HHS Public Access

Author manuscript

Med Image Anal. Author manuscript; available in PMC 2022 May 01.

Published in final edited form as:

Med Image Anal. 2022 May ; 78: 102413. doi:10.1016/j.media.2022.102413.

An attention-based hybrid deep learning framework integrating brain connectivity and activity of resting-state functional MRI data

Min Zhao[a,b], **Weizheng Yan**[c], **Na Luo**[a], **Dongmei Zhi**[d], **Zening Fu**[c], **Yuhui Du**[e], **Shan Yu**[a,b], **Tianzi Jiang**[a,b], **Vince D. Calhoun**[c], **Jing Sui**[c,d,*]

[a]Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China

[b]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China

[c]Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA, USA

[d]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

[e]School of Computer and Information Technology, Shanxi University, Taiyuan, China

Abstract

Functional magnetic resonance imaging (fMRI) as a promising tool to investigate psychotic disorders can be decomposed into useful imaging features such as time courses (TCs) of independent components (ICs) and functional network connectivity (FNC) calculated by TC cross-correlation. TCs reflect the temporal dynamics of brain activity and the FNC characterizes temporal coherence across intrinsic brain networks. Both features have been used as input to deep learning approaches with decent results. However, few studies have tried to leverage their complementary information to learn optimal representations at multiple facets. Motivated by this, we proposed a Hybrid Deep Learning Framework integrating brain Connectivity and Activity (HDLFCA) together by combining convolutional recurrent neural network (C-RNN) and deep neural network (DNN), aiming to improve classification accuracy and interpretability simultaneously. Specifically, C-RNN^{AM} was proposed to extract temporal dynamic dependencies with an attention module (AM) to automatically learn discriminative knowledge from TC nodes, while DNN was applied to identify the most group-discriminative FNC patterns with layer-wise

*Corresponding author at: State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, China., jsui@bnu.edu.cn (J. Sui).

CRediT authorship contribution statement

Min Zhao: Data curation, Investigation, Writing – original draft. **Weizheng Yan:** Writing – review & editing, Conceptualization. **Na Luo:** Writing – review & editing. **Dongmei Zhi:** Writing – review & editing. **Zening Fu:** Conceptualization. **Yuhui Du:** Writing – review & editing. **Shan Yu:** Writing – review & editing. **Tianzi Jiang:** Data curation. **Vince D. Calhoun:** Writing – original draft, Data curation. **Jing Sui:** Data curation, Writing – original draft.

Declaration of Competing Interest

The authors report no biomedical financial interests or potential conflicts of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.media.2022.102413.

relevance propagation (LRP). Then, both prediction outputs were concatenated to build a new feature matrix, generating the final decision by logistic regression. The effectiveness of HDLFCA was validated on both multi-site schizophrenia (SZ, n ~ 1100) and public autism datasets (ABIDE, n ~ 1522) by outperforming 12 alternative models at 2.8–8.9% accuracy, including 8 models using either static FNC or TCs and 4 models using dynamic FNC. Appreciable classification accuracy was achieved for HC vs. SZ (85.3%) and HC vs. Autism (72.4%) respectively. More importantly, the most group-discriminative brain regions can be easily attributed and visualized, providing meaningful biological interpretability and highlighting the great potential of the proposed HDLFCA model in the identification of valid neuroimaging biomarkers.

Keywords

Attention mechanism; Deep learning; Brain connectivity and activity; fMRI

1. Introduction

Functional magnetic resonance imaging (fMRI) has been a promising tool to provide novel insights into the brain function abnormalities of psychotic disorders (Andreou, 2020). Based on multivariate decomposition such as independent component analysis (ICA) (Du and Fan, 2013), useful imaging features such as independent components (ICs), their corresponding time courses (TCs) and functional network connectivity (FNC) (Calhoun and Adali, 2006; Jafri et al., 2008; Smith et al., 2009) can be easily extracted and widely used in studies of mental disorders (Fig. 1 A). Specifically, TCs reflect the temporal fluctuations of each IC, *i.e.,* the spatially distinct brain regions, while FNC characterizes the temporal coherence across the selected ICs by correlating their TCs, representing the intrinsic connectivity networks (Calhoun and Adali, 2012; Seeley et al., 2007; Supekar et al., 2009). Both features have been widely used in brain disorder comparison and classification.

On the other hand, with the ability to characterize discriminative patterns and learn optimal representations automatically from neuroimaging data, deep learning (DL) methods have received growing attention in fMRI-based diagnosis of mental disorders. One of the most commonly used DL input features is functional (network) connectivity calculated based on either a brain atlas or ICA (Du et al., 2018). For example, Kim et al. trained a deep neural network (DNN) based on FNC, with L1-norm to monitor weight sparsity, achieved substantial performance improvement (Kim et al., 2016). Zeng et al. presented a sparse autoencoder to learn imaging site-shared FCs, which was then used to guide SVM training on multi-site datasets for schizophrenia (SZ) diagnosis (Zeng et al., 2018). Similarly, in order to exploit the wealth of temporal dynamic information in BOLD signals, recurrent neural networks (RNN)-based approaches have also been proposed to work on fMRI time series. Particularly, Yan et al. proposed multi-scale RNN on the TCs (Yan et al., 2017) and Dakka et al. adopted a recurrent convolutional neural network (R-CNN) on 4-D fMRI recordings at the whole-brain voxel level (Dakka et al., 2017) to distinguish patients with SZ from healthy controls (HCs). Moreover, dynamic FNC (dFNC) has also been adopted with or without combining with static FNCs to discriminate brain disorders, which can further improve prediction accuracy (Cetin et al., 2016; Du et al., 2017; Rashid et al., 2016).

However, despite the significant advances in fMRI-based classification, the complementary information between spatial-temporal coherence (FNC) and temporal dynamics of brain activity (TCs) have not been fully leveraged to take advantage of fMRI data. To our knowledge, there are no deep models yet combining both functional connectivity and activity as input features. To address this issue, we are motivated to propose a Hybrid Deep Learning framework integrating brain Connectivity and Activity (HDLFCA) together by combining DNN and C-RNN (convolutional recurrent neural network), aiming to enhance the classification performance for brain disorders by capitalizing on multi-domain neuroimaging information. The prediction outputs of the two neural networks were then concatenated to build a new feature matrix, generating the final decision by logistic regression (Fig. 1B).

Another point that needs to mention is the lack of interpretability of DL methods, which often limited their use in clinical contexts due to the 'black-box' nature of deep layers (Kohoutová et al., 2020). To this end, the attention mechanism, inspired by human perception, was developed to improve the interpretability of DL models, and has been employed in various medical imaging data mining cases. For instance, Lian et al. developed an attention-guided DL framework for dementia diagnosis (Lian et al., 2020), including a full CNN to localize the discriminative regions and a hybrid network to fuse multi-level spatial information. Similarly, Jin et al. proposed an attention-based 3D CNN for Alzheimer's disease diagnosis (Jin et al., 2020). However, most existing attention-guided DL studies focused on structural images such as structural MRI (sMRI) and Computed Tomography (CT) (Chen et al., 2020; Dong et al., 2019; Lei et al., 2020), less attention has been paid to fMRI data due to its higher dimensionality. In this work, we propose two schemes to improve the interpretability: 1) to develop an attention-guided C-RNN for TCs, i.e., C-RNN^{AM}, which enables learning of temporal dynamics and identification of the most discriminative TC nodes (ICs) integrated into a unified framework (Fig. 1C). 2) In parallel, layer-wise relevance propagation (LRP) was applied to DNN layers, searching for the most discriminative FNC patterns. Taken together, the most contributing fMRI features for group discrimination were identified and visualized, improving the whole model interpretability.

To validate the effectiveness of our proposed method, HDLFCA, rigorous comparisons have been made with 12 popular methods. Specifically, we compared with 8 alternative models based on static FNC or TCs and 4 DL methods using dynamic FNC, which also characterized functional connectivity and dynamics of BOLD signals simultaneously. These tests were performed using In-House multi-site dataset (558 SZ and 541 HCs) and public ABIDE datasets (743 ASD and 779 HCs). Experimental results showed our method outperformed 12 alternative models by 2.8–8.9%, achieving SZ-HC classification accuracy at 85.1% and 81.0% for the multi-site pooling and leave-one-site-out respectively, and 72.4% for ABIDE dataset with multi-site pooling. More importantly, the most group discriminative brain regions can be easily traced back with convincing biological interpretability, suggesting the great promise of HDLFCA to identify potential imaging biomarkers.

