temporal convolution and 8 spectral filters designed by the Chebyshev polynomials of order 4 in the spectral graph convolution. Both the temporal and spatial pooling is designed with a stride of 2. In the FC-conv network, each layer has 128 edge filters, 256 node filters, and a bottleneck ratio of 4 in the MLP. The output layer contains 3 fully connected layers with 256, 256 and 1 hidden node, respectively, and two dropout layers with a dropout rate of 0.2. Leaky ReLU with a leak rate of 0.33 is used in all layers since negative time series in the ST-graph-conv network and negative functional connectivity in the FC-conv network are considered to be meaningful in this study.

The ST-DAG-Att model is trained using NIVIDIA Tesla V100-SXM2 GPU with 32GB RAM and Intel Xeon Gold 5118 CPU with 2.30GHz and by the stochastic gradient descent algorithm with mini-batch size of 32. Each epoch takes 543.58 sec when the training sample size is 4615. Training is in general convergent after 10 epochs.

3.7. Evaluation metrics an~~d~~ cross-validation

This study employs root mean square error (RMSE) to quantify the difference between actual and predicted values of testing samples. In addition, we also employ mean absolute error (MAE) and Pearson's correlation between actual and predicted values of testing samples.

4. Datasets and MRI preprocessing

This study includes two datasets, Adolescent Brain Cognitive Development (ABCD) and Open Access Series of Imaging Study-3 (OASIS-3), to predict fluid intelligence and age, respectively. We briefly describe the datasets and MRI preprocessing.

4.1. Adolescent brain cognitive development (ABCD)

The Adolescent Brain Cognitive Development (ABCD) study is an ongoing study and its data are publically available (release 2.0.1; https://abcdstudy.org/). In the first visit, the study includes youth 9–11 years of age with a similar proportion of males and females living in the United States ($n = 11875$). The sample selection criteria are targeted to reflect the sociodemographic proportion of the U.S. population as described in the ABCD study design Barch et al. (2018). All participants are administered assessments to obtain data on the respective youth's brain imaging, demographics, environment factors, and behavioral assessment. Written informed consent is obtained from all parents, and all children provided assent to a research protocol approved by the institutional review board at each data collection site (https://abcdstudy.org/study-sites/).

Images are acquired across 21 sites in the United States with harmonized imaging protocols for Philips, GE, and Siemens scanners. The detail of image acquisition can be found in Casey et al. (2018). This study only uses structural T1-weighted and resting-state fMRI (rs-fMRI) images. Rs-fMRI is acquired for multiple runs with 2.4 mm^3 isotropic voxels and temporal resolution (TR) of 800 ms. The structural T1-weighted image is with 1 mm^3 spatial resolution. This study excludes the Philips scans due to released issues (see detailed errors in the ABCD release notes) and also excludes one site with only 24 subjects, resulting in 18 sites and 7693 subjects. We only use one run per subject and hence include 7693 rs-fMRI data.

This study predicts fluid intelligence that is defined as the average of 5 NIH Toolbox cognition scores, including Dimensional Change Card Sort, Flanker, Picture Sequence Memory, List Sorting Working Memory, and Pattern Comparison Processing Speed Akshoomoff et al. (2013). The fluid intelligence score of the 7693 subjects is from 64 to 123 with mean and standard deviation of 95.3 ± 7.3.

4.2. Open access series of imaging study-3 (OASIS-3)

The Open Access Series of Imaging Study-3 (OASIS-3) study is an ongoing longitudinal study on Alzheimer's disease (AD) and its data are publically available (https://www.oasis-brains.org). Due to the limited sample size of rs-fMRI scans in AD and mild cognitive impairment (MCI) patients, this study only includes the data of 468 normal controls. The number of visits per subject varies from 1 to 7, forming 1786 total rs-fMRI scans included in this study.

4.3. MRI preprocessing

This study employs the same analysis pipelines to analyze T1-weighted images and rs-fMRI. The structural image is further processed using FreeSurfer 5.3.0 to segment the brain image into three tissue types, gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Post-processing quality check is conducted according to the instruction on https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/TroubleshootingData. Non-linear image normalization is achieved by aligning individual T1-weighted MRI images to an atlas space via large deformation diffeomorphic metric mapping (LDDMM) Du et al. (2011); Tan and Qiu (2016).

The rs-fMRI is further processed with intensity normalization. The rs-fMRI scans with mean framewise displacement (FD $>$ 0.5 mm) of head motion are excluded from the study Power et al. (2012). Notably, no frame censoring or scrubbing is applied in this study. Six motion parameters, whole brain, WM, and CSF signals are further partialled out from rs-fMRI signals. Global signal regression is an appropriate preprocessing step particularly for studying pediatric to eliminate artefactual variance due to head motion (0.009–0.08 Hz) is then applied. Within each run, the mean functional volume is aligned to the corresponding structural T1-weighted image via rigid-body alignment. The rs-fMRI is finally transformed to the atlas space via LDDMM obtained based on the T1-weighted image. This study uses a whole-brain parcellation with 268 brain regions of interest (ROIs) Shen et al. (2017); Finn et al. (2015) and averages the time series over each ROI as the input of our deep learning framework. We compute the functional connectivity (FC) between any two ROIs via Pearson's correlation of their averaged time series.

5. Results

In this section, we first perform experiments to understand the ST-graph-conv network on functional time series and the FC-conv network on functional connectivity for the prediction of fluid intelligence using the ABCD dataset. We employ the ABCD dataset to predict fluid intelligence via 5-fold and leave-one-site-out cross-validation and then employ the OASIS-3 dataset to predict age via 5-fold cross-validation. Moreover, we compare our proposed ST-DAG-Att framework with existing BrainNetCNN Kawahara et al. (2017) and support vector regression (SVR) using both the ABCD and OASIS-3 datasets. Finally, we compare the intelligence prediction results obtained from SVR, elastic net's mixture with random forest Pornpattananangkul et al. (2021b), spatio-temporal graph CNN Gadgil et al. (2020), BrainNetCNN Kawahara et al. (2017), and our ST-DAG-Att method.

