EBioMedicine 47 (2019) 543–552

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site FMRI data

Weizheng Yan [a,b], Vince Calhoun [c], Ming Song [a,b], Yue Cui [a,b], Hao Yan [d,e], Shengfeng Liu [a,b], Lingzhong Fan [a,b], Nianming Zuo [a,b], Zhengyi Yang [a,b], Kaibin Xu [a,b], Jun Yan [d,e], Luxian Lv [f,g], Jun Chen [h], Yunchun Chen [i], Hua Guo [j], Peng Li [d,e], Lin Lu [d,e], Ping Wan [j], Huaning Wang [i], Huiling Wang [h], Yongfeng Yang [f,g,k], Hongxing Zhang [f,l], Dai Zhang [d,e,m], Tianzi Jiang [a,b,k,n,o,⁎], Jing Sui [a,b,o,⁎]

[a] *National Laboratory of Pattern Recognition and Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[b] *University of Chinese Academy of Sciences, Beijing 100049, China*
[c] *Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS) Center, Atlanta 30303, GA, USA*
[d] *Peking University Sixth Hospital, Institute of Mental Health, Beijing 100191, China*
[e] *Key Laboratory of Mental Health, Ministry of Health, Peking University, Beijing 100191, China*
[f] *Department of Psychiatry, Henan Mental Hospital, The Second Affiliated Hospital of Xinxiang Medical University, Xinxiang 453002, Henan, China*
[g] *Henan Key Lab of Biological Psychiatry, Xinxiang Medical University, Xinxiang 453002, Henan, China*
[h] *Department of Radiology, Renmin Hospital of Wuhan University, Wuhan 430060, Hubei, China*
[i] *Department of Psychiatry, Xijing Hospital, The Fourth Military Medical University, Xi'an 710032, Shaanxi, China*
[j] *Zhumadian Psychiatric Hospital, Zhumadian 463000, Henan, China*
[k] *Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, Sichuan, China*
[l] *Department of Psychology, Xinxiang Medical University, Xinxiang 453002, Henan, China*
[m] *Center for Life Sciences/PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China*
[n] *Queensland Brain Institute, University of Queensland, Brisbane 4072, QLD, Australia*
[o] *CAS Center for Excellence in Brain Science and Intelligence Technology, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*

ARTICLE INFO

Article history:
Received 1 March 2019
Received in revised form 9 August 2019
Accepted 9 August 2019
Available online 13 August 2019

Keywords:
Recurrent neural network (RNN)
Schizophrenia
Multi-site classification
fMRI
Striatum
Cerebellum
Deep learning

ABSTRACT

Background: Current fMRI-based classification approaches mostly use functional connectivity or spatial maps as input, instead of exploring the dynamic time courses directly, which does not leverage the full temporal information.
Methods: Motivated by the ability of recurrent neural networks (RNN) in capturing dynamic information of time sequences, we propose a multi-scale RNN model, which enables classification between 558 schizophrenia and 542 healthy controls by using time courses of fMRI independent components (ICs) directly. To increase interpretability, we also propose a leave-one-IC-out looping strategy for estimating the top contributing ICs.
Findings: Accuracies of 83·2% and 80·2% were obtained respectively for the multi-site pooling and leave-one-site-out transfer classification. Subsequently, dorsal striatum and cerebellum components contribute the top two group-discriminative time courses, which is true even when adopting different brain atlases to extract time series.
Interpretation: This is the first attempt to apply a multi-scale RNN model directly on fMRI time courses for classification of mental disorders, and shows the potential for multi-scale RNN-based neuroimaging classifications.
Fund: Natural Science Foundation of China, the Strategic Priority Research Program of the Chinese Academy of Sciences, National Institutes of Health Grants, National Science Foundation.

1. Introduction

Functional magnetic resonance imaging (fMRI), as a non-invasive imaging technique, has been extensively applied to study psychiatric disorders [1]. Due to the high-dimensional and low signal-to-noise ratio properties of the fMRI data, efficient feature selection procedures are usually required to reduce the redundancy before modeling. Two types of approaches, data-driven [2] and seed-based [3], have been extensively applied to decompose 4D fMRI data, resulting in spatial brain regions/independent components (ICs) and their corresponding time courses (TCs). Currently, existing fMRI-based classification models mostly adopt either subject-specific spatial maps [4] or functional

⁎ Corresponding authors at: Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China.
E-mail addresses: jiangtz@nlpr.ia.ac.cn (T. Jiang), jing.sui@nlpr.ia.ac.cn (J. Sui).

https://doi.org/10.1016/j.ebiom.2019.08.023

Research in context

Evidence before this study

Current fMRI-based classification approaches mostly use functional connectivity or spatial maps as input, instead of exploring the dynamic time courses directly, which does not leverage the full temporal information. In addition, the excellent feature-representation ability of deep learning methods provides us a way to capture spatiotemporal information from time courses.

Added value of this study

In the present study, we contributed a new deep learning-based framework which can directly work on fMRI time courses for identifying brain disorders. In addition, by using our proposed deep learning-interpretation method, dorsal striatum and cerebellum are discovered as the top two discriminative brain regions.

Implications of all the available evidence

To the best of our knowledge, this is the first attempt to enable deep learning directly to work on time courses of fMRI components in schizophrenia classification, which promise great potentials of deep-chronnectome-learning and a broad utility on neuroimaging applications, *e.g.*, the extension to MEG, EEG learning.

(network) connectivity calculated by TC correlations as input features [5,6], though have achieved substantial progress, the sequential temporal dynamics were generally missed. The field is still striving to understand how to diagnose and discriminate complex mental illness, *e.g.,* schizophrenia *versus* bipolar disorder, while ignoring the temporal information time-point by time-point is likely missing a critical, but available, part of the puzzle.

The power of deep learning models lies in enabling automatic discovery of latent or abstract higher-level information from high-dimensional neuroimaging data, which can be an important step to understand complex mental disorders [7–14]. Specifically, convolutional neural network (CNN) which is “deep in space” and recurrent neural network (RNN) which is “deep in time” are two classic deep learning branches. It is natural to use CNN as an ‘encoder’ for obtaining correlations between brain regions and simultaneously employ RNN for sequence classification. RNN models such as long short-term memory (LSTM) [15] and gated recurrent unit (GRU) [16] have been firmly established as state-of-the-art approaches in sequence modeling, such as identifying autism using fMRI [17], diagnosing brain disorder by analyzing electroencephalograms [18], detecting temporally dynamic functional state translations [13,14,19].