2. Materials and methods

2.1. Participants

For In-House dataset, participants (558 schizophrenia patients and 542 HCs) were recruited from 7 hospitals, including Peking University Sixth Hospital (PKU6), Beijing Huilongguan Hospital (HLG), Xinxiang Hospital Simens (XX#1), Xinxiang Hospital GE (XX#2), Xijing Hospital (XJ), Renmin Hospital of Wuhan University (RWU) and Zhumadian Psychiatric Hospital (ZMD). Demographic and clinical information of subjects were listed in Table 1 and Table S1. All patients with SZ are diagnosed by experienced psychiatrists using the Structured Clinical Interview for DSM-IV-TR Disorders. All HCs are interviewed using the SCID-Non-Patient Version and excluded if their first-degree relatives had any psychotic disorders. Besides, none of the participants had neurological disorders, substance abuse or dependence, pregnancy, and prior electroconvulsive therapy or head injury resulting in loss of consciousness. The severity of positive and negative symptoms was assessed according to PANSS scores. Two sample t-test and Chi-square test were performed to measure the difference of age and gender between HCs and patients respectively. This study has been approved by the ethical committees and all subjects provided written informed consent, including permission to share data between centers.

For public ABIDE dataset (743 patients with ASD and 779 HCs), the detailed demographic information of datasets was listed in Table S14.

2.2. Image acquisition

For all sites in In-House datasets, scanning parameters are as follows: repetition time (TR) = 2000 ms; echo time (TE) = 30 ms; flip angle (FA) = 90°; field of view (FOV) = 220 × 220mm; matrix = 64 × 64; slice thickness = 4 mm; gap = 0.6 mm; slices = 33. The resting-state fMRI data were collected on a 3T Tim Trio scanner (Siemens) in PKU6, HLG and XJ sites, Verio scanner (Siemens) in XX#1 site, 3T Signa HDx GE scanner (General Electric) in the other sites. Subjects were instructed to lie still, keep their eyes closed, stay awake, and minimize head movement with foam padding and earplugs. Details of all sites were listed in Table S2.

2.3. Data preprocessing

All resting-state fMRI data were preprocessed with the same procedures as we did in Liu et al. (2019) using the SPM software package (http://www.fil.ion.ucl.ac.uk/spm/). The first ten volumes of each scan time series were discarded for magnetization equilibrium. The following processing pipeline was then performed: 1) slice timing correction to the middle slice; 2) motion correction to the first image; 3) normalization into the standard Montreal Neurological Institute (MNI) space, and resliced to 3×3×3 mm; 4) denoising and spatially smoothing using an 8 mm full width half max (FWHM) Gaussian kernel.

To control the effects of motion artifacts, each subject has been evaluated with a maximum displacement that did not exceed ± 3 mm (translation) or ± 3° (rotation). The group difference in the mean framewise displacement (FD) between HC and SZ groups was not significant (HC: 0.137 ± 0.071, SZ: 0.142 ± 0.085, two-sample t-test: $p = 0.98$).

2.4. Feature extraction

Imaging data were decomposed into spatial functional networks and back-reconstructed using Group-guided independent component analysis (GIG-ICA) (Calhoun et al., 2001; Du et al., 2016; Du and Fan, 2013; Du et al., 2020) in the GIFT software (http://trendscenter.org/software/gift). We chose a high model order ICA (number of components = 100) to decompose the functional networks showing temporally coherent activity as our previous work (Luo et al., 2020; Zhi et al., 2018). For subject-level data, 150 principal components were retained by principal component analysis (PCA). For group-level data, acquired by concatenating subject data across time, 100 principal components were retained using PCA again. Afterward, the Infomax ICA algorithm was repeated 20 times using ICASSO followed by selection of the most representative result, to improve the reliability of the decomposition, resulting in 100 stable group ICs (Du et al., 2014; Yan et al., 2021). 50 ICs were further selected and characterized as intrinsic connectivity networks, which showed higher low-frequency spectral power and presented minimal overlap with white matter, ventricles, and edge regions (Allen et al., 2011). The 50 spatial maps are sorted into eight domains as listed in Fig. S1. Furthermore, subject-specific time courses and spatial maps were back-reconstructed using GIG-ICA (Du et al., 2016; Du and Fan, 2013). The following additional post-processing steps were performed on the selected component TCs: linear, quadratic and cubic detrending, regressing out six realignment parameters and their temporal derivatives, despiking, and low-pass filtering (<0.15 Hz).

As shown in Fig. 1, the subject-level TCs with a size of 50×170 (ICs × time points) are used as the input of the RNN-based model. Pearson's correlation between TCs of each pair of ICs was calculated, yielding a symmetric connectivity matrix of 50×50. The FNC matrix was further reshaped into a vector with a dimension of (50 × 49)/2 = 1225 using the upper triangle elements, which were used as input features of DNN.

2.5. Methods

2.5.1. Hybrid deep learning framework integrating brain connectivity and activity (HDLFCA)—

As shown in Fig. 1B, we proposed a Hybrid Deep Learning Framework integrating brain Connectivity and Activity (HDLFCA) to enhance the performance for brain disorder classification by taking advantage of both temporal coherence and dynamic neuroimaging information. In the first stage, different DL models were designed to characterize heterogeneous features and leverage complementary information between TCs and FNC. Specifically, we used the C-RNN^{AM} to capture time-varying fluctuations in fMRI time series, with the attention module integrated to automatically extract the most discriminative TCs. Meanwhile, we used DNN to learn functional interaction between ICs, where LRP was performed to identify the most group-discriminative FNC patterns. In the second stage, the outputs from the above two models were concatenated to create a new feature matrix to train a logic regression, whose output is the final decision. 10-fold cross-validation was conducted to evaluate the performance of models. The implementation details were depicted in section 2.6.

2.5.2. Convolutional recurrent neural network with attention module (C-RNN^{AM})

1) Overview: As shown in Fig. 1C, the C-RNN^{AM} network consists of an attention module, three 1D convolutional layers with different kernel sizes, one concatenation layer, one max pooling layer, two gated recurrent unit (GRU) layers, and a fully connected layer. The processed TCs were fed to the C-RNN^{AM} network to generate the intermediate prediction $P_1 \in R^{N\times 1}$, where N is the number of training samples.

Although RNN has great power in sequence modeling, it is still challenging for it to deal with high dimension spatiotemporal fMRI data with lots of redundant information. To solve this problem, we first used Conv1D layers as an 'encoder' to learn correlations between brain regions, followed by max-pooling layer. The Conv1D layers extract local information from neighboring time points in the space dimension and the pooling layer downsample data in the time dimension (Roy et al., 2019; Yan et al., 2019). Considering the brain dynamics at different timescales can capture distinct aspects of human behavior (Liegeois et al., 2019), we expanded simple convolution layers by applying multiple Conv1D layers with different kernel sizes so that the next stage would aggregate dynamic brain activity from multiple time scales simultaneously. Since the filter lengths vary exponentially rather than linearly (Szegedy et al., 2015), we set the size of three convolutional filters as 32×2×50 (number of filters × time scales × ICs), 16×4×50 and 16×8 ×50, resulting in three feature maps with a size of 170×32 (time scales× ICs×number of filters), 170×16 and 170×16 respectively. A concatenation layer was followed to integrate features with different time scales. Furthermore, a max-pooling layer was performed to downsample along the time axis with 3×1 kernel size, resulting in 56×64 features (time points ×feature dimension) as the input of GRU layers.

Considering the brain activity is characterized by long-range temporal dependence such that signal fluctuations at the present time influence signal dynamics up to several minutes in the future (Dhamala et al., 2020; Guclu and van Gerven, 2017), while conventional RNNs often fail to learn long-term dependencies due to the gradient exploding and vanishing problems during the back-propagation (Bengio et al., 1994). Therefore, we proposed to utilize GRU layers to learn useful representations of brain activity patterns, which can mitigate the gradients problem by controlling information flow with gating mechanisms (Roy et al., 2019). In this study, two GRU layers were stacked in the HDLFCA to capture both short- and long-term dependencies in BOLD time series. It is worth noting that each GRU layer was densely connected to the other GRU layers to mitigate the degradation problem, which provided short-cut paths during back-propagation (Huang et al., 2017). The size of hidden states units was set as 32. To make full use of brain activity throughout the scan, the GRU outputs were further averaged, and two fully-connected layers were followed to give the intermediate prediction, which was then concatenated for the final decision.

2) Attention Module: The attention module was proposed to increase representation power and improve interpretability by focusing on important brain regions and suppress unnecessary ones. The schematic of attention module is illustrated in Fig. 1C. Given the previously processed TCs $X \in R^{170\times 50}$ as input, where 170 and 50 are the number of time

points and ICs, the attention module generated an attention map $M(X) \in R^{50\times1\times1}$. The attention process can be defined as follows:

$$X' = B(M(X)) \otimes X$$

where $\otimes$ denotes element-wise multiplication and $B(\cdot)$ denotes broadcast operations : the attention values $M(X)$ was copied along time dimension accordingly and then reshaped into the same size with X' is the refined feature.