5.1. Spatio-temporal directed acyclic graph learning

This study uses the ABCD dataset for the prediction of fluid intelligence to understand the components of the spatio-temporal directed graph learning framework. We employ 5-fold cross-validation, in which one fold of the samples is left out for testing and four folds are used for training (75%) and validation (25%). Given the family information of the subjects, we keep the rs-fMRI scans from the same family in the training, validation, or in the testing set to avoid potential data leakage. We first compare the proposed ST-DAG-Att framework with the two major networks in the ST-DAG-Att, the ST-graph-conv network (Fig. 6A) and the FC-conv + FC-SAtt model (Fig. 6B). We train these networks with a

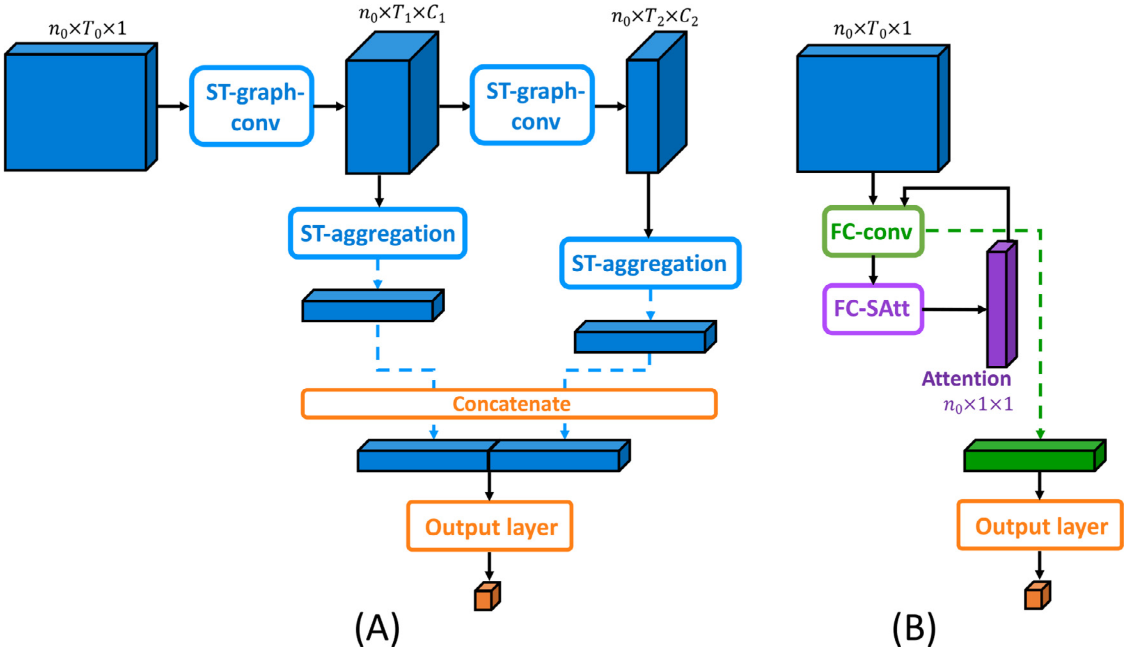


Fig. 6. Schematic architectures of (A) ST-graph-conv network and (B) FC-conv network. Abbreviations: ST-graph-conv, spatio-temporal graph convolution; ST-aggregation, spatio-temporal aggregation; FC-conv, functional connectivity convolution; FC-SAtt, functional connectivity based spatial attention; output layer, output 3-layer MLP.

Table 1
Prediction accuracy of fluid intelligence obtained from 5-fold cross-validation on the ABCD dataset.

Model	Correlation	MAE	RMSE
ST-graph-conv	0.223 ± 0.005	5.657 ± 0.008	7.165 ± 0.008
FC-conv	0.277 ± 0.003	5.595 ± 0.010	7.065 ± 0.007
ST-DAG-Att	0.288 ± 0.003	5.582 ± 0.012	7.046 ± 0.010

learning rate of 10^{-3}, 10 epochs, and l_2-norm regularization rate of 1×10^{-3}. We perform 5-fold cross-validation 10 times.

Fig. 8 shows the scatter plots of the actual and predicted fluid intelligence obtained from the ST-graph-conv, FC-conv, and ST-DAG-Att networks. Table. 1 lists the Pearson's correlation, MAE, and RMSE for these experiments. The ST-DAG-Att framework significantly outperforms the ST-graph-conv network (correlation: $p = 3.4 \times 10^{-18}$; MAE: $p = 3.4 \times 10^{-12}$; RMSE: $p = 8.6 \times 10^{-17}$) and the FC-conv network (correlation: $p = 8.8 \times 10^{-7}$; MAE: $p = 0.02$, RMSE: $p = 8.1 \times 10^{-5}$).

5.2. Fluid intelligence prediction via leave-one-site-out cross-validation

This experiment employs the ABCD dataset to predict fluid intelligence via leave-one-site-out cross-validation to demonstrate the generalizability against detecting spurious behavior relationships with the functional brain Sripada et al. (2020). The ABCD dataset employed in this study comprises of 18 sites with 480, 520, 515, 293, 465, 259, 205, 342, 452, 357, 446, 453, 496, 282, 915, 229, 571, and 413 rs-fMRI scans, respectively. We train the proposed ST-DAG-Att framework with a learning rate of 10^{-3}, 10 epochs, and l_2-norm regularization rate of 1×10^{-3}. We perform leave-one-site-out cross-validation in which the training and validation data are randomly divided into 75% and 25% of the samples of 17 sites and leave one site as the testing set. We repeat the experiment 10 times.