In particular, GRU is a particular RNN-based model which can effectively solve the long-term dependency problem by controlling information flow with several gates, which may fit the fMRI brain voxel-wise changes along with time series. Moreover, multi-scale convolution layers can be complementary for CNN feature extraction, because it can account for different temporal scales (from seconds to minutes) of brain activity. Therefore, we combine the strengths of CNN and RNN models and develop a *Multi-scale RNN (MsRNN)* model, which can directly work on fMRI time courses for classifying brain disorders, thus avoids the second-level calculation (*e.g.,* correlation analysis) of time courses and takes advantage of the high-level spatiotemporal information of fMRI data. Such a design of classification framework relies on two assumptions: *1)* underlying dynamics of fMRI data, *i.e.*, rules by which neural activities involved in time; *2)* brain disorders may have different patterns of temporal changes recorded by fMRI.

In this work, based on a large-scale Chinese Han resting-state fMRI data consisting of 558 schizophrenia patients (SZ) and 542 healthy controls (HC) that were recruited from seven sites with compatible MRI scanning parameters and imaging quality, we tested the power of the proposed MsRNN model for deep chronnectome learning on multiple facets, with comparison of three classic classification algorithms and eight varietal deep-learning models. Furthermore, to improve the result's interpretability, which is the most challenging issue of deep learning in neuroimaging applications, we propose a leave-one-IC-out strategy for estimating the contribution of each IC on classifying schizophrenia. Subsequently, components of dorsal striatum and cerebellum contributed the top two group-discriminating time courses. Finally, the time courses extracted by using seed-based strategies, *e.g.*, using brain atlases such as AAL [20] or Brainnetome Atlas [21], were compared further with ICA results. To the best of our knowledge, this is the first attempt to enable CNN + RNN directly to work on time courses of fMRI components in mental disorder classification, which promise great potentials of deep-chronnectome-learning and a broad utility on neuroimaging applications, *e.g.*, the extension to MEG, EEG learning.

2. Materials and methods

Fig. 1 presents an overview framework of the *MsRNN* classification method. Resting-state fMRI data from 1100 Chinese subjects (558 SZs, 542 HCs, from 7 sites) were used, which were preprocessed using the standard procedure [6]. Details of the demographic information are shown in Table S1. Time courses were extracted using group ICA [2]. Each subject was then represented with the TC features (*No*. time points × *No*. ICs, Fig. 1a, c). The proposed *MsRNN* model was directly applied on TCs of the selected non-artificial ICs to identify SZs from HCs using two types of classification strategies (Fig. 1b): *1*) Multi-site pooling classification, in which all 1100 subjects from seven sites were pooled together, which were split into training set, validation set and testing set. Moreover, the classification performance was measured using k-fold cross-validation strategy; *2*) Leave-one-site-out transfer classification, in which the subjects of a given site were left for testing, and the samples of all other sites were used for training and validation. These two types of classification strategies were independent of each other [9]. We trained the *MsRNN* using the TCs in training and validation sets with their corresponding labels (Fig. 1c). The learnable parameters of the *MsRNN* were iteratively adjusted using the error backpropagation algorithm. The validation samples were simultaneously used for monitoring the training process and avoid overfitting. The performance of the trained *MsRNN* was finally tested using held out TCs.

2.1. Participants and demographics

Table S1 lists the demographic and clinical information of all 1100 participants (558 SZs and 542 age and gender-matched HCs) in this study. The subjects were within the 18–45 age range, right-handed who were screened for ethical clearance, with only Chinese Han people recruited from seven sites in China with the same recruitment criterion, including Peking University Sixth Hospital (Site 1); Beijing Huilongguan Hospital (Site 2); Xinxiang Hospital Simens (Site 3); Xinxiang Hospital GE (Site 4); Xijing Hospital (Site 5); Renmin Hospital of Wuhan University (Site 6); Zhumadian Psychiatric Hospital (Site 7). Each site received approval from their respective research ethics boards and written informed consents were obtained from all study participants. All the SZ patients were evaluated based on the Structured Clinical Interview for DSM disorders (SCID) and diagnosed by experienced psychiatrists according to the criteria of DSM-IV-TR. All the HCs were recruited from the same local geographical areas as the patients cohort through local advertisement and were free of Axis I or II disorders (SCID-Nonpatient) Additional exclusion criteria include factors such as current or past