To construct the attention module, TCs inputs were reshaped into a matrix of size 50×1×170. The average-pooling calculates the mean value of all elements in the pooling region, and may reduce the contrast of the new feature map, while max-pooling only uses the maximum element and ignores the others, which may be useful for classification tasks (Yu et al., 2014). Therefore, we adopted both of these along the time axis to learn temporal statistics and aggregate temporal information fully (Woo et al., 2018). After that, two temporal context descriptors: $F^{\max}$and F^{avg}, which denote max-pooled features and average-pooled features respectively, were generated and were concatenated to produce an efficient feature descriptor. We applied a convolution layer and sigmoid activation to produce an attention map. Note that the size of filter is 50×1, which has the same dimension as the number of ICs rather than a smaller size to extract global relations among ICs. And the number of filters is 50, each of them was responsible for learning the importance of one IC. Integrated in the unified framework, the attention map tells 'which region' is an informative part, namely, the greater the weight of the attention map, the higher the discrimination power of the brain region. To sum up, the attention module can be denoted as follows:

$$\begin{aligned} M(X) &= \sigma(conv([AvgPool(X);\ MaxPool(X)])) \\ &= \sigma\left(conv\left(F^{avg};\ F^{\max}\right)\right) \end{aligned}$$

where σ is the sigmoid function.

2.5.3. Deep neural network (DNN)—Given the FNC as input, the deep neural network was applied to learn high-level hierarchical feature representation and give the intermediate prediction $P_2 \in R^{N\times1}$. DNN was composed of one input layer, two hidden layers, and one output layer. The size of hidden notes was set 32 and 16 respectively. L_2 norm regularization and dropout strategies were used to avoid overfitting as reported in (Srivastava et al., 2014).

Based on the trained models, LRP was introduced to identify important FNC patterns for classification decisions, and it decomposed the prediction of DNN over a test sample down to relevance scores for the single input dimensions such as each FNC here. Supposing there arelayers in total, the relevance of output neuron can be obtained in a feed-forward fashion: $R_1^{(M)} = f(x)$. $\beta - rule$ was performed to compute the propagation of relevance from layer $l+1$ to layer l

$$R_i^{(l,l+1)} = \left((1+\beta)\frac{z_{ij}^+}{z_j^+} - \beta\frac{z_{ij}^-}{z_j^-} \right) R_j^{(l+1)}$$
$$z_{ij} = x_i w_{ij}, z_j^+ = \sum_i z_{ij}^+ + b_j^+, z_j^- = \sum_i z_{ij}^- + b_j^-$$

where z_{ij}^+ and z_{ij}^- denotes positive and negative activations respectively. b_j^+ and b_j^- denote the positive and negative part of the bias item $b_j \cdot R_j^{(l+1)}$ and $R_{i \leftarrow j}^{(l,l+1)}$ denotes the relevance of a neuron j at layer $l+1$, and message between neurons i at the layer l and neurons i at layer $l+1$ respectively. β controls how much inhibition is incorporated into the relevance redistribution. Then the relevance of a neuron i at layer l was defined by summing messages from neurons at layer $l+1$:

$$R_i^{(l)} = \sum_{j \in (l+1)} R_{i \leftarrow j}^{(l,l+1)}$$

Therefore, the relevance score $R_d^{(1)}$ of each FNC was determined by this rule. For more details on LRP, please refer to (Bach et al., 2015).

2.6. Implementation details

The HDLFCA was implemented via nested cross-validation using the Keras package (https://keras.io/). In each one of the 10 fold experiment, the 3-fold cross-validation was performed further to avoid overfitting. Specifically, training data was divided into three folds further in the training stage, where two folds were used for training and validation, and the remaining one was used for prediction. After 3-fold cross-validation, predictions from three DNN models were concatenated to constitute intermediate prediction P1 and so does C-RNN^{AM} to generate P2, which were used for the final decision. In the testing stage, the outputs of three DNN models and three C-RNN models were first averaged respectively, then two predictions were concatenated to build the final decision by logistic regression. The procedures of the training and testing phase were illustrated in Fig. S4. An implementation for HDLFCA is available at https://github.com/minzhaoCASIA/HDLFCA.

The C-RNN model was trained by the Adam optimizer with an initial learning rate of 0.001 and decayed with the rate of 0.01. Dropout (0.5) and $L_{1,2}$-norm regularization (L1 = 0.0001, L2 = 0.0001) were performed to control weight sparsity. The batch size was set at 64. The DNN model was trained with the cross-entropy loss by the Adam optimizer with an initial learning rate of 0.001. The performance of methods was evaluated by five metrics including accuracy (ACC), specificity (SPE), sensitivity (SEN), F1-score (F1) and area under the receiver operating characteristic curve (AUC). The performance of different algorithms was compared via a two-sample t-test.

3. Results

3.1. Multi-site pooling classification

Ten-fold multi-site pooling experiments were conducted to evaluate classification performance, where fMRI data from all sites were pooled together and ten-fold cross-validation was performed. All experiments were repeated 10 times to generate mean and standard deviations of metrics. We compare HDLFCA with eight competing methods on both In-House and ABIDE datasets. The quantitative results in the task of classification are reported in Table 2, Table 3 and Fig. 2.

As shown in Fig. 2, *first*, the HDLFCA reported a mean classification accuracy of 85.3% and 72.4% on In-House and ABIDE datasets, indicating a significant improvement over the other classical classifiers ($p<0.01$). For instance, HDLFCA achieved an improvement of 8.9%, 8.3% and 3.8% in ACC compared with Random Forest, AdaBoost and SVM, respectively on In-House datasets. This implied the significant effectiveness of learning high-level, "deep" features from fMRI data. *Second*, compared with BrainNetCNN, DNN, C-RNN and C-RNN^{AM} that adopted features of either FNC or TC only, the proposed HDLFCA that exploits complementary information between them led to a better diagnostic performance on two datasets. For example, in terms of ACC, an improvement of 5.2%, 4.4%, 2.8% and 1.8% was achieved on HC-SZ datasets respectively, and an improvement of 3.9%, 2.0%, 3.3% and 3.0% was achieved for ABIDE datasets, suggesting the necessity and validity of integrating functional dependency between brain regions and temporal dynamics of brain activity. *Third*, the comparative performance of C-RNN^{AM} and C-RNN in SZ classification showed that C-RNN^{AM} achieved an improvement of about 1% in terms of ACC, SPE, SEN and F1 values, demonstrating that incorporation of discriminative IC localization and disease classification into a unified framework boosts the final performance. It should be noted that although the attention module identified the discriminative ICs as well as improved performance, it did not cause an increase in model complexity. *Forth*, our HDLFCA outperformed the connectivity-based graph convolutional network (cGCN) (Wang et al., 2021) significantly on two datasets as well, which also used TCs and FCs to extract similar connectome features.

Furthermore, to validate the generalizability of HDLFCA, we reproduce the experiments based on TCs obtained from Automated Anatomical Labeling (AAL) template instead of ICA, where the mean regional TCs were calculated by averaging the voxel-wise fMRI time series in each of brain regions of interests (ROI). Pearson's correlation between TCs of each pair of ROIs was calculated, yielding a symmetric connectivity matrix of 116×116. The results were reported in Table S5 and Fig. 2C. We can draw a similar conclusion as above. Particularly, HDLFCA outperformed single feature-based deep learning models (i.e., DNN, C-RNN and C-RNN$^{AM)}$ largely, demonstrating the superiority of utilizing complementary information between FNC and TCs. The attention module also yielded better classification performance (3.6% in ACC) compared with C-RNN. The HDLFCA based on ICA showed a little better performance (85.3%) than fixed AAL (84.9%), this is likely due to the ability of ICA to capture variability in the components among subjects.

3.2. Leave-one-site-out classification

In the leave-one-site-out transfer classification, one imaging site was considered as the testing dataset while the other sites were used for training, with 10% of the samples chosen randomly for validation in the HDLFCA. The quantitative results on In-House dataset were shown in Table 4, Table S3 and Fig. 2C. We can draw a similar conclusion as that in Section 3.1. That is, compared with the conventional machine learning approaches (i.e., Random Forest, AdaBoost and SVM), the proposed HDLFCA largely improved the diagnostic performance, suggesting that automatically learning high-level fMRI features is beneficial for SZ classification. Besides, HDLFCA resulted in ACC improvement at 5.7%, 4.7%, 3.9%, and 2.6% respectively compared to single-feature-based deep learning models (i.e., BrainNetCNN, DNN, C-RNN and C-RNN^{AM}). This demonstrated the superiority of integrating FNC and TCs. In addition, from the Table 4, the embedded attention module still yielded better classification performance, which is consistent with the results reported in Section 3.1. It further indicated that it not only identified the discriminative ICs but also improved the classification performance. The HDLFCA still outperformed cGCN, suggesting our method are more powerful to capture functional connectivity and dynamic brain activity underlying the fMRI data.