Fig. 7 A (orange bars) shows the correlation coefficient between the actual and predicted fluid intelligence for each study site. The correlation coefficient ranges from 0.188 to 0.383 and with mean of 0.281 and standard deviation of 0.004. Moreover, Fig. 7B, C (orange bars) show MAE and RMSE between the actual and predicted fluid intelligence for each site. MAE ranges from 5.076 to 6.156 and with mean of 5.597 and standard deviation of 0.012, while RMSE ranges from 6.447 to 7.679 and with mean of 7.063 and standard deviation of 0.011.

We employ the spatial attention and pooling modules to create the attention map in Fig. 9A. The attention map is created from the first block after the computation of the spatial attention graph pooling module where all the samples are averaged over the 18 sites and 10 repetitions. It shows the brain regions whose time series and functional connectivity most contribute to the prediction of fluid intelligence. These brain regions are mainly located in the frontal and parietal regions as well as the cerebellar and visual regions. In this study, fluid intelligence is computed based on the five NIH Toolbox cognitive tests, including Dimensional Change Card Sort, Flanker, Picture Sequence Memory, List Sorting Working Memory, and Pattern Comparison Processing Speed Akshoomoff et al. (2013). Fluid intelligence measures individual capacity for new learning and information processing in novel situations. Existing functional MRI studies suggest that neural activity in the frontal and parietal regions during the aforementioned cognitive tasks is related to individual differences in intelligence Duncan (2010); Jung and Haier (2007). Moreover, rs-fMRI studies show that intelligence is underpinned by communication between frontal and parietal regions Song et al. (2008). Our attention map is highly consistent with these findings and provides evidence on the fronto-parietal integration theory of intelligence Yoon et al. (2017).

5.3. Age prediction

This study also applies the proposed ST-DAG-Att framework to predict age using the OASIS-3 dataset. The age range of the OASIS-3 dataset is from 42 to 96 years (mean ± standard deviation: 69.6 ± 8.5). We employ 5-fold cross-validation, in which one fold of the samples is left out for testing and four folds are used for training (75%) and validation (25%). We train the ST-DAG-Att with

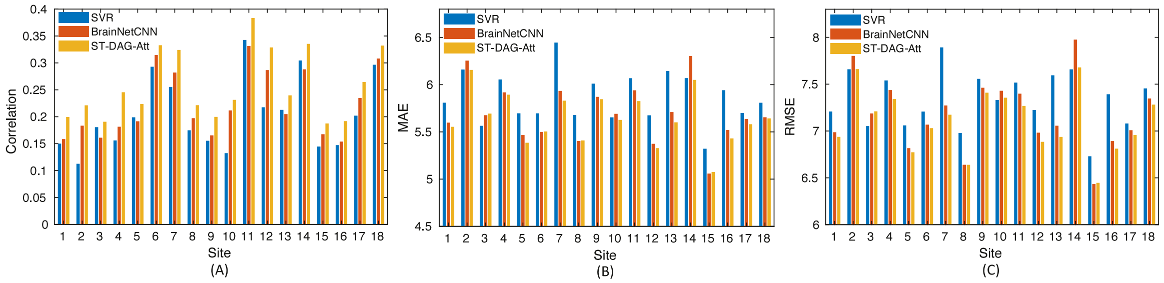


Fig. 7. Prediction accuracy of fluid intelligence via leave-one-site-out cross-validation using the ST-DAG-Att (orange), BrainNetCNN (red), and SVR methods (blue). Panels (A-C) illustrate the correlation, MAE, and RMSE between the actual and predicted fluid intelligence of the ABCD dataset.

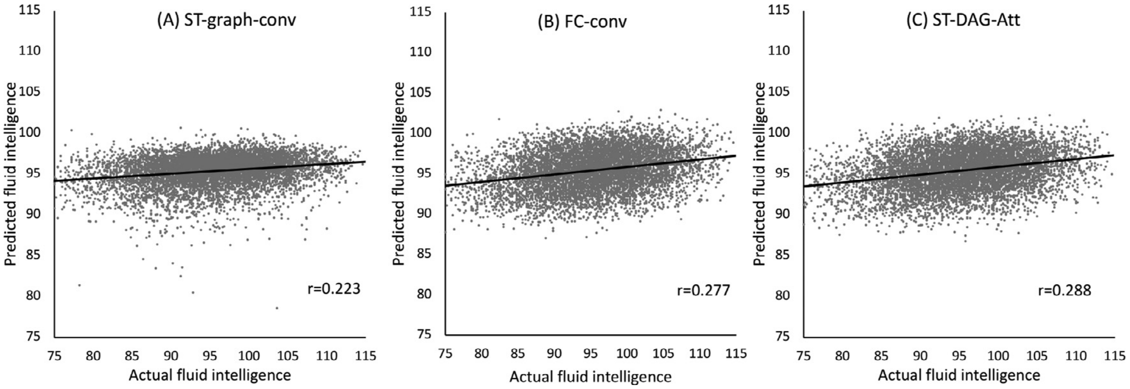


Fig. 8. Scatter plots of the actual and predicted fluid intelligence obtained from the ST-graph-conv (A), FC-conv (B), and ST-DAG-Att (C).