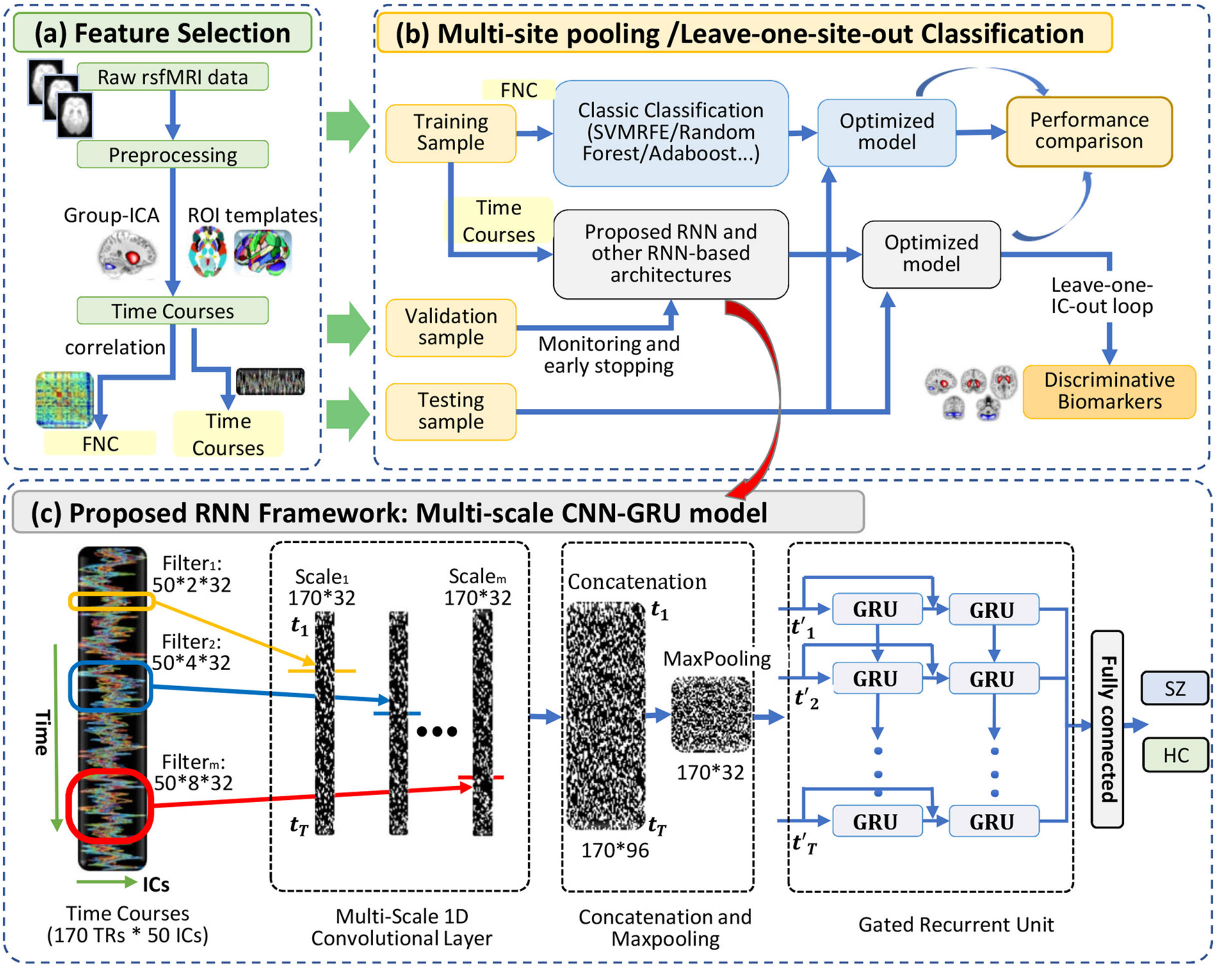

Fig. 1. The framework of the Multi-scale RNN model in distinguishing schizophrenia patients from healthy controls. (a) Data preprocessing and feature selection. All rsfMRI data were preprocessed using the standard procedure. Time courses were then extracted using group-ICA/AAL/Brainnetome Atlas respectively. (b) The TCs/FNC data were randomly split into training, validation and testing sets. In multi-site pooling classification, all seven datasets were pooled together, and then k-fold cross-validation strategies were used for evaluating classification performance. In leave-one-site-out transfer prediction, the samples of a given imaging site were left for testing, and the samples of other sites were used for training. The performance of conventional methods (including Adaboost, Random Forest and SVM) and various RNN-based models were used for comparison. The most discriminative components were found by using leave-one-IC-out method. (c) Details of the MsRNN classification model. Three different scales convolutional filters were used for extracting of spatial features from time courses. The extracted features were then concatenated, pooling, and sent to stacked GRU module.

neurological illness, substance abuse or dependence, pregnancy, and prior electroconvulsive therapy or head injury resulting in loss of consciousness.

2.2. Image acquisition

The resting state fMRI data were collected with the following three different types of scanners: 3·0 T Siemens Trio Tim Scanner (Siemens; Site 1, 2 & 5), 3·0 T Siemens Verio Scanner (Siemens; Site 3), and 3·0 T Signa HDx GE Scanner (General Electric; Site 4, 6 & 7). To ensure equivalent, coincident and high-quality data acquisition, the scanning protocols for all the seven sites were set up by the same experienced experts [6]. Subjects were instructed to relax and lie still in the scanner while remaining calm and awake. More details of scanning parameters are listed in Supplementary Table S4.

2.3. Data preprocessing and IC extraction

The rsfMRI data were preprocessed according to the procedures which were the same as we did in [6] using SPM8 software (http://www.fil.ion.ucl.ac.uk/spm/). For each participant, the first ten volumes of each scan time series were discarded to ensure magnetization equilibrium. The remaining resting state volumes were first corrected by the acquisition time delay of different slices and then realigned to the first volume for head-motion correction [22]. For each subject, the translation of head motion was <3 mm and the rotation of head motion did not exceed 3° in all axes through the whole scanning process. Subsequently, the images were spatially normalized to EPI template conforming to the Montreal Neurological Institute (MNI) space. The data (originally collected at 3·44 mm × 3·44 mm × 4·60 mm) were then resliced to a voxel size of 3 mm × 3 mm × 3 mm, resulting in 53 × 63 × 46 voxels for each image. Subsequently, group ICA toolbox (GIFT, http://mialab.mrn.org/software/gift) was used to perform GIG-ICA [23] on the preprocessed fMRI data. 50 ICs were characterized as intrinsic connectivity networks (ICNs) after removing those ICs corresponding to physiological, movement-related or imaging artifacts, and their spatial maps (SMs) are listed in the Supplementary file Fig. S3. According to previous work [24,25], the control of movement-related artifacts should be stringent for the analysis of time courses of fMRI data. We compared the mean of framewise displacement (FD) for HC and SZ groups. The mean FD for HC and SZ are 0·137±0·071 and 0·142±0·085 respectively, with no significant group differences ($P=.98$, two-sample t-tests) existing. In our preprocessing, as did in previous work, nuisance covariates including six head motion parameters, mean FD, white matter signal, cerebrospinal fluid signal, and global mean signal were all regressed out [24,26,27]. Two covariants (age and gender) which may have potential confounding effects were also regressed out. Then the time courses were stacked to form a matrix with dimensions of [No. Subjects] × [No. Time courses] × [No. Independent components or ROIs)] which was then used to calculate the FNC matrix or to train the *MsRNN* model directly.

2.4. Multi-scale CNN-GRU (MsRNN)

As shown in Fig. 1c, *MsRNN* consists of 3 different scales of 1D convolutional filters (2TR, 4TR and 8TR, TR = 2 s), one concatenation layer, one max-pooling layer, a two-layer stacked gated recurrent unit (GRU) which are densely connected in a feed-forward manner, and an averaged layer which integrate the whole sequence. The time courses were fed into the proposed *MsRNN* model for parameter optimization. After optimizing the parameters, the model was saved for testing and comparison. Equations are listed in the Supplementary files for a precise definition of the *MsRNN* model.

2.4.1. Multi-scale convolutional layer

Multi-scale convolution layers may be helpful in feature extraction because it can account for different scales (from seconds to minutes) of brain activity. Inspired by 1D convolution (Conv1D) layers [28], we designed an architecture which expands upon simple convolutional layers by including multiple filters of varying sizes in each Conv1D layer. This architecture allows the network to extract information over multiple time scales. The filter lengths used in the Conv1D were drawn from a logarithmic instead of a linear scale, leading to exponentially varying filter lengths (2TR, 4TR, and 8TR). Therefore, the size of 3 different scales of convolutional filters are 50 (ICs) × 2 × 32 (number of filters), 50 × 4 × 32, 50 × 8 × 32 in our experiment. A concatenation layer then concatenates the incoming features among the depth axis, resulting in feature maps whose size are 170 (time points) ×96 (feature dimension). Whereafter, a max-pooling layer performs downsampling operation along the time dimensions with filter size 3, resulting in feature such as 57(time points) × 96(feature dimension).The downsampled features are as the input of the following GRU layers.

2.4.2. Densely connected GRU layer

A two-layer stacked GRU may capture higher-level dynamic information than single-layer GRU model. The size of the GRU's hidden state was set as 32. However, one of the central challenges of training a deep GRU-based network the gradient exploding/vanishing problem. It is worthy to note that the densely-connected structure may effectively prohibit the “gradient exploding/vanishing” problem by connecting each layer to every other layer in a feed-forward manner [29].