3.3. Most HC-SZ discriminative FNC

The contribution of each FNC was rendered using the LRP algorithm by propagating the correlation layer by layer. The top 50, 70 and 100 contributing FNC features in the task of SZ diagnosis were presented in the circle diagram (Fig. 3A), where the 50 ICs were divided into eight functional networks (Fig. S1). The discriminative FNC showed diffuse patterns widely across the entire brain, implying widely impaired brain regions in SZ patients. Despite the complexity, we observed that default-mode networks with connections to frontal, and attentional networks shared a high proportion in the top 50 contributing connectivity, which are reported to be highly associated with SZ. In Fig. 3A, the comparison of top 50 and top 70 contributing FNC revealed a substantial increase in connections within visual networks. Connections between frontal and default mode networks, frontal and attention networks, and connections within visual networks indicated the most contributing influence when presenting the top 100 contributing FNC, suggesting that schizophrenia is characterized by impairments in high-level cognitive and emotional processing circuits.

3.4. Most discriminative independent components captured by attention module

The attention module can automatically identify discriminative brain regions by learning which regions to focus or suppress. An attention value map with a 50×1×1 size was obtained for each subject and the mean attention map was generated by averaging them, where a higher value indicates the greater discrimination power of the IC. To obtain more robust imaging markers, we repeated the 10-fold cross-validation experiments 10 times (10*10 trained models in total) and counted the frequency of the top 10 discriminative ICs. Fig. 3B displays the frequency distribution histogram, where only ICs with an occurring frequency greater than 10% are shown. Fig. 3B also displays the spatial maps of the top 10 discriminative ICs, in which the striatum, cerebellum and anterior cingulate were highlighted as the three most SZ-discriminating ICs by the attention module, suggesting

that the attention scheme can effectively extract useful information from whole-brain fMRI features. It should be noted that Fig. 3B presents the group-discriminative ICs by averaging the attention maps for each subject, but they are not totally the same across all subjects, for example, the same ICs may be emphasized differently, implicating the potential for individualized localization of brain regions.

3.5. Comparison with dynamic FNC features(dFNC)

Since dFNC also simultaneously characterized functional dependency and temporal dynamics of spontaneous BOLD signal, we also compared other deep learning methods using dFNC with our proposed HDLFCA, which also integrated dynamic FCs and TCs to improve classification performance. The dFNC was computed by the sliding window method in steps of 1 TR. We conducted multiple experiments under different settings, where the window length varies from the 30s to 70s at intervals of 10s (15–35 TR). A comparison of classification performance was reported in Table 5. More details are available in the supplementary materials (Table S4 and Figure S2).

From Table 5 and Table S4, we can observe that the proposed HDLFCA outperformed the best performing dFNC-based DL methods in all metrics significantly ($p<0.01$). For instance, in terms of ACC, HDLFCA achieved an improvement of 4.6%, 4.9%, 4.5% and 5.5% compared with the best results achieved by LSTM, BiLSTM, GRU, and C-LSTM respectively, suggesting the superiority of our method. The lower performance of C-LSTM compared to LSTM may be attributed to the high dimension of the FNC vector (1225, compared to 50 in previous TC-based methods), which largely increased the parameters of the model. Furthermore, GRU based on dFNC outperformed the same neural network based on TCs significantly, which only contains temporal dynamics of brain activity, suggesting the effectiveness to integrate brain connectivity and activity of rs-fMRI data.

3.6. Comparison with different DL architectures

In this section, we compared the proposed C-RNN^{AM} with eight alternative deep learning models in multi-site pooling experiments on In-House datasets. The results were reported in Table 6. Considering the great power in sequence modeling of RNN and the rich temporal dynamics of brain activity in time series of BOLD-signal, we first directly applied simple RNN and GRU in the same settings to classify brain disorders. The results showed the GRU models achieved an improvement of 23.6% in ACC, possibly because simple RNN is difficult to learn long-term dependencies due to the vanishing and exploding gradient problem (Bengio et al., 1994) and the brain activity is characterized by long-range temporal dependence such that signal fluctuations at the present time influence signal dynamics up to several minutes in the future (Dhamala et al., 2020; Guclu and van Gerven, 2017). The C-RNN further outperformed GRU and C-MLP, potentially because the convolutional and GRU layers were responsible for capturing spatial and temporal information respectively. The C-RNN with multi-scale convolution kernel size outperformed the S_C-RNN with single-scale convolution kernel, suggesting that extracting dynamics from a variety of timescales is useful in fMRI data.

Moreover, we designed 4 variants of attention mechanism integrated into C-RNN models. The architectures were illustrated in Fig. S5. Specifically, C-RNN^{AM} achieved a light increase compared with AM_1, suggesting capturing global relations between brain networks is more effective than local relations. AM_3 performed worse than others, showing that the emphasizing important brain regions play an essential role in brain disorder classification.

4. Discussion

In this study, we proposed a novel unified DL framework by integrating temporal coherence and dynamics effectively to classify brain disorders. The classification accuracy of 85.1% and 81.0% were achieved in multi-site pooling and leave-one-site-out respectively in the task of HC-SZ discrimination. Moreover, when using publicly accessible ABIDE dataset, ACC of 72.4% was achieved in the multi-site pooling classification of HC vs. ASD, which significantly outperformed multiple single feature-based methods. The competitive result is comparable to, if not better than, the recent studies on large multi-site fMRI datasets (Kim et al., 2016; Yan et al., 2019; Zeng et al., 2018). Additionally, LRP and an attention module were introduced to identify the most discriminative FNC patterns and brain regions for SZ. To the best of our knowledge, this is the first attempt to integrate identification of discriminative brain regions and diagnosis of brain disorders into a unified framework based on fMRI data using an attention mechanism-based network.

Recently, numerous studies have applied deep learning methods for SZ classification and achieved high performance. Compared with previous studies (Dakka et al., 2017; Rozycki et al., 2018; Skåtun et al., 2017), this work achieved an improvement (>5.0%) in accuracy on multi-site pooling and leave-one-site-out classification. The promising results may derive from the following aspects: First, we combined different powerful deep learning models to leverage complementary information between TCs and FNC, where the TCs neglects the functional dependency between brain regions and FNC discards sequential temporal dynamics. The experimental results demonstrated the superiority of combing multiple features. Second, the attention module helps to refine and optimize feature representation by focusing on more important brain regions instead of the full feature. The experimental results also showed the attention module improved classification performance. Third, since the convolutional neural network (CNN) is ‘deep in space’ and RNN is ‘deep in time’, both of them were applied to make full use of the spatial and temporal information underlying the spontaneous BOLD signal. Furthermore, to validate the superiority of our method, the HDLFCA was compared with other deep learning methods based on dFNC, which also takes dynamic fluctuation and temporal coherence into consideration. Our method achieved an improvement (>4.0%) of average accuracy. Importantly, the goal of our method is not only to focus on high performance, but also to provide results that are interpretable and provide insight into the brain. The attention module provides an effective way to explore underlying biomarkers in DL methods. It allows for the integration of discriminative ICs localization and SZ diagnosis into a unified framework, since the isolated informative region identification may lead to suboptimal performance. What’s more, the discriminative ICs are not totally the same across all subjects, showing the importance of individualized localization of brain regions associated with schizophrenia.

The results revealed that the attention module highlighted brain regions at the locations of the striatum, cerebellum and anterior cingulate. The striatum, including putamen and caudate, has been proved to play a vital role in the pathophysiology of schizophrenia (Yan et al., 2019). Compelling evidence has shown that the striatum was involved in cognition domains, including motor, decision-making, and stimulus-response learning (Yager et al., 2015). Recently, numerous findings converged on evidence for both an increase in striatal dopamine and striatal dopamine receptors. The dopaminergic hyperfunction in the striatum may contribute to cognitive deficits in SZ (McCutcheon et al., 2019). Moreover, the increase of D2 receptors was found to be predictive for treatment response and the popular antipsychotics usually blocks the dopamine D2 receptors in the striatum (Li et al., 2020; Sarpal et al., 2016). Another highlighted component was the cerebellum. Many studies showed significant evidence for cerebellar abnormalities in SZ, such as impairment white matter integrity and blood flow decrease in the cerebellum during cognition tasks(Andreasen and Pierson, 2008; Kim et al., 2014; Luo et al., 2018; Yan et al., 2021). In addition, the other important component identified by attention module was located in the anterior cingulate cortex (ACC). Previous studies have demonstrated that a failure of functional ACC is associated with disturbed cognitive control and working memory deficits in SZ greatly (Fletcher et al., 1999; Fletcher et al., 1996) and SZ patients exhibit significantly reduced ACC activation (Schultz et al., 2012). Overall, the most group discriminative brain regions can be easily traced back with convincing biological interpretability, implying that the attention module emphasized important ICs effectively and our method showed great promise to identify potential imaging biomarkers.

Although the proposed HDLCD achieved high performance in discriminative ICs localization and psychotic disorder classification, several limitations should be considered in the future. First, C-RNN^{AM} and DNN were trained independently and then their predictions were fed into meta-learner to utilize complementary information between TCs and FNC, which makes the later fusion stage couldn't help refine feature representations in the first stage. A promising direction is to integrate the two stages into a purely end-to-end framework to provide complementary guidance for each other. Second, static FNC as the most commonly used functional connectivity feature, was combined with brain activity (TCs) as input features in this work. Nevertheless, it is interesting to investigate whether combining dynamic connectivity and brain activity can further advance classification performance in the future.