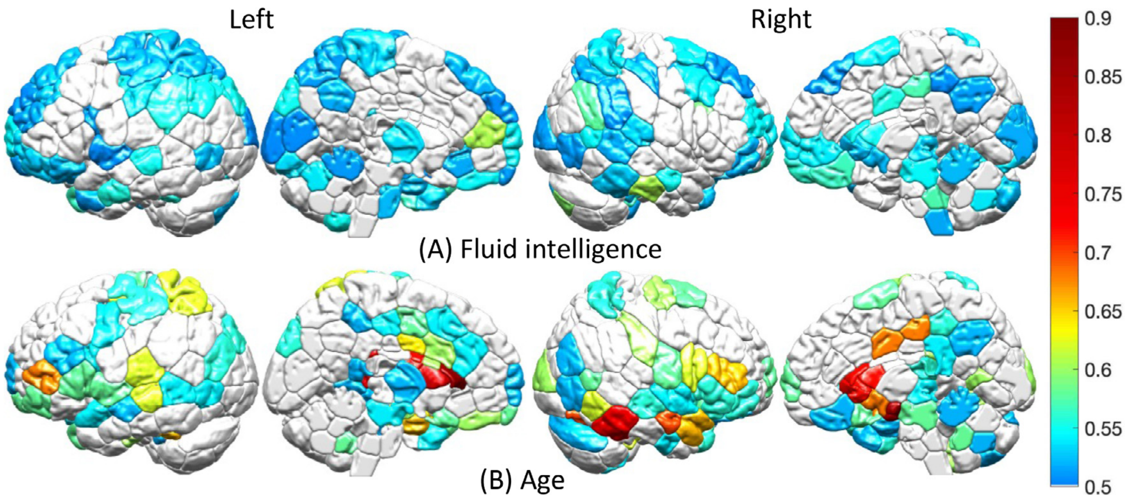


Fig. 9. Attention maps for fluid intelligence (A) and age prediction (B).

a learning rate of 10^{-2} and l_2-norm regularization rate of 1×10^{-4}. We perform 5-fold cross-validation 10 times.

Fig. 10 (orange bars) shows the Pearson's correlation, MAE and RMSE between the actual and predicted age. The predicted age is highly correlated with the actual age ($r = 0.597 \pm 0.011$). The MAE between the actual and predicted age is 5.377 ± 0.049 years and the RMSE is 6.820 ± 0.064.

Fig. 9 B shows the attention map for age prediction. The attention map is created as averaged all the samples over the 5 folds and 10 repetitions. The brain regions whose time series and functional connectivity most contribute to age include the medial temporal lobe and subcortical regions as well as the regions of default mode network (DMN). The brain regions emphasized in this study are highly consistent to those related to aging Wen et al. (2020b).

5.4. Comparisons with BrainNetCNN and SVR in the prediction of fluid intelligence and age

We compare our proposed ST-DAG-Att framework with BrainNetCNN Kawahara et al. (2017) and SVR for the prediction of fluid intelligence of the ABCD dataset and age prediction of the OASIS-3 dataset. We choose BrainNetCNN Kawahara et al. (2017) to compare with our ST-DAG-Att framework because several studies show that BrainNetCNN is one of the state-of-the-art deep learning approach for the brain functional network D'Souza et al. (2019); Li and Duncan (2020); Lee et al. (2021). For BrainNetCNN, we utilize the same architecture as that used in Kawahara et al. (2017), that is, 32, 64, and 256 output channels in the edge-to-edge, edge-to-node, and node-to-graph layers and 128 and 30 hidden nodes in the 1st and 2nd fully connected layers with dropout rate 0.5.

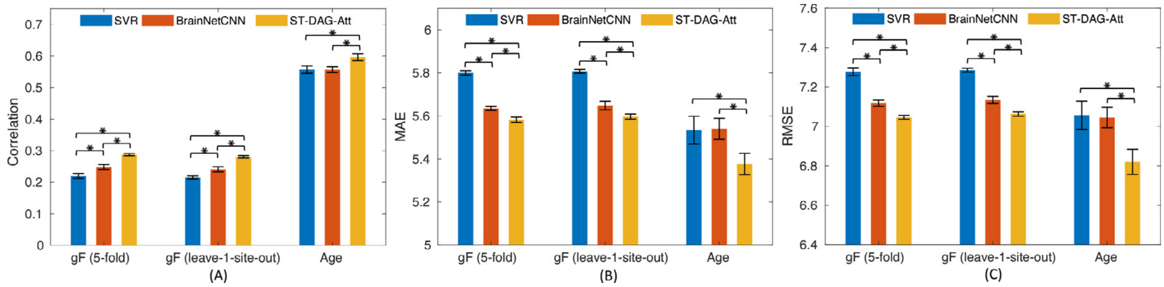


Fig. 10. Prediction accuracy of fluid intelligence (gF) prediction with 5-fold and leave-one-site-out cross-validations using the ABCD dataset and age prediction using the OASIS-3 dataset. $*$ denotes $p < 0.001$ in two-sample t-test. Blue, red, and orange bars show the results obtained from SVR, BrainNetCNN, and ST-DAG-Att, respectively.

We directly apply it to resting-state functional connectivity matrix. BrainNetCNN is trained via stochastic gradient descent algorithm with mini-batch size of 32, an initial learning rate of 10^{-3}, learning rate decay of 0.05 for every 20 epoches and momentum of 0.9. The number of epochs is 20 for fluid intelligence prediction and 150 for age prediction because the sample size of the OASIS-3 dataset is much smaller than that of the ABCD dataset. The l_2-norm regularization rate is 5×10^{-4} and 1×10^{-4} for the fluid intelligence and age predictions, respectively. Similarly, we apply SVR to the functional connectivity matrix, where its upper-triangle elements are vectorized as the input of SVR. The SVR is trained with an epsilon-insensitive loss. The validation set is used to tune the regularization and epsilon hyperparameters through a grid parameter search. All the experiments use the same training, validation, and testing sets as those used in the ST-DAG-Att to train and evaluate the accuracy of BrainNetCNN and SVR.

Fluid Intelligence Fig. 10 shows the Pearson's correlation, MAE and RMSE for the fluid intelligence prediction via 5-fold cross-validation. The proposed ST-DAG-Att framework shows the highest correlation coefficient and the lowest MAE and RMSE values than SVR and BrainNetCNN Kawahara et al. (2017). Two-sample t-tests suggest that the proposed model performs significantly better than BrainNetCNN (correlation: $p = 6.6 \times 10^{-12}$; MAE: $p = 1.8 \times 10^{-9}$; RMSE: $p = 4.0 \times 10^{-4}$) and SVR (correlation: $p = 1.0 \times 10^{-15}$; MAE: $p = 1.0 \times 10^{-19}$; RMSE: $p = 1.1 \times 10^{-17}$), while the BrainNetCNN is significantly better than SVR (correlation: $p = 8.7 \times 10^{-8}$; MAE: $p = 8.7 \times 10^{-19}$; RMSE: $p = 8.7 \times 10^{-14}$).