2.4.3. Averaged layer

Even with the best experimental fMRI design, it is infeasible to control the random thoughts of the subjects during the resting-state fMRI scanning because they depend on too many subject-specific factors. Also, it is not possible to label the beginning and the end of brain activities. Hence combining all fMRI steps by averaging all of the GRU outputs is a compromised solution [10]. In this way, all activities of the brain during scanning may be leveraged for obtaining better classification performance.

In summary, the proposed *MsRNN* classification model consists of multiple-scale Conv1D layers, stacked GRU layers which are densely connected in a feed-forward manner, an averaged layer which integrates the context of the whole sequence, and fully-connected layers. More details of the model can be found in Supplementary Fig. S2.

2.5. MsRNN model implementation

The time courses of ICs described above were used as the inputs for training the Multi-scale RNN model. The model was trained by minimizing the cross-entropy loss using *Adam* optimizer. The training batch size was set as 64. The learning rate started from 0.001 and decayed after each epoch with the decay rate of $10^{-2}10^{-2}$. To improve the generalization performance of the model and overcome overfitting, dropout(dropout = 0.5) and $L_{1,2}$-norm regularization (L1 = 0.0005, L2 = 0.0005) were also applied for regulating the model parameters. The training process was stopped when the validation loss stopped decreasing for 50 epochs or when the maximum epochs (1000 epochs) had been executed. In our experiment, the training time for *MsRNN* was around five minutes, while the testing time for a new subject is <0.01 s. The intermediate model which achieved the highest accuracy on the validation dataset was reserved for testing. Also, the proposed models were implemented on the platform of Keras (https://keras.io/) and ScikitLearn (https://scikit-learn.org/).

The visualization of *MsRNN* codes was performed by the unsupervised dimensionality reduction technique t-SNE, which embeds high-dimensional data into a low-dimensional space while preserving the pairwise distances of the data points, implemented in MATLAB. The activation strengths of individual neurons at the last hidden layer by the training and testing samples were used as the raw variables. The parameters for the stochastic optimization for t-SNE [30] were as follows [31]: The perplexity was 30, and the dimension for initial principal components analysis was 30.

2.6. Estimating the discriminative power of independent components (leave-one-IC-out)

The basic idea is that the feature whose elimination lead to the most significant damage of classification performance should be regarded as the top contributing features. More specifically, as shown in Fig. 3b, each subject is represented with a $T \times D$ matrix, where T is the length of time courses and D is the number of independent components (ICs). A specific element in the matrix can be denoted by v_{td}. To quantify the classification contribution of the d_{th} IC, we replace the time courses of d_{th} IC with its averaged value $\frac{1}{T}\sum_{t=1}^{T} v_{td}$ while keep other ICs' time courses as they were. This is equivalent to eliminating the contribution of d_{th} component. All the testing samples are processed in the same way and subsequently fed to the trained *MsRNN* model. The classification performance of the trained model which is fed with reduced features may decrease compared to that using all features. The variation of the classification performance (*i.e.*, accuracy, sensitivity, specificity) when removing d_{th} dimension are recorded and sorted. The features which maximize the decrease of the classification performance are further selected as the most discriminative features. Specifically, the 1100 samples were randomly split into five folds. 880 samples (four folds) were used for optimizing the parameters of *MsRNN*, and 220 samples (one-fold) were used for further finding the contribution of each IC during each cross-validation. The specific procedures are as follow: *1)* After optimizing the trained model with 880 samples, the parameters of the trained model were saved; *2)* The time courses of 220 subjects without removing any component were fed to the model to obtain a baseline classification performance; *3)* The 220 subjects which have removed the contribution of one specific IC were fed to the model to obtain the classification performance repeatedly. The decrease of sensitivity/specificity when removing a specific component was recorded and sorted; *4)* Repeat step 3 until each IC has been removed once.

2.7. Statistics

The performance of identifying schizophrenia from normal controls was evaluated by five metrics including accuracy (ACC), sensitivity (SEN), specificity (SPE), F-score (F1) and area under curve (AUC) based on the results of cross-validation (k-fold or leave-one-site-out). They are defined as below:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, SEN = \frac{TP}{TP + FN}, SPE = \frac{TN}{TN + FP},$$

$$PPV = \frac{TP}{TP + FP}, F1 = 2\frac{SEN \times PPV}{SEN + PPV}$$

where TP, TN, FP, FN, PPV denote true positive, true negative, false positive, false negative and positive predictive value respectively, *e.g.*, SEN represents the percentage of SZ are classified as SZ correctly. The full k-fold cross-validation procedure was repeated ten times to generate the means and standard deviations of accuracy, sensitivity, and specificity. We used two-sample *t*-test to compare classification performances between different algorithms and hyperparameter settings.

2.8. Data availability

All data needed to evaluate the conclusions are present in the paper and/or the supplementary materials. Additional data related to this paper may be requested from the authors.

3. Results

3.1. Multi-site pooling classification

We compared the *MsRNN* with three traditional popular classifiers (*SVM* [32], *Adaboost* [33], *Random Forest* [34]), one multi-layer perception model, and seven RNN-based alternative deep learning models (Table 1). The detailed hyperparameters and the time complexity of these methods can be found in Supplementary Table S5. All the above models were implemented on a desktop computer (Intel(R) Xeon (R) CPU E5–1650 v4 @ 3.60GHz, 6 CPU cores) with a single GPU (12GB NVIDIA GTX TITAN 12GB), and can be trained within five minutes. Note that the three conventional classification methods usually work on the FNC matrix that was computed using the correlation of TCs of selected components instead of the TCs themselves. Therefore, in performance comparison, FNCs were used as the input of conventional methods while TCs were used as the input of *MsRNN,* multi-layer perception, and other RNN-based deep learning methods. All models were trained using the training dataset and tested using testing dataset, embedded in nested five-fold cross-validation cycles. Fig. 1c shows the architecture of the proposed *MsRNN* model.

Table 1 and Fig. 2a listed the averaged accuracy and variance of classification performance achieved by all 11 methods in multi-site pooling condition. In the deep learning classification frameworks (including *MsRNN,* multi-layer perception, and other RNN-based architectures), we used four folds as the training set (10% samples of the training set were further randomly selected as validation dataset), and one-fold as the testing dataset. As for conventional classification models (*Adaboost, Random Forest* and *SVM*), four folds were used for training and one-fold for testing.