5. Conclusions

In this work, we proposed HDLFCA, a unified framework that takes fully advantage of temporal coherence (FNCs) and time-varying fluctuations (TCs) jointly to classify psychiatric disorders based on rs-fMRI data. The method was validated on both In-House SZ dataset (n = 1100) and the public ABIDE datasets (n = 1552), with 2.8–8.9% increase compared to 12 popular classifiers, suggesting the superiority of combining multiple features. To the best of our knowledge, this is the first attempt to introduce an attention module into a C-RNN based framework to improve the classification performance and automatically identify discriminative brain regions. Such a method shows the potential

for deep learning to provide utility for both predicting and understanding the healthy and disordered brain.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by The Natural Science Foundation of China (82022035), the National Key Research and Development Program of China (2017YFA0105203), and the Natural Science Foundation of China (61773380, 82001450, 62076157), Beijing municipal science and technology commission (Z181100001518005), China Postdoctoral Science Foundation (BX20200364), the National Institute of Health grants (R01EB005846, R01MH117107, R01MH118695) and the National Science Foundation (2112455).

References

Allen EA, Erhardt EB, Damaraju E, Gruner W, Segall JM, Silva RF, Havlicek M, Rachakonda S, Fries J, Kalyanam R, 2011. A baseline for the multivariate comparison of resting-state networks. Front. Syst. Neurosci 5 (2).

Andreasen NC, Pierson R, 2008. The role of the cerebellum in schizophrenia. Biol. Psychiatry 64, 81–88. [PubMed: 18395701]

Andreou C, Borgwardt Stefan, 2020. Structural and functional imaging markers for susceptibility to psychosis. Mol. Psychiatry 1–13.

Bach S, Binder A, Montavon G, Klauschen F, Müller K-R, Samek W, 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One 10, e0130140. [PubMed: 26161953]

Bengio Y, Simard P, Frasconi P, 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Networks 5, 157–166. [PubMed: 18267787]

Calhoun VD, Adali T, 2006. In: Unmixing fMRI with Independent Component Analysis, 25. IEEE Engineering in Medicine Biology Magazine, pp. 79–90.

Calhoun VD, Adali T, 2012. Multisubject independent component analysis of fMRI: a decade of intrinsic networks, default mode, and neurodiagnostic discovery. IEEE Rev. Biomed. Eng 5, 60–73. [PubMed: 23231989]

Calhoun VD, Adali T, Pearlson GD, Pekar J, 2001. A method for making group inferences from functional MRI data using independent component analysis. Hum. Brain Mapp 14, 140–151. [PubMed: 11559959]

Cetin MS, Houck JM, Rashid B, Agacoglu O, Stephen JM, Sui J, Canive J, Mayer A, Aine C, Bustillo JR, 2016. Multimodal classification of schizophrenia patients with MEG and fMRI data using static and dynamic connectivity measures. Front. Neurosci. 10, 466. [PubMed: 27807403]

Chen X, Yao L, Zhang Y, 2020. Residual attention u-net for automated multi-class segmentation of covid-19 chest ct images. arXiv preprint arXiv:.05645.

Dakka J, Bashivan P, Gheiratmand M, Rish I, Jha S, Greiner R, 2017. Learning neural markers of schizophrenia disorder using recurrent neural networks. arXiv preprint arXiv:.00512.

Dhamala E, Jamison KW, Sabuncu MR, Kuceyeski A, 2020. Sex classification using long-range temporal dependence of resting-state functionalMRItime series. Hum. Brain Mapp. 41, 3567–3579. [PubMed: 32627300]

Dong X, Lei Y, Tian S, Wang T, Patel P, Curran WJ, Jani AB, Liu T, Yang X, 2019. Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network. Radiother. Oncol. 141, 192–199. [PubMed: 31630868]

Du W, Ma S, Fu G-S, Calhoun VD, Adali T, 2014. A novel approach for assessing reliability of ICA for FMRI analysis. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 2084–2088.

Du Y, Allen EA, He H, Sui J, Wu L, Calhoun VD, 2016. Artifact removal in the context of group ICA: A comparison of single-subject and group approaches. Hum. Brain Mapp. 37, 1005–1025. [PubMed: 26859308]

Du Y, Fan Y, 2013. Group information guided ICA for fMRI data analysis. Neuroimage 69, 157–197. [PubMed: 23194820]

Du Y, Fu Z, Calhoun VD, 2018. Classification and prediction of brain disorders using functional connectivity: promising but challenging. Front. Neurosci. 12, 525. [PubMed: 30127711]

Du Y, Fu Z, Sui J, Gao S, Xing Y, Lin D, Salman M, Abrol A, Rahaman MA, Chen J, 2020. NeuroMark: an automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders. NeuroImage 28, 102375. [PubMed: 32961402]

Du Y, Pearlson GD, Lin D, Sui J, Chen J, Salman M, Tamminga CA, Ivleva EI, Sweeney JA, Keshavan MS, 2017. Identifying dynamic functional connectivity biomarkers using GIG-ICA: Application to schizophrenia, schizoaffective disorder, and psychotic bipolar disorder. Hum. Brain Mapp. 38, 2683–2708. [PubMed: 28294459]

Fletcher P, McKenna PJ, Friston KJ, Frith CD, Dolan RJ, 1999. Abnormal cingulate modulation of fronto-temporal connectivity in schizophrenia. Neuroimage 9, 342.

Fletcher PC, Frith CD, Grasby PM, Friston KJ, Dolan RJ, 1996. Local and distributed effects of apomorphine on fronto-temporal function in acute unmedicated schizophrenia. J. Neurosci. Methods 16, 7062.

Guclu U, van Gerven MAJ, 2017. Modeling the dynamics of human brain activity with recurrent neural networks. Front. Comput. Neurosci. 11. [PubMed: 28326032]

Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

Jafri MJ, Pearlson GD, Stevens M, Calhoun VD, 2008. A method for functional network connectivity among spatially independent resting-state components in schizophrenia. Neuroimage 39, 1666–1681. [PubMed: 18082428]

Jin D, Zhou B, Han Y, Ren J, Han T, Liu B, Lu J, Song C, Wang P, Wang D, 2020. Generalizable, reproducible, and neuroscientifically interpretable imaging biomarkers for Alzheimer's disease. Adv. Sci, 2000675.

Kim D-J, Kent JS, Bolbecker AR, Sporns O, Cheng H, Newman SD, Puce A, O'Donnell BF, Hetrick WP, 2014. Disrupted modular architecture of cerebellum in schizophrenia: a graph theoretic analysis. Schizophr. Bull. 40, 1216–1226. [PubMed: 24782561]

Kim J, Calhoun VD, Shim E, Lee J-H, 2016. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. Neuroimage 124, 127–146. [PubMed: 25987366]

Kohoutová L, Heo J, Cha S, Lee S, Moon T, Wager TD, Woo C-W, 2020. Toward a unified framework for interpreting machine-learning models in neuroimaging. Nat. Protoc. 15, 1399–1435. [PubMed: 32203486]

Lei Y, Dong X, Tian Z, Liu Y, Tian S, Wang T, Jiang X, Patel P, Jani AB, Mao H, 2020. CT prostate segmentation based on synthetic MRI-aided deep attention fully convolution network. Med. Phys. 47, 530–540. [PubMed: 31745995]

Li A, Zalesky A, Yue W, Howes O, Yan H, Liu Y, Fan L, Whitaker KJ, Xu K, Rao G, 2020. A neuroimaging biomarker for striatal dysfunction in schizophrenia. Nat. Med. 26, 558–565. [PubMed: 32251404]

Lian C, Liu M, Pan Y, Shen D, 2020. Attention-guided hybrid network for dementia diagnosis with structural MR images. IEEE Trans. Cybern..

Liegeois R, Li J, Kong R, Orban C, Van De Ville D, Ge T, Sabuncu MR, Yeo BTT, 2019. Resting brain dynamics at different timescales capture distinct aspects of human behavior. Nat. Commun. 10. [PubMed: 30602777]

Liu S, Wang H, Song M, Lv L, Cui Y, Liu Y, Fan L, Zuo N, Xu K, Du Y, Yu Q, Luo N, Qi S, Yang J, Xie S, Li J, Chen J, Chen Y, Wang H, Guo H, Wan P, Yang Y, Li P, Lu L, Yan H, Yan J, Wang H, Zhang H, Zhang D, Calhoun VD, Jiang T, Sui J, 2019. Linked 4-way multimodal brain

differences in Schizophrenia in a large Chinese Han population. Schizophr. Bull. 45, 436–449. [PubMed: 29897555]

Luo N, Sui J, Abrol A, Chen J, Turner JA, Damaraju E, Fu Z, Fan L, Lin D, Zhuo C, 2020. Structural brain architectures match intrinsic functional networks and vary across domains: a study from 15 000+ individuals. Cereb. Cortex 30, 5460–5470. [PubMed: 32488253]

Luo N, Sui J, Chen J, Zhang F, Tian L, Lin D, Song M, Calhoun VD, Cui Y, Vergara VM, 2018. A schizophrenia-related genetic-brain-cognition pathway revealed in a large Chinese population. EBioMedicine 37, 471–482. [PubMed: 30341038]

McCutcheon RA, Abi-Dargham A, Howes OD, 2019. Schizophrenia, dopamine and the striatum: from biology to symptoms. Trends Neurosci. 42, 205–220. [PubMed: 30621912]

Rashid B, Arbabshirani MR, Damaraju E, Cetin MS, Miller R, Pearlson GD, Calhoun VD, 2016. Classification of schizophrenia and bipolar patients using static and dynamic resting-state fMRI brain connectivity. Neuroimage 134, 645–657. [PubMed: 27118088]

Roy S, Kiral-Kornek I, Harrer S, 2019. ChronoNet: a deep recurrent neural network for abnormal EEG identification. In: Conference on Artificial Intelligence in Medicine in Europe. Springer, pp. 47–56.