Fig. 10 also shows the three evaluation metrics of the proposed ST-DAG-Att, SVR, and BrainNetCNN for fluid intelligence prediction via leave-one-site-out cross validation. Overall, two-sample $t-$tests show that the proposed model performs significantly better than BrainNetCNN (correlation: $p = 3.8 \times 10^{-11}$; MAE: $p = 1.2 \times 10^{-6}$; RMSE: $p = 3.8 \times 10^{-9}$) and SVR (correlation: $p = 2.7 \times 10^{-17}$; MAE: $p = 3.6 \times 10^{-20}$; RMSE: $p = 3.3 \times 10^{-20}$) and BrainNetCNN performs significantly better than SVR (correlation: $p = 1.1 \times 10^{-7}$; MAE: $p = 6.8 \times 10^{-15}$; RMSE: $p = 5.1 \times 10^{-15}$). Moreover, Fig. 7A shows that the ST-DAG-Att consistently outperforms the SVR and BrainNetCNN over 18 sites in correlation. Nevertheless, the BrainNetCNN has higher correlation in 13 sites but lower correlation in the other 5 sites when compared to SVR. Fig. 7B and C show that the ST-DAG-Att has lower or comparable MAE and RMSE to the SVR and BrainNetCNN except for the 3rd site, while the BrainNetCNN has higher MAE and RMSE than the SVR in 4 of the 18 sites.

Age Fig. 10 shows that the proposed ST-DAG-Att framework has the highest correlation coefficient and the lowest MAE and RMSE in the age prediction using the OASIS-3 dataset. The ST-DAG-Att significantly outperforms the SVR (correlation: $p = 2.3 \times 10^{-8}$; MAE: $p = 6.3 \times 10^{-7}$; RMSE: $p = 8.1 \times 10^{-8}$) and BrainNetCNN (correlation: $p = 1.5 \times 10^{-7}$; MAE: $p = 8.7 \times 10^{-6}$; RMSE: $p = 3.5 \times 10^{-7}$) but BrainNetCNN and SVR have the same performance (correlation: $p = 0.52$; MAE: $p = 0.79$; RMSE: $p = 0.68$).

5.5. Comparisons with elastic net's mixture with random forest, spatio-temporal graph convolution, and BrainNetCNN

In this study, we further employ the ABCD dataset and compare our ST-DAG-Att with the most promising machine learning methods, including SVR and elastic net's mixture with random forest Pornpattananangkul et al. (2021b), as well as deep learning methods, including spatio-temporal graph convolution Gadgil et al. (2020) and BrainNetCNN Kawahara et al. (2017).

We choose to compare our ST-DAG-Att method with elastic net's mixture with random forest Pornpattananangkul et al. (2021b) that was recently published and predicted fluid intelligence using the same ABCD dataset. Table 2 lists the prediction result of elastic net's mixture with random forest obtained from the study in Pornpattananangkul et al. (2021b) based on 1998 subjects from the ABCD dataset. The correlation coefficient between the actual and predicted fluid intelligence is 0.230, which is slighly higher than that obtained using SVR and the sample size of 7693.

In this study, we further compare our ST-DAG-Att method with two deep learning methods (spatio-temporal graph convolution Gadgil et al., 2020 and BrainNetCNN Kawahara et al., 2017). As discussed in Section of "Related Work", the spatio-temporal graph convolution Gadgil et al. (2020) is the most superior deep learning method on brain functional time signals because it incorporates the dynamics of brain functional networks as well as functional time signals. In contrast, BrainNetCNN Kawahara et al. (2017) is designed for brain functional networks and its performance is equal or superior to a number of deep learning models, such as multi-layer perceptron (MLP), CNN with self-attention mechanism Vaswani et al. (2017), graph convolutional network (GCN) Parisot et al. (2018), graph neural network (GNN) Li et al. (2019), and neural networks with dictionary learning D'Souza et al. (2019). Again, this study utilizes the same architecture as that used in Gadgil et al. (2020) and Kawahara et al. (2017) for spatio-temporal graph convolution and BrainNetCNN, respectively. For spatio-temporal graph convolution, the architecture is 64, 64 and 64 output channels in three spatio-temporal convolution layers and one fully connected layer with dropout rate of 0.5. For BrainNetCNN, the architecture is 32, 64, and 256 output channels in the edge-to-edge, edge-to-node, and node-to-graph layers and two fully connected layers with dropout rate 0.5. The spatio-temporal graph convolution is applied on functional time series matrix, and BrainNetCNN is directly applied on functional connectivity matrix. Both the spatio-temporal graph convolution and BrainNetCNN are trained via stochastic gradient descent algorithm with mini-batch size of 32. Specifically, BrainNetCNN is trained with an ini-

Table 2
Comparison among SVR, elastic net's mixture with random forest Pornpattananangkul et al. (2021b), spatio-temporal graph convolution Gadgil et al. (2020), BrainNetCNN Kawahara et al. (2017), and ST-DAG-Att based on the fluid intelligence prediction obtained from 5-fold cross-validation on the ABCD dataset.

Model	Year of publication	Sample size	Correlation
machine learning methods			
SVR	–	7693	0.220
elastic net's mixture with random forest	2021	1998	0.233
deep learning methods			
spatio-temporal graph convolution	2020	7693	0.220
BrainNetCNN	2017	7693	0.248
ST-DAG-Att	-	7693	0.288

tial learning rate of 10^{-3}, learning rate decay of 0.05 for every 20 epoches, momentum of 0.9, and l_2-norm regularization rate is 5×10^{-4}. Spatial-temporal graph convolutional network is trained with a learning rate of 10^{-3}, 10 epoches, and l_2-norm regularization rate is 1×10^{-3}. Table 2 lists the correlation coefficients between the actual and predicted fluid intelligence via 5-fold cross-validation, suggesting that our ST-DAG-Att ($r = 0.288$) provides the highest correlation in comparison with spatio-temporal graph convolution ($r = 0.220$) and BrainNetCNN ($r = 0.248$).