The accuracy of 83·2 ± 3·2% was obtained by using the *MsRNN* method, which is significantly higher than those obtained by using the *Adaboost, Random Forest* and *SVM* ($P = 2 \cdot 1e\text{-}4$, $1 \cdot 9e\text{-}4$, 1.1e-2, two-sample *t*-tests, df = 18). Also, the ROC curves of these methods are shown in Fig. 2b. The proposed *MsRNN* achieved an AUC of 0.906, while the AUC of *Adaboost*, *Random Forest* and *SVM* ranges from 0·840–0·868. To validate the advantage of the proposed model, other RNN architectures based on GRU and one similar network architecture based on LSTM were also compared with *MsRNN*. As shown in Table 1, a single layer GRU model can easily reach a higher classification performance than the classic FNC-based methods. The improvement may be due to the ability of GRU in extracting dynamic information from time sequences. In addition, the performance of *GRU_1_ave* is better than *GRU_1_last* because the former one made full use of temporal information at every time point. Furthermore, combining the GRU layer with Conv1D layer is a remedy for improving the classification performance because CNN-GRU model is "double deep" which include both spatial and temporal layers. Thus it can be jointly trained to learn convolutional perceptual representations and temporal dynamics simultaneously.

Finally, the proposed multi-scale convolution is even better than a single-scale convolution layer because it can extract dynamics from a variety of timescales. In summary, multi-site pooling results indicated that our proposed *MsRNN* model achieved the best performance by smartly integrating the advantages of CNN and RNN, while the LSTM-based model can reach competitive performance compared with the GRU-based model.

3.2. Leave-one-site-out transfer classification

In the leave-one-site-out classification, we left each of the seven sites as the testing data and used the other six sites for training and validation, in which 10% samples were randomly selected as validation dataset and the other 90% were used for training *MsRNN* or other deep learning architectures. In the *Adaboost, Random Forest* and *SVM* classification frameworks, we used the samples of the given imaging site for testing and the samples of other sites for training. The leave-one-site-out transfer classification results are shown in Table 2 and Table S2. The averaged classification performance of the seven sites was used to represent the overall performance of cross-site prediction. The accuracy

Table 1
Performance comparison in multi-site pooling classification.

Methods		ACC	SEN	SPE	F1	AUC
CON	Adaboost	75.6(3.8)**	77.0(4.4)**	74.2(4.4)**	76.2(3.8)**	84.2(3.6)**
CON	Random Forest	76.0(3.5)**	81.0(3.9)o	71.4(5.5)**	77.4(3.5)**	84.0(3.4)**
CON	SVM	79.4(3.1)*	80.4(3.5)o	78.4(3.9)*	79.6(3.3)*	86.8(3.2)*
RNN	GRU_1_last	51.6(3.6)**	52.0(5.3)**	51.2(4.3)**	52.0(3.8)**	51.2(3.6)**
RNN	GRU_1_ave	77.8(3.4)**	78.4(3.8)**	77.0(3.5)**	78.2(3.4)**	86.8(3.5)*
RNN	GRU_2_ave	78.0(3.9)**	80.8(5.1)o	76.0(4.2)**	78.8(3.9)*	86.8(4.1)*
CMLP	Multi_CNN_MLP	77.8(3.4)**	76.2(4.0)**	79.2(4.8)○	77.2(3.4)**	86.4(3.1)**
CRNN	Simple_CNN_GRU_2_ave	80.8(3.0)○	80.2(4.3)○	82.0(3.5)○	80.8(3.1)○	89.2(2.8)○
CRNN	Multi_CNN_GRU_1_ave	80.6(3.5)○	80.8(4.1)○	80.6(4.3)○	80.8(3.3)○	88.2(3.6)○
CRNN	Multi_CNN_GRU_2_ave	81.2(3.4)○	81.4(4.1)○	81.0(4.9)○	81.0(3.5)○	88.6(3.7)○
CRNN	Multi_CNN_LSTM_2_ave	81.6(2.9)○	82.6(3.6)○	80.4(3.8)○	82.0(2.7)○	89.4(2.8)○
CRNN	MsRNN(Proposed)	83.2(3.2)	83.1(3.7)	83.5(3.7)	83.3(3.2)	90.6(3.0)

CON: conventional classification methods; RNN: RNN-based methods; CMLP: CNN linked with multi-layer perception; CRNN: CNN-RNN based methods; SVM: Support vector machine with Gaussian kernel; LSTM: Long short-term memory network; GRU: gated recurrent unit. GRU_1: one layer of GRU; GRU_2: two-layer stacked GRU; #_last: the output of the last GRU step is connected to the next layer. #_ave: the average of the outputs of all GRU steps is connected to the next layer; SimpleCNN: Convolutional layer has fixed kernel size; Multi_CNN: Convolutional layer has different kernel size; ○ denotes that the methods have no significant difference (two-sample *t*-test) with the proposed. */** denote respectively that the methods are significantly worse than the proposed model with *P* value = .05/0.01. Details of all these mentioned architectures are shown in Supplementary file Fig. S2. The last row is our proposed method.