Rozycki M, Satterthwaite TD, Koutsouleris N, Erus G, Doshi J, Wolf DH, Fan Y, Gur RE, Gur RC, Meisenzahl EM, 2018. Multisite machine learning analysis provides a robust structural imaging signature of schizophrenia detectable across diverse patient populations and within individuals. Schizophr. Bull. 44, 1035–1044. [PubMed: 29186619]

Sarpal DK, Argyelan M, Robinson DG, Szeszko PR, Karlsgodt KH, John M, Weissman N, Gallego JA, Kane JM, Lencz T, 2016. Baseline striatal functional connectivity as a predictor of response to antipsychotic drug treatment. Am. J. Psychiatry 173, 69–77. [PubMed: 26315980]

Schultz CC, Koch K, Wagner G, Nenadic I, Schachtzabel C, Guellmar D, Reichenbach JR, Sauer H, Schlosser RGM, 2012. Reduced anterior cingulate cognitive activation is associated with prefrontal-temporal cortical thinning in schizophrenia. Biol. Psychiatry 71, 153.

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD, 2007. Dissociable intrinsic connectivity networks for salience processing and executive control. J. Neurosci. 27, 2349–2356. [PubMed: 17329432]

Skåtun KC, Kaufmann T, Doan NT, Alnæs D, Córdova-Palomera A, Jönsson EG, Fatouros-Bergman H, Flyckt L, KaSP I. Melle, 2017. Consistent functional connectivity alterations in schizophrenia spectrum disorder: a multisite study. Schizophr. Bull. 43, 914–924. [PubMed: 27872268]

Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR, 2009. Correspondence of the brain's functional architecture during activation and rest. Proc. Natl. Acad. Sci. 106, 13040–13045. [PubMed: 19620724]

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R, 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Supekar K, Musen M, Menon V, 2009. Development of large-scale functional brain networks in children. PLoS Biol. 7, e1000157. [PubMed: 19621066]

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A, 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9.

Wang L, Li K, Hu XP, 2021. Graph convolutional network for fMRI analysis based on connectivity neighborhood. Netw. Neurosci. 5, 95.

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

Yager LM, Garcia AF, Wunsch AM, Ferguson SM, 2015. The ins and outs of the striatum: role in drug addiction. Neuroscience 301, 529–541. [PubMed: 26116518]

Yan W, Calhoun V, Song M, Cui Y, Yan H, Liu S, Fan L, Zuo N, Yang Z, Xu K, 2019. Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site FMRI data. EBioMedicine 47, 543–552. [PubMed: 31420302]

Yan W, Plis S, Calhoun VD, Liu S, Jiang R, Jiang T-Z, Sui J, 2017. Discriminating schizophrenia from normal controls using resting state functional network connectivity: a deep neural network and

layer-wise relevance propagation method. In: 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, pp. 1–6.

Yan W, Zhao M, Fu Z, Pearlson GD, Sui J, Calhoun VD, 2021. Mapping relationships among schizophrenia, bipolar and schizoaffective disorders: a deep classification and clustering framework using fMRI time series. Schizophr. Res..

Yu D, Wang H, Chen P, Wei Z, 2014. Mixed pooling for convolutional neural networks. In: International Conference on Rough Sets and Knowledge Technology. Springer, pp. 364–375.

Zeng L-L, Wang H, Hu P, Yang B, Pu W, Shen H, Chen X, Liu Z, Yin H, Tan Q, 2018. Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI. EBioMedicine 30, 74–85. [PubMed: 29622496]

Zhi D, Calhoun VD, Lv L, Ma X, Ke Q, Fu Z, Du Y, Yang Y, Yang X, Pan M, 2018. Aberrant dynamic functional network connectivity and graph properties in major depressive disorder. Front. Psychiatry 9, 339. [PubMed: 30108526]

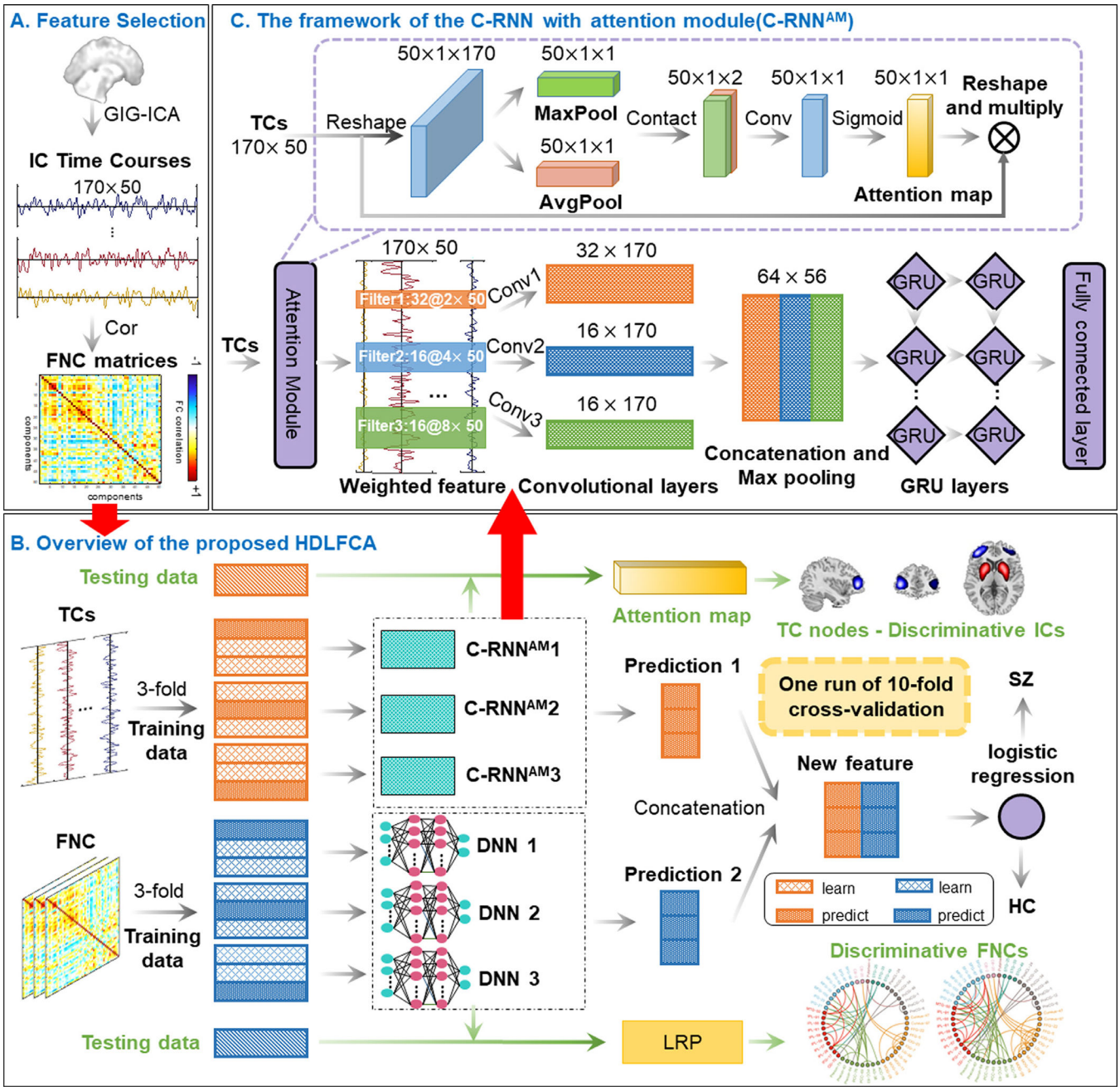

Fig. 1.