6. Discussion

This study develops the new interpretable deep learning framework, ST-DAG-Att, on functional time series and connectivity for the prediction of cognition and age. The architecture of the ST-DAG-Att framework is a directed acyclic graph that incorporates the features of functional time series and functional connectivity respectively derived from the ST-graph-conv and FC-conv networks at multiple spatial and temporal scales. The ST-DAG-Att framework incorporates the graph configuration of the brain functional network and FC-based spatial attention mechanism in spatio-temporal convolution network to embed functional time series and connectivity into low dimensional spaces. Moreover, this framework takes multi-scale embedded functional signals to construct the brain functional connectivity and transform it to FC features. Furthermore, our spatial attention module provides an interpretable brain map (e.g., Fig. 9) that shows brain regions whose functional connectivity strength has the most discriminative power to classification or prediction. Based on large-scale datasets, our experiments demonstrate that the integration of functional time series and connectivity is superior to using them alone for cognition and age prediction. Compared to the functional time series, the functional connectivity in our framework may have a greater discriminative power to cognition prediction. Nevertheless, the ST-graph-conv network provides functional time series at multiple temporal and spatial scales that allow the construction of functional connectivities at multiple scales as well. Our results further show that the proposed ST-DAG-Att framework performs similarly for cognition prediction on the independent samples from one site to the other.

This study conducted the comparison of the ST-DAG-Att with two deep learning methods, spatio-temporal graph convolution Gadgil et al. (2020) and BrainNetCNN Kawahara et al. (2017) that are respectively based on brain functional time signals and functional networks. These two deep learning methods are relatively advanced deep learning approaches for functional MRI data as discussed in Section of "Related Work". Our experiments demonstrate that the ST-DAG-Att can achieve a better prediction of fluid intelligence when compared to spatio-temporal graph convolution and BrainNetCNN. Our results also implicate the importance of incorporating both functional networks and time signals at multiple scales for the cognitive prediction.

Moreover, the most recent study Pornpattananangkul et al. (2021a) employed machine learning approaches on the rs-fMRI of the same ABCD dataset as that in our study and predicted fluid intelligence at correlation of $r = 0.233$, which was lower than the ST-DAG-Att performance ($r = 0.288$). Furthermore, other studies employed rs-fRMI and predicted IQ in children at correlation of 0.20 Langeslag et al. (2013) and classified adults with low and high IQ at accuracy of 65.6% Xiao et al. (2019). Likewise, a study with the largest sample size from the UK Biobank cohort ($n = 14,701$) and the similar age group as that in our study showed that rs-fMRI can predict age at correlation of $r = 0.444$, which was also lower than that obtained from our ST-DAG-Att framework.

This study provides several components, including the ST-graph-conv, FC-conv, FC-SAtt, and FC-based spatial attention graph pooling, which allows the construction of deep learning frameworks for graph-structured data and provides the attention map for understanding brain regions contributed to outcomes. Our study highlights frontal and parietal regions that most contribute to the prediction of fluid intelligence, which is consistent to existing findings in similar age groups Langeslag et al. (2013); Li and Tian (2014). Intelligence involves the ability to reason, plan, solve problems, think abstractly, learn quickly and learn from experience. The frontal and parietal regions form executive functional networks that are engaged in high-level cognitive processes. Therefore, the attention map from the ST-graph-conv identifies the meaningful regions related to intelligence. Moreover, the attention map identifies temporal regions (e.g., Wen et al., 2020a), sensorimotor cortex (e.g., He et al., 2017; Wen et al., 2020a), anterior cingulate cortex (e.g., Cao et al., 2014; Wen et al., 2020a), and regions involved in the default mode network (DMN), such as the posterior cingulate, precuneus, and medial frontal cortex (see review in Dennis and Thompson, 2014), that are related to aging. These regions have been well studied in normal aging using rs-fMRI (e.g., Wen et al., 2020a; He et al., 2017; Cao et al., 2014; Dennis and Thompson, 2014). Hence, our ST-DAG-Att can provide attention maps that can facilitate the intepretation of brain regions in relation to classification or prediction outcomes.

In conclusion, the ST-DAG-Att framework is an interpretable deep learning model that explain how it makes decisions and has reasonable performance. The FC-conv and FC-SAtt components can be further extended to incorporate dynamic functional connectivity, which needs further investigation.

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.

CRediT authorship contribution statement

Shih-Gu Huang: Conceptualization, Data curation, Methodology, Validation, Writing – original draft. **Jing Xia:** Data curation, Validation, Writing – review & editing. **Liyuan Xu:** Writing – review & editing. **Anqi Qiu:** Conceptualization, Data curation, Methodology, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Acknowledgments

This research/project is supported by the National Science Foundation MDS-2010778, the Singapore Ministry of Education Tier 1, and NUS Institute of Data Science. This research was also supported by the A*STAR Computational Resource Centre through the use of its high-performance computing facilities.

References

Achard, S., Bullmore, E., 2007. Efficiency and cost of economical brain functional networks. PLoS Comput. Biol. 3, e17.

Akshoomoff, N., Beaumont, J.L., Bauer, P.J., Dikmen, S.S., Gershon, R.C., Mungas, D., Slotkin, J., Tulsky, D., Weintraub, S., Zelazo, P.D., et al., 2013. VIII. NIH toolbox cognition battery (CB): composite scores of crystallized, fluid, and overall cognition. Monogr. Soc. Res. Child Dev. 78 (4), 119–132.