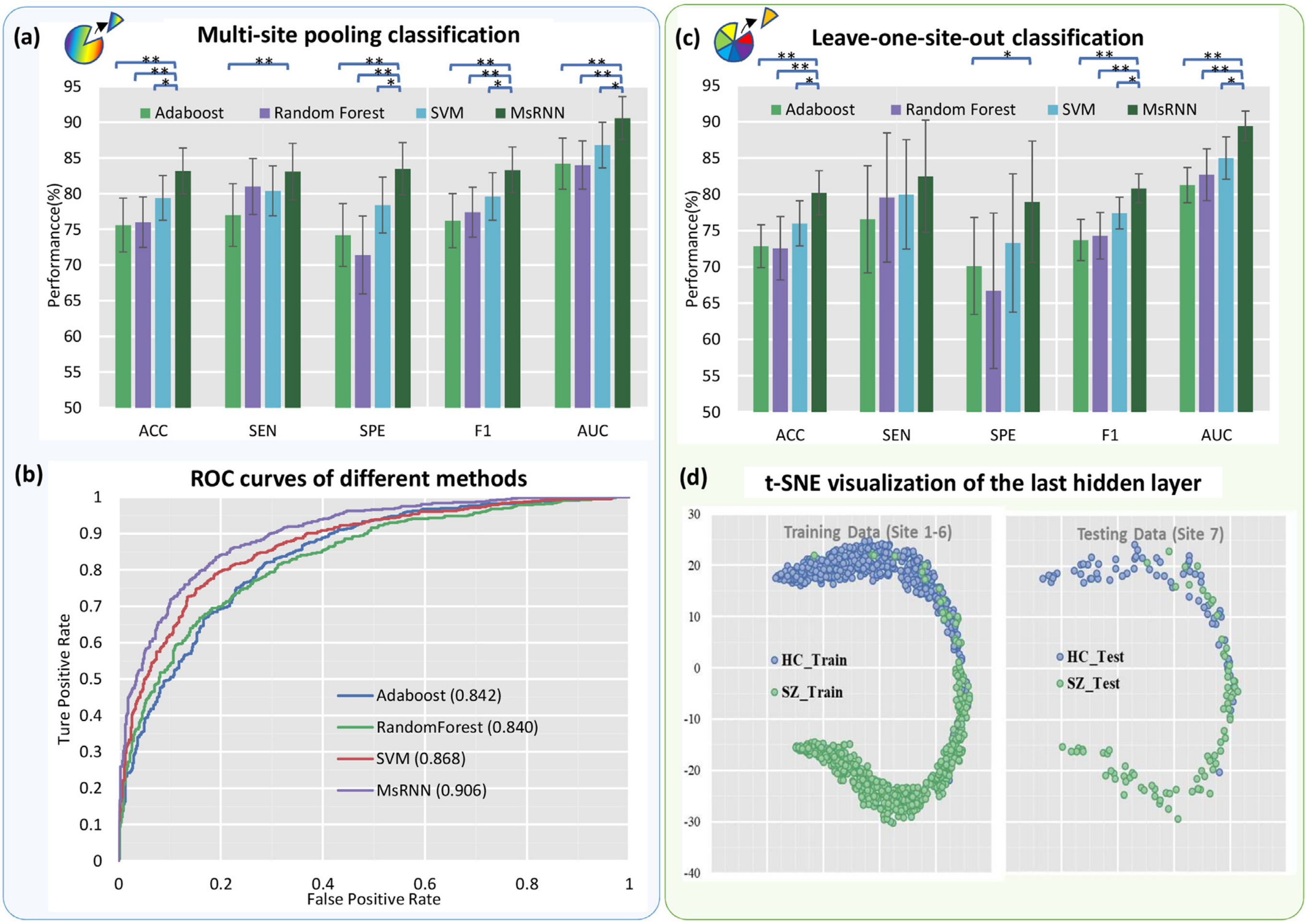

Fig. 2. Classification results of multi-site pooling and leave-one-site-out transfer classification. (a) 5-fold multi-site pooling classification results. $^{**}P < .01$(two-sample *t*-test), $^{*}P < .05$ (two-sample *t*-test). (b) The comparison of receiver operating characteristic curves of different methods. (c) Leave-one-site-out transfer classification results. (d) t-SNE visualization of the last hidden layer representation in the MsRNN for SZ/HC classification. Here we show the MsRNN's internal representation of SZ and HC by applying t-SNE, a method for visualizing high-dimensional data, to the last hidden layer in the MsRNN of training (Site 1–6: 951 subjects) and testing (Site 7: 149 subjects) samples.

of 80·2% was achieved by using the *MsRNN* method, which was significantly higher than the accuracies obtained by using the *Adaboost*, *Random Forest* and *SVM* ($P = 6{\cdot}2\text{e-}4$, 7.0e-3, 1.9e-2, two-sample *t*-test, df = 12) (Fig. 2c). To visualize the performance of *MsRNN* classifier, we used *t*-Distributed Stochastic Neighbor Embedding (*t*-SNE) to project the 32-dimensional representations of subjects extracted from the hidden layer of the trained *MsRNN* model to a 2D plane. As shown in Fig. 2d, samples from six sites (951 subjects, site 1–6) were used as the training/validation set, and the samples from site 7 (149 subjects) were used for testing. The tSNE result indicates that the proposed *MsRNN* model can successfully distill features and separate the SZ and HC apart.

Table 2
Performance comparison in leave-one-site-out classification.

Methods		ACC	SEN	SPE	F1	AUC
CON	Adaboost	72.9(3.0)**	76.6(7.4)○	70.1(6.7)*	73.7(2.8)**	81.3(2.4)**
CON	Random Forest	72.6(4.4)**	79.6(8.9)○	66.7(10.7○	74.3(3.2)**	82.7(3.6)**
CON	SVM	76.0(3.1)*	80.0(7.5)○	73.3(9.5)○	77.4(2.2)*	85.0(2.9)*
RNN	GRU_1_last	47.7(3.2)**	50.6(6.8)**	44.7(7.1)**	49.3(3.7)**	46.7(2.4)**
RNN	GRU_1_ave	78.7(2.8)○	80.9(7.3)○	77.4(7.4)○	79.4(1.9)○	86.9(2.3)*
RNN	GRU_2_ave	77.9(3.9)○	79.0(9.2)○	77.9(7.5)○	78.1(2.7)○	87.7(3.0)○
CMLP	Multi_CNN_MLP	76.1(3.2)*	79.7(8.2)○	73.4(9.8)○	77.0(2.1)**	85.4(2.7)*
CRNN	Simple_CNN_GRU_2_ave	79.1(3.7)○	82.4(7.9)○	76.7(10.7)○	80.1(2.3)○	89.1(2.3)○
CRNN	Multi_CNN_GRU_1_ave	80.3(3.0)○	82.9(7.3)○	79.0(9.4)○	81.1(1.8)○	88.7(2.3)○
CRNN	Multi_CNN_GRU_2_ave	79.7(3.0)○	80.4(7.2)○	79.6(7.7)○	79.9(2.7)○	88.6(2.3)○
CRNN	Multi_CNN_LSTM_2_ave	78.7(3.9)○	83.1(8.3)○	75.3(9.7)○	79.7(2.6)○	89.6(3.0)○
CRNN	MsRNN(Proposed)	80.2(3.0)	82.5(7.7)	79.0(8.4)	80.8(2.0)	89.4(2.1)

CON: conventional classification methods; RNN: RNN-based methods; CMLP: CNN linked with multi-layer perception; CRNN: CNN-RNN based methods; SVM: Support vector machine with Gaussian kernel; LSTM: Long short-term memory network; GRU: gated recurrent unit. GRU_1: one layer of GRU; GRU_2: two-layer stacked GRU; #_last: the output of the last GRU step is connected to the next layer. #_ave: the average of the outputs of all GRU steps is connected to the next layer; SimpleCNN: Convolutional layer has fixed kernel size; Multi_CNN: Convolutional layer has different kernel size; Details of all these mentioned architectures are shown in Supplementary file Fig. S2. The last row is our proposed method. ○ denotes that the methods have no significant difference (two-sample t-test) with the proposed. */** denote respectively that the methods are significantly worse than the proposed model with P = .05/0.01.

3.3. Comparison of TC-extracting strategies

Besides using ICA to extract TCs, we further tested the performance of the *MsRNN* by using TCs obtained from brain parcellation using both AAL template and Brainnetome Atlas, where the TCs of each brain regions of interests (ROI) were calculated by averaging the voxel-wise time series within each ROI. The dimension of TCs for AAL atlas is 170(time points) × 116(ROIs) and[a] 170(time points) × 273 (ROIs) for Brainnetome Atlas. *MsRNN* models were separately trained and evaluated, as shown in Fig. 3a and Table S3, the TCs generated from ICA achieved the best performance, surpassing the AAL feature extraction strategies by at least 7% on AUC ($P = 3{\cdot}0\text{e-}2$, two-sample *t*-test). This is likely due to the ability of ICA to capture variability in the components among subjects and is also consistent with earlier work showing that ICA time courses show better performance than fixed ROIs for graph theory metrics [35].