The framework of the proposed HDLFCA in psychotic disorder classification. (A) Data preprocessing and Feature extraction. TCs was obtained by decomposing fMRI data using GIG-ICA, and FNCs was estimated from the TCs. (B) Overview of our proposed HDLFCA. C-RNN^{AM} and DNN were used to characterize temporal dynamics in TCs and learn functional dependency between brain regions respectively. Then their predictions were concatenated to build a new feature matrix, generating the final decision by logistic regression. For model interpretability, attention module and layer-wise relevance propagation (LRP) were applied to identify the most discriminative ICs and FNC patterns

respectively. (C) Details of the C-RNN$^{AM.}$ It consists of an attention module, multiple 1D convolutional (Conv1D) layers, one concatenation and max pooling layer, two gated recurrent unit (GRU) layers and a fully connected layer. The purple frame shows the scheme of the attention module, which is trainable along with other modules. The greater the weight of the attention map, the more important the component was. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

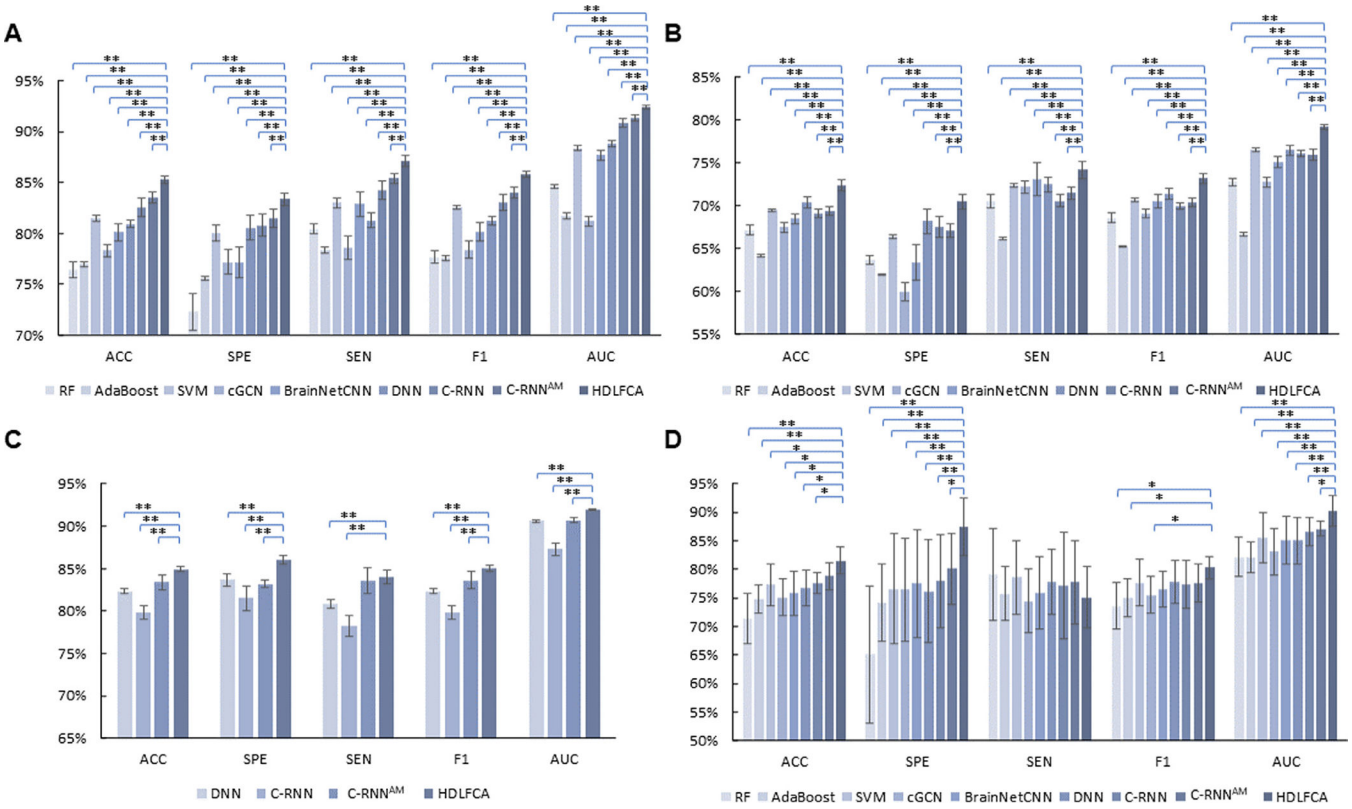

Fig. 2.

The classification results of (A) multi-site pooling classification in in-house SZ datasets, (B) multi-site pooling classification in public ABIDE datasets, (C) multi-site pooling classification based on TCs or FNCs extracted by AAL atlas in in-house SZ datasets, and (D) leave-one-site-out classification in HC-SZ datasets. */** denote that the proposed HDLFCA method achieves significantly better performance than the listed ones, with P value=0.05/0.01.

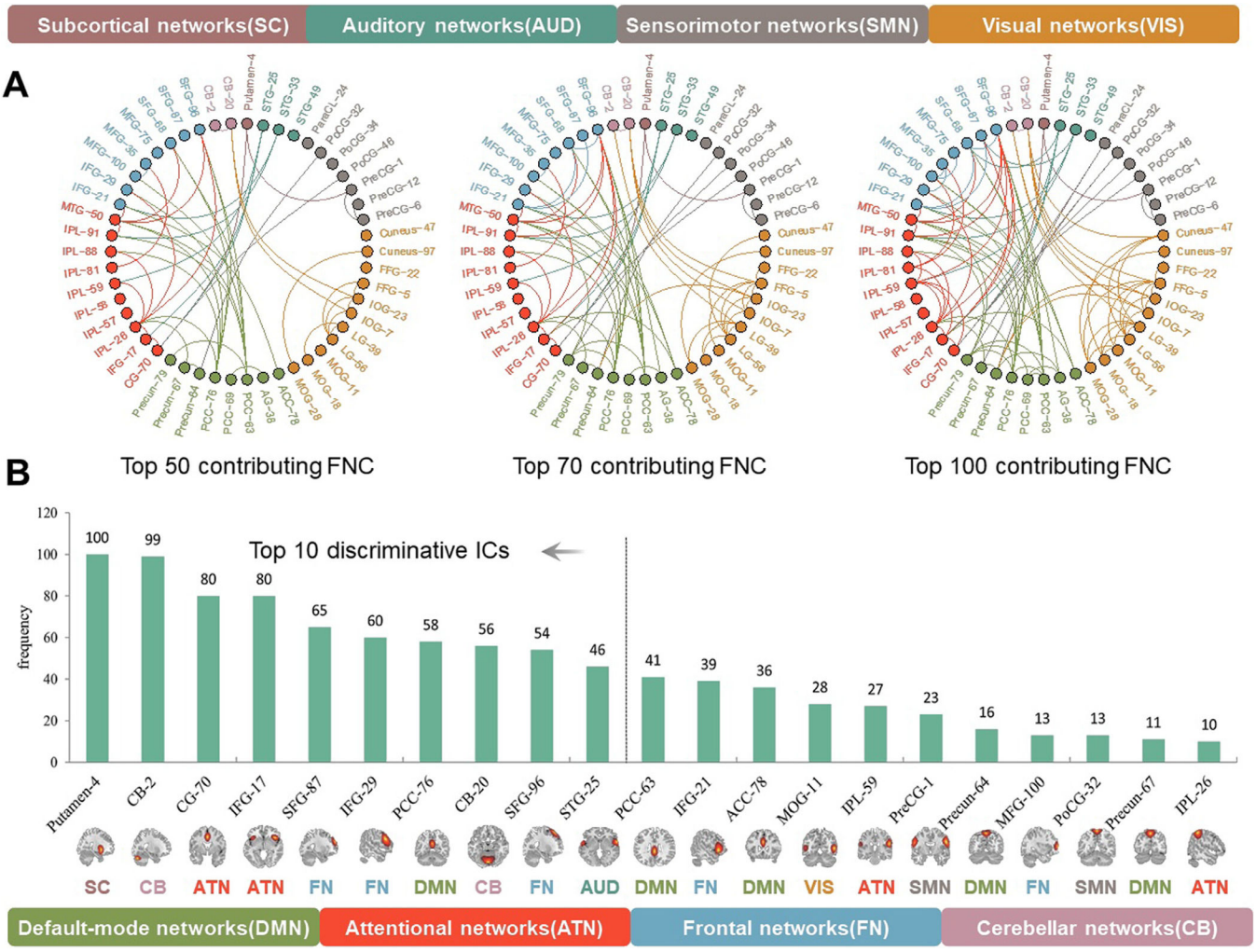

Fig. 3.

The most HC-SZ discriminative features localization. (A) Illustration of the top 50, 70 and 100 contributing functional network connectivities identified by LRP. Connections between frontal network and default mode networks, frontal network and attention networks, and connections within visual networks indicate the most contributing influence, suggesting that schizophrenia is characterized by impairment in high-level cognitive and emotional processing circuits. (B) The frequency distribution histogram of top 10 ICs identified by attention module in 100 experiments. The striatum, cerebellum, anterior cingulate stand out as the top three most discriminating brain regions. Putamen-4 represents the ICs showing subcortical regions such as caudate and putamen (striatum). The spatial maps of all 50 ICs were displayed in Figure S1.