Azevedo, T., Passamonti, L., Lio, P., Toschi, N., 2020. A deep spatiotemporal graph learning architecture for brain connectivity analysis. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 1120–1123.

Barch, D.M., Albaugh, M.D., Avenevoli, S., Chang, L., Clark, D.B., Glantz, M.D., Hudziak, J.J., Jernigan, T.L., Tapert, S.F., Yurgelun-Todd, D., et al., 2018. Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: rationale and description. Dev. Cogn. Neurosci. 32, 55–66.

Bruna, J., Zaremba, W., Szlam, A., LeCun, Y., Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203

Cao, W., Luo, C., Zhu, B., Zhang, D., Dong, l., Gong, J., Gong, D., He, H., Tu, S., Yin, W., Li, J., Chen, H., Yao, D., 2014. Resting-state functional connectivity in anterior cingulate cortex in normal aging. Front. Aging Neurosci. 6, 280. doi:10.3389/fnagi.2014.00280.

Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., et al., 2018. The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. Dev. Cogn. Neurosci. 32, 43–54.

Defferrard, M., Bresson, X., Vandergheynst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering. Adv. Neural Inf. Process. Syst. 29, 3844–3852.

Dennis, E.L., Thompson, P.M., 2014. Functional brain connectivity using fMRI in aging and Alzheimer's disease. Neuropsychol. Rev. 24, 49–62.

D'Souza, N.S., Nebel, M.B., Wymbs, N., Mostofsky, S., Venkataraman, A., 2019. Integrating neural networks and dictionary learning for multidimensional clinical characterizations from functional connectomics data. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), pp. 709–717.

Du, J., Younes, L., Qiu, A., 2011. Whole brain diffeomorphic metric mapping via integration of sulcal and gyral curves, cortical surfaces, and images. NeuroImage 56 (1), 162–173.

Duncan, J., 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends Cogn. Sci. 14, 172–179.

Dvornek, N.C., Ventola, P., Pelphrey, K.A., Duncan, J.S., 2017. Identifying autism from resting-state fMRI using long short-term memory networks. In: International Workshop on Machine Learning in Medical Imaging, pp. 362–370.

Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat. Neurosci. 18 (11), 1664–1671.

Fornito, A., Zalesky, A., Breakspear, M., 2015. The connectomics of brain disorders. Nat. Rev. Neurosci. 16 (3), 159–172.

Gadgil, S., Zhao, Q., Pfefferbaum, A., Sullivan, E., Adeli, E., Pohl, K., 2020. Spatio-temporal graph convolution for resting-state fMRI analysis. In: Medical Image Computing and Computer-Assisted Intervention: MICCAI... International Conference on Medical Image Computing and Computer-Assisted Intervention, 12267, pp. 528–538.

Glover, G.H., 2011. Overview of functional magnetic resonance imaging. Neurosurg. Clin. 22 (2), 133–139.

He, H., Luo, C., Chang, X., Shan, Y., Cao, W., Gong, J., Klugah-Brown, B., Bobes Leon, M., Biswal, B., Yao, D., 2017. The functional integration in the sensory-motor system predicts aging in healthy older adults. Front. Aging Neurosci. 8. doi:10.3389/fnagi.2016.00306.

He, T., Kong, R., Holmes, A.J., Nguyen, M., Sabuncu, M.R., Eickhoff, S.B., Bzdok, D., Feng, J., Yeo, B.T.T., 2020. Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. NeuroImage 206, 116276.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141.

Huang, J., Zhou, L., Wang, L., Zhang, D., 2020. Attention-diffusion-bilinear neural network for brain network analysis. IEEE Trans. Med. Imaging 39 (7), 2541–2552.

Huang, S.-G., Chung, M.K., Qiu, A., 2021. Revisiting convolutional neural network on graphs with polynomial approximations of laplace-beltrami spectral filtering. Neural Comput. Appl. 33 (20), 13693–13704.

Huang, S.-G., Lyu, I., Qiu, A., Chung, M.K., 2020. Fast polynomial approximation of heat kernel convolution on manifolds and its application to brain sulcal and gyral graph pattern analysis. IEEE Trans. Med. Imaging 39 (6), 2201–2212.

Huettel, S.A., Song, A.W., McCarthy, G., et al., 2004. Functional Magnetic Resonance Imaging, vol. 1. Sinauer Associates Sunderland, MA.

Jiang, H., Cao, P., Xu, M., Yang, J., Zaiane, O., 2020. Hi-GCN: a hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction. Comput. Biol. Med. 127, 104096.

Jung, R.E., Haier, R.J., 2007. The parieto-frontal integration theory (p-FIT) of intelligence: converging neuroimaging evidence. Behav. Brain Sci. 30, 135–154.

Kawahara, J., Brown, C.J., Miller, S.P., Booth, B.G., Chau, V., Grunau, R.E., Zwicker, J.G., Hamarneh, G., 2017. BrainNetCNN: convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage 146, 1038–1049.

Khosla, M., Jamison, K., Ngo, G.H., Kuceyeski, A., Sabuncu, M.R., 2019. Machine learning in resting-state fMRI analysis. Magn. Reson. Imaging 64, 101–121.

Kipf, T. N., Welling, M., Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907

Langeslag, S., Schmidt, M., Ghassabian, A., Jaddoe, V., Hofman, A., Lugt, A., Verhulst, F., Tiemeier, H., White, T., 2013. Functional connectivity between parietal and frontal brain regions and intelligence in young children: the generation r study. Hum. Brain Mapp. 34. doi:10.1002/hbm.22143.

Lee, J., Ko, W., Kang, E., Suk, H.-I., 2021. A unified framework for personalized regions selection and functional relation modeling for early MCI identification. NeuroImage 236, 118048.

Li, C., Tian, L., 2014. Association between resting-state coactivation in the parieto-frontal network and intelligence during late childhood and adolescence. AJNR. Am. J. Neuroradiol. 35, 1150–1156.