3.4. Estimating the most discriminating ICs

The ultimate goal of fMRI classification studies is to identify a collection of statistical features that can serve as reliable imaging biomarkers for disease diagnosis and are reproducible across multiple datasets. Despite extraordinary classification performance, the lack of interpretability often restricts the application of deep learning methods. Some previous work tried to open the black box of deep learning by analyzing the weight matrix of the trained model [9,12]. Generally speaking, the most important features are those whose removal can cause the most significant performance decrease compared to other features. Here we proposed a leave-one-IC-out method to leave one IC's time course out, and used the remaining 49 ICs' time course to train the model. After that, we compared the alteration of classification performances by looping all 50 ICs (shown in Fig. 3b). As a result, TCs from two components: *1)* putamen and caudate which are parts of striatum; *2)* declive and uvula which are parts of the cerebellum (Fig. 3c), contributed the top 2 group-discriminating time courses. Table 3 listed the Talairach labels of the two components. Note that similar findings of the most group-discriminating ROIs were obtained from both AAL and Brainnetome atlas.

4. Discussion

As known, the current clinical diagnosis of schizophrenia is based solely on clinical manifestations. In recent years, many studies attempted to find stable neuroimaging-based biomarkers by machine learning techniques. To the best of our knowledge, this is the first attempt to apply an RNN model directly on fMRI time courses for schizophrenia diagnosis, which avoids second-level correlation analysis and make full use of time-varying functional network information. Accuracies of 83·2% and 80·2% were obtained in the multi-site pooling classification and leave-one-site-out transfer prediction between schizophrenia patients and healthy controls respectively, yielding 4% improvement of accuracy compared to conventional approaches, suggesting a remarkable increase of the discriminative power *via* deep learning in neuroimaging predictions. The promising results may benefit from the following two aspects: *1)* the proposed *MsRNN* can learn both temporal and spatial information simultaneously based on time

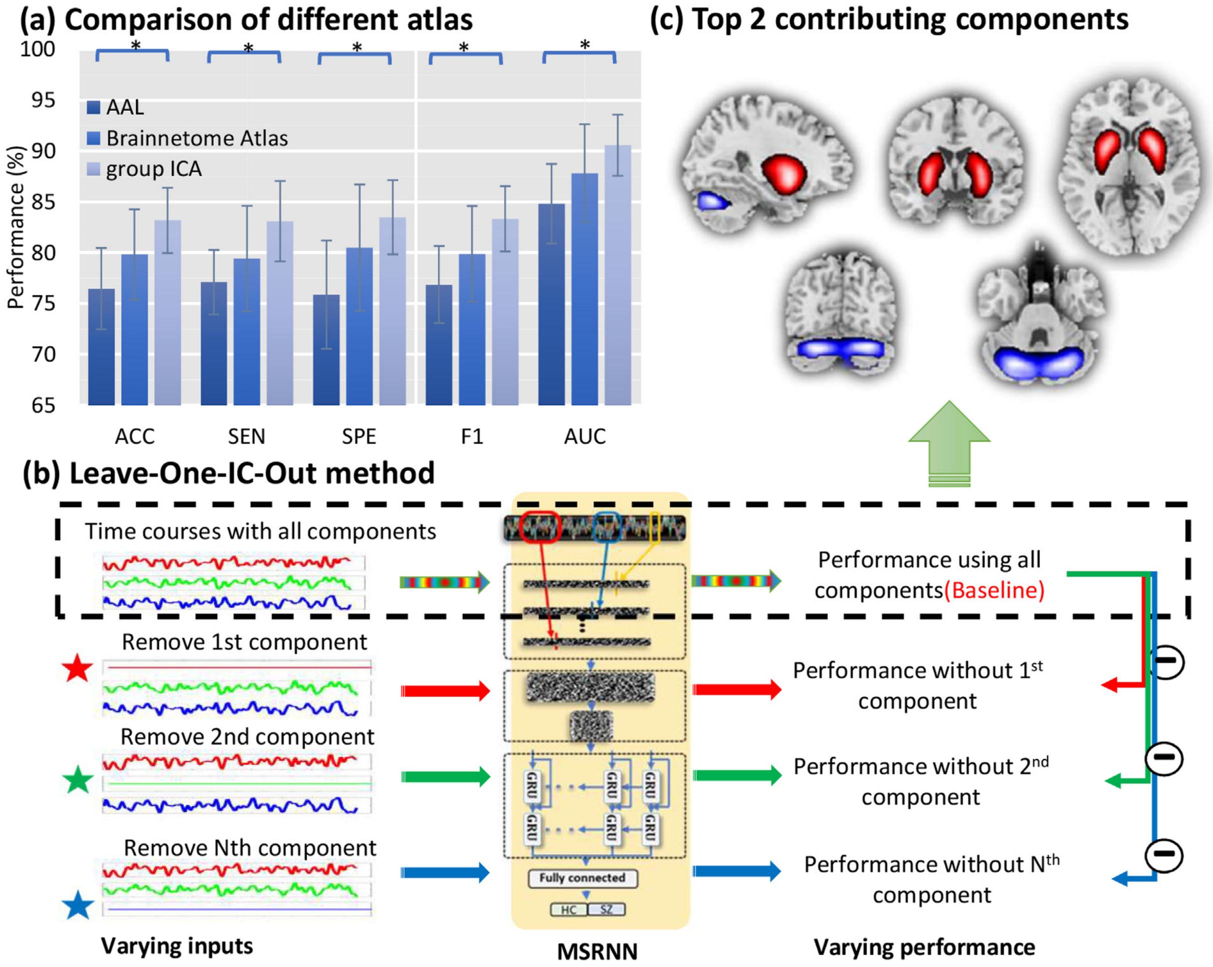
(a) Comparison of different atlas
Performance (%)
100
95
90
85
80
75
70
65
*
AAL
Brainnetome Atlas
group ICA
ACC
SEN
SPE
F1
AUC
(c) Top 2 contributing components
(b) Leave-One-IC-Out method
Time courses with all components
Remove 1st component
Remove 2nd component
Remove Nth component
Varying inputs
GRU
Fully connected
HC
SZ
MSRNN
Performance using all components(Baseline)
Performance without 1st component
Performance without 2nd component
Performance without Nth component
Varying performance

Fig. 3. Comparison of different atlas and Leave-One-IC-Out method. (a) The MsRNN classification results using three different feature selection methods. * $P < .05$ (two-sample *t*-test). (b) Leave-one-IC-out method for estimating the contribution of each IC. (c) Top two discriminative independent components discovered using the leave-one-IC-out method.

Table 3
Talairach labels of the peak activations in spatial maps of selected ICs.