Table 1

Demographic information of datasets.

Mean±SD	**SZ**	**HC**	***P*-value**
Number	558	542	NA
Age	27.6±7.1	28.0±7.2	0.06
Gender(M/F)	292/266	276/266	1.96
PANSS positive	23.9±4.2	NA	NA
PANSS negative	20.1±5.9	NA	NA
PANSS general	39.7±7.2	NA	NA
PANSS total	83.6±12.3	NA	NA

Notes: *P*-value: the significance value of two sample t-test. NA: not applicable.

Table 2

Performance comparison in multi-site pooling classification on In-House schizophrenia datasets.

Methods	Feature	ACC	SPE	SEN	F1	AUC
RF	FNC	76.4 ±0.8**	72.3±1.8**	80.4±0.5**	77.6± 0.5**	84.6±0.2**
AdaBoost	FNC	77.0 ±0.2**	75.6±0.2**	78.3±0.3**	77.6± 0.2**	81.8±0.3**
SVM	FNC	81.5 ±0.3**	80.0±0.8**	83.0±0.5**	82.6±0.2 **	88.4±0.2**
BrainNetCNN	FNC	80.1 ±0.8**	77.2±1.5**	82.9±1.2**	80.1±0.9 **	87.7±0.5**
DNN	FNC	80.9 ±0.4**	80.6±1.2**	81.3±0.7**	81.3±0.4 **	88.8±0.3**
C-RNN	TCs	82.5 ±0.9**	80.8±1.1**	84.2±0.9**	83.1±0.8 **	90.8±0.4**
C-RNN^{AM}	TCs	83.5 ±0.5**	81.5±0.9**	85.4±0.5**	84.0±0.5 **	91.4±0.3**
cGCN	FNC+TCs	78.3 ±0.6**	77.2±1.2**	78.6±1.1**	78.4±0.8 **	81.2±0.5**
HDLFCA	**FNC+TCs**	**85.3±0.4**	**83.4±0.6**	**87.1±0.5**	**85.8±0.3**	**92.4±0.2**

Notes: RF: random forest.

*/** denote that the proposed HDLFCA method achieves significantly better performance than the listed ones, with P value=0.05/0.01.

Table 3

Performance comparison in multi-site pooling classification on HC-ASD using ABIDE sites.

Methods	Feature	ACC	SPE	SEN	F1	AUC
RF	FNC	67.2±0.6 **	63.7±0.5 **	70.5±0.8 **	68.6±0.6 **	72.8±0.4
AdaBoost	FNC	64.2±0.1 **	62.0±0.1 **	66.2±0.2 **	65.3±0.1 **	66.7±0.2
SVM	FNC	69.5±0.1 **	66.4±0.2 **	72.4±0.2 **	70.7±0.2 **	76.6±0.2
BrainNetCNN	FNC	68.5±0.6 **	63.4±2.1 **	73.1±1.9 **	70.5±0.8 **	75.1±0.6
DNN	FNC	70.4±0.6 **	68.2±1.4 **	72.5±0.9 **	71.4±0.6 **	76.5±0.6
C-RNN	TCs	69.1±0.5 **	67.6±1.2 **	70.6±0.7 **	70.0±0.4 **	76.1±0.4
C-RNN^{AM}	TCs	69.4±0.5 **	67.1±0.8 **	71.5±0.7 **	70.4±0.5 **	76.0±0.6
cGCN	FNC+TCs	67.5±0.6 **	60.0±1.1 **	72.2±0.7 **	69.1±0.5 **	72.8±0.6
HDLFCA	**FNC+TCs**	**72.4±0.6**	**70.5±0.9**	**74.2±1.0**	**73.2±0.6**	**79.2±0.3**

Notes

*/** denote that the proposed HDLFCA method achieves significantly better performance than the listed ones, with P value = 0.05/0.01.

Table 4

Performance comparison in leave-one-site-out classification between HC and SZ.

Methods	Feature	ACC	SPE	SEN	F1	AUC
RF	FNC	71.4±4.4 **	65.1±12 **	79.1±8.1	73.5±4.0 *	82.1±3.5 **
AdaBoost	FNC	74.8±2.5 **	74.1±6.7 **	75.7±4.7	75.1±3.3 *	82.1±2.6 **
SVM	FNC	77.2±3.6 *	76.6±9.7 **	78.5±6.5	77.6±4.0	85.5±4.4 **
BrainNetCNN	FNC	75.8±3.8 *	77.5±9.5 **	75.8±6.3	76.5±3.2	85.1±4.2 **
DNN	FNC	76.8±3.1 *	76.2±9.0 **	77.8±5.7	77.8±3.7	85.0±4.0 **
C-RNN	TCs	77.6±1.9 *	77.9±8.1 **	77.1±9.3	77.3 ±4.2	86.5±2.4 **
C-RNN^{AM}	TCs	78.9±2.1	80.0±6.5 *	77.9±7.8	77.8±3.0	87.2±2.1 *
cGCN	FNC+TCs	75.1±3.2 *	76.5±9.0 **	74.4±5.6	75.5±3.2 *	83.1±4.1 **
HDLFCA	**FNC+TCs**	**81.5±2.2**	**87.5±6.0**	**75.1±5.8**	**80.3±1.7**	**90.2±2.4**

Note

*/** denote that the proposed HDLFCA method achieves significantly better performance than the listed ones, with P vaiue=0.05/0.01.

Table 5

Comparison with alternative classification methods using dynamic FNC on HC-SZ classification.

Methods	Feature	ACC	SPE	SEN	F1	AUC
GRU	TCs	76.9±0.5 **	74.4±1.0 **	79.3±0.7 **	77.8±0.5 **	84.3±0.3 **
LSTM	DFNC	80.5±0.5 **	81.5±1.2 **	79.6±1.0 **	80.6±0.5 **	88.8±0.3 **
BiLSTM	DFNC	80.2±0.5 **	81.1±2.0 **	79.4±1.6 **	80.3±0.5 **	88.7±0.4 **
GRU	DFNC	80.6±0.9 **	80.5±1.1 **	81.1±2.3 **	81.1 ± 1.2 **	88.7±0.6 **
C-LSTM	DFNC	79.6±0.7 **	80.2±2.0 **	78.9±1.2 **	79.7±0.6 **	88.0±0.4 **
HDLFCA	**FNC+TCs**	**85.1 ±0.4**	**82.8±0.8**	**87.3±0.8**	**85.6±0.3**	**92.1 ±0.2**

Notes: LSTM: Long short-term memory network; BiLSTM: Bidirectional LSTM; GRU: gated recurrent unit; C-LSTM: CNN+LSTM

*/** denote that the proposed HDLFCA method achieves significantly better performance than the listed ones with p=0.05/0.01.

Table 6

Performance comparison of different DL architectures on SZ classification based on multi-site pooling

Methods	Feature	ACC	SPE	SEN	F1	AUC
S_RNN	TCs	53.3 ±0.9 **	43.7±1.1 **	62.5±0.9 **	57.7±0.8 **	53.8±0.4 **
GRU	TCs	76.9 ±0.5 **	74.4±1.0 **	79.3±0.7 **	77.8±0.5 **	84.3 ±0.3 **
C-MLP	TCs	77.1 ± 0.4 **	75.7±0.8 **	78.4±0.7 **	77.7±0.4 **	86.7±0.3 **
S_C-RNN	TCs	80.5 ±0.5 **	79.4±1.0 **	81.4±0.9 **	80.9±0.5 **	88.5±0.4 **
C-RNN	TCs	82.5 ±0.9 *	80.8±1.1	84.2±0.9 *	83.1±0.8 *	90.8±0.4
AM_1	TCs	83.4 ±0.5	81.6±1.0	85.1±0.7	83.9±0.5	91.0±0.3
AM_2	TCs	83.4 ±0.4	81.6±0.8	85.2±1.1	83.9±0.4	91.3 ±0.3
AM_3	TCs	54.8 ±0.6 **	54.4±0.6 **	55.3±1.2 **	55.5±0.8 **	57.3±0.4 **
C-RNN™	**TCs**	**83.5±0.5**	**81.5±0.9**	**85.4±0.5**	**84.0±0.5**	**91.4±0.3**

Notes

*/** denote that the proposed HDLFCA method achieves significantly better performance with P value=0.05/0.01. S_RNN: simple RNN. C-MLP: the convolutional layer (CON) has different kernel size as C-RNN and the fully connected layers was followed. S_C-RNN: the CON has fixed kernel size and the other architecture was the same as C-RNN. AM_1: the CON in AM was one kernel with 4* 1 size. AM_2: the CON in AM was replaced by the shared MLP, including three fully connected layers with 50, 10 and 50 hidden nodes respectively. AM_3: a spatial-temporal attention module based on the proposed attention module (AM) in this work to emphasize important time points and regions simultaneously.