Li, H., Fan, Y., 2018. Brain decoding from functional MRI using long short-term memory recurrent neural networks. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), pp. 320–328.

Li, H., Satterthwaite, T.D., Fan, Y., 2018. Brain age prediction based on resting-state functional connectivity patterns using convolutional neural networks. In: IEEE 15th International Symposium on Biomedical Imaging, pp. 101–104.

Li, X., Duncan, J., 2020. BrainGNN: interpretable brain graph neural network for fMRI analysis. bioRxiv.

Li, X., Dvornek, N.C., Zhou, Y., Zhuang, J., Ventola, P., Duncan, J.S., 2019. Graph neural network for interpreting task-fMRI biomarkers. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), pp. 485–493.

Mahmood, U., Fu, Z., Calhoun, V.D., Plis, S., 2021. A deep learning model for data–driven discovery of functional connectivity. Algorithms 14 (3), 75.

Mao, Z., Su, Y., Xu, G., Wang, X., Huang, Y., Yue, W., Sun, L., Xiong, N., 2019. Spatio-temporal deep learning method for ADHD fMRI classification. Inf. Sci. 499, 1–11.

Parisot, S., Ktena, S.I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., Rueckert, D., 2018. Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer's disease. Med. Image Anal. 48, 117–130.

Parisot, S., Ktena, S.I., Ferrante, E., Lee, M., Moreno, R.G., Glocker, B., Rueckert, D., 2017. Spectral graph convolutions for population-based disease prediction. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), pp. 177–185.

Parmar, H., Nutter, B., Long, R., Antani, S., Mitra, S., 2020. Spatiotemporal feature extraction and classification of Alzheimer's disease using deep learning 3D-CNN for fMRI data. J. Med. Imaging 7 (5), 056001.

Pervaiz, U., Vidaurre, D., Woolrich, M.W., Smith, S.M., 2020. Optimising network modelling methods for fMRI. NeuroImage 211, 116604.

Pornpattananangkul, N., Wang, Y., Stringaris, A., 2021a. Multimodal-neural predictive models of children's general intelligence that are stable across two years of development. bioRxiv. 10.1101/2021.02.21.432130

Pornpattananangkul, N., Wang, Y., Stringaris, A., 2021b. Multimodal-neural predictive models of children's general intelligence that are stable across two years of development. bioRxiv.

Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage 59 (3), 2142–2154.

Qiao, C., Shi, Y., Diao, Y.-X., Calhoun, V.D., Wang, Y.-P., 2020. Log-sum enhanced sparse deep neural network. Neurocomputing 407, 206–220.

Qu, G., Hu, W., Xiao, L., Wang, Y.-P., 2020. A graph deep learning model for the classification of groups with different IQ using resting state fMRI. In: Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 11317, p. 113170A.

Shen, X., Finn, E.S., Scheinost, D., Rosenberg, M.D., Chun, M.M., Papademetris, X., Constable, R.T., 2017. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. Nat. Protoc. 12 (3), 506–518.

Song, M., Zhou, Y., Li, J., Liu, Y., Tian, L., Yu, C., Jiang, T., 2008. Brain spontaneous functional connectivity and intelligence. NeuroImage 41, 1168–1176.

Sripada, C., Rutherford, S., Angstadt, M., Thompson, W.K., Luciana, M., Weigard, A., Hyde, L.H., Heitzeg, M., 2020. Prediction of neurocognition in youth from resting state fMRI. Mol. Psychiatry 25 (12), 3413–3421.

Sui, J., Jiang, R., Bustillo, J., Calhoun, V., 2020. Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: methods and promises. Biol. Psychiatry 88 (11), 818–828.

Tan, M., Qiu, A., 2016. Large deformation multiresolution diffeomorphic metric mapping for multiresolution cortical surfaces: a coarse-to-fine approach. IEEE Trans. Image Process. 25 (9), 4061–4074.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I., Attention is all you need. arXiv preprint:1706.03762

Wang, L., Li, K., Chen, X., Hu, X.P., 2019. Application of convolutional recurrent neural network for individual recognition based on resting state fMRI data. Front. Neurosci. 13, 434.

Wen, X., Dong, l., Chen, J., Xiang, J., Yang, J., Li, H., Liu, X., Luo, C., Yao, D., 2020. Detecting the information of functional connectivity networks in normal aging using deep learning from a big data perspective. Front. Neurosci. 13, 1435. doi:10.3389/fnins.2019.01435.

Wen, X., He, H., Dong, L., Chen, J., Yang, J., Guo, H., Luo, C., Yao, D., 2020. Alterations of local functional connectivity in lifespan: a resting-state fMRI study. Brain Behav. 10 (7), e01652.

Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. CBAM: convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19.

Xiao, L., Stephen, J.M., Wilson, T.W., Calhoun, V.D., Wang, Y.-P., 2019. Alternating diffusion map based fusion of multimodal brain connectivity networks for IQ prediction. IEEE Trans. Biomed. Eng. 66 (8), 2140–2151. doi:10.1109/TBME.2018.2884129.

Yan, W., Calhoun, V., Song, M., Cui, Y., Yan, H., Liu, S., Fan, L., Zuo, N., Yang, Z., Xu, K., et al., 2019. Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site FMRI data. EBioMedicine 47, 543–552.

Yang, S., Ramanan, D., 2015. Multi-scale recognition with DAG-CNNs. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1215–1223.

Yi, L., Su, H., Guo, X., Guibas, L.J., 2017. Syncspeccnn: synchronized spectral CNN for 3D shape segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2282–2290.

Yoon, Y.B., Shin, W., Lee, T.Y., Hur, J.W., Cho, K.I.K., Sohn, W.S., Kim, S.G., Lee, K.H., Kwon, J.S., 2017. Brain structural networks associated with intelligence and visuomotor ability. Sci. Rep. 7, 1–9.