Area	Brodmann area	Volume (cc)	Random effects: Max Value (x, y, z)
IC_4			
Putamen		4.2/4.9	1.4 (−24, 12, 15)/1.4 (29, −8, 14)
Lentiform nucleus		1.6/1.3	1.4 (−28, −17, 13)/1.4 (14, −1, −2)
Parahippocampal Gyrus	34	0.8/0.8	1.4 (−23, −8, −16)/1.4 (32, −10, −13)
Claustrum		0.8/1.0	1.4 (−36, −13, 2)/1.4 (34, 1, 9)
Inferior Frontal Gyrus	13, 47	0.6/0.1	1.4 (−32, 10, −15)/1.4 (30, 13, −12)
Caudate		1.7/1.8	1.4 (−11, 17, 7)/1.4 (16, −8, 19)
IC_2			
Declive		2.9/3.0	1.9 (−27, −71, −22)/1.9 (21, −71, −22)
Uvula		0.5/0.8	1.6 (−27, −71, −25)/1.8 (24, −71, 24)
Pyramis		0.0/0.1	NA/1.6 (27, −71, −27)

courses rather than the second-level FNC features. Specifically, the multi-scale CNN module can capture the spatial correlation of components from different time scales (2TR~8TR), and the RNN module can leverage temporal information; *2)* the large-scale dataset (1100 subjects) provide us the opportunity to train the deep learning model sufficiently. From this view of point, the present study may mark a significant breakthrough for enhancing the capabilities of psychiatrists by bringing RNN-based deep learning method to the task of diagnosing brain disorders across sites. Such applications would be critical and useful in clinical practice to predict for the new imaging sites or subjects. We also noticed a recently published multi-center study using deep learning method to diagnose schizophrenia [9]. The deep discriminant autoencoder network proposed by Zeng et al., aiming at learning imaging site-shared functional connectivity features, achieved desirable discrimination of schizophrenia across multiple independent imaging sites. To clarify, the current study used an entirely different deep learning architecture (AutoEncoder [Zeng et al.] *vs.* MsRNN [ours]) and different input features for classification (functional connectivity [Zeng et al.] *vs.* time courses [ours]), which avoid the second-level computation of fMRI data.

As to the identified brain regions, the dominating component is related to the dorsal striatum in the classification of schizophrenia. The dorsal striatum, comprising caudate and putamen, primarily mediates cognition involving motor function, certain executive functions (*e.g.*, inhibitory control), and stimulus-response learning. It receives input from cortex, thalamus, hippocampus and amygdala, then projects its output information to thalamus. The thalamus, which projects back to the cortex, thereby completing the circuit is also a component of the reward system that may suffer severely in SZ [36–38]. A similar impairment in SZ was verified in multiple resting-state fMRI studies [39] and cognitive studies [40]. For example, Yoon et al. [41] observed a link between impaired prefrontal-basal ganglia functional connectivity and the severity of psychosis, and Sarpal et al. [42] found a negative relationship between the functional connectivity of striatal regions and reduction in psychosis.

Another cerebellum component consist of declive, uvula and pyramis. The cerebellum is engaged in basic cognitive function such as attention, working memory, verbal learning and sensory discrimination, has led to an emerging interest in the role of the cerebellum in schizophrenia [43]. Structural and functional cerebellar abnormalities have been observed in schizophrenia, with evidence the impairment in white matter integrity in specific cerebellar lobes [44], as well as the abnormal size and a significant decrease in cerebral blood flow during a broad range of cognitive tasks [43,45]. Besides, researchers have posited the role of the cerebellum in reinforcement learning, allowing for more direct convergence between the theories of cognitive dysmetria and impaired reinforcement learning in schizophrenia [46].

Across several studies, altered connectivity patterns between the striatum and cerebellum have been frequently found in schizophrenia. Abnormalities in the relationship between cortical and sub-cortical regions, in particular, the prefrontal cortex, thalamus, basal ganglia, and cerebellum, were observed in patients with schizophrenia and correlated primarily with deficits in executive functioning, as well as deficits in processing speed and working memory [45]. Su et al. [47] and Repovs et al. [48] provided evidence that the connectivity strength between cerebellum and caudate is associated with executive functioning loss in schizophrenia. Also, reduced functional connectivity between the cerebellum and medial dorsal nucleus of the thalamus in schizophrenia providing evidence of abnormalities in this portion of the cortico-cerebellar-thalamic-cortico circuit [9,12,45,49]. Our results suggest that the temporal dynamics in the two identified brain regions and their connectivity are highly different between HC and SZ, which may serve as potential biomarkers for SZ discrimination.

The proposed model is stable and robust. Fig. S1 shows the learning curves on training and validation data while optimizing the parameters of *MsRNN*. The model convergent quickly during the first 100 epochs and reached a steady point after around 300 epochs. Since the number of hidden nodes in GRU layer may directly affect the learning capacity of a GRU model., we compared the performance of *MsRNN* model with a varying number of hidden units (*i.e.* [2^1, 2^2, 2^3 …, 2^{10}]) to validate the influence of the number of hidden notes in GRU layer. The statistical results indicate that our proposed model is not sensitive to the number of hidden units (Fig. S1b). The model can reach an over 80% classification accuracy with a range of 2^3~2^9 GRU hidden nodes. More hyperparameters about MsRNN including batch size, number of filters, scales of filters were analyzed thoroughly (Table S6-S8). The results show that the proposed MsRNN model is quite robust and not sensitive to these hyperparameters. Moreover, the hyperparameters combination we used in this work is close to an optimal solution. We also compared the influence of multiple training-testing ratios (Table S9), results show that the higher training-testing ratio is, the better performance MsRNN model achieves, which is consistent with the previous finding [9], suggesting further potential improvement of our proposed method when gathering more samples for modeling. Finally, to study the influence of the number of ICs, we further compared four different ICA component settings (Table S10). The two-sample *t*-test results show that only when the number of ICs is 16, the classification accuracy is less attractive than using 50 ICs(proposed), however, using more ICs does not show significant improvement, and many previous studies use a similar number of ICs as we did [50,51].

The current study has a few limitations. One is that information on antipsychotic or mood stabilizing medications for part of the patients were unavailable, which makes it difficult to assess the medication effect that may result in specific functional changes [6]. Secondly, the time courses were filtered within the range of 0·01–0·1HZ during the preprocessing step. However, the discriminative functional activity in the human brain may occur in a higher frequency range. Since the proposed *MsRNN* model can be applied to classify using either magnetoencephalogram (MEG) or electroencephalography (EEG) data due to its feasibility to higher temporal resolution data [6], therefore, a more stable and generative deep learning classification model may be designed by fusing multi-modalities to extract fused features and apply them to the RNN classification model in the future [52]. Another limitation is that even though head motion effect has been substantially attenuated through preprocessing procedures, it may not be completely removed and may remain certain influences. Moreover, the fMRI data acquisition protocols for all sites were set up by the same experienced experts and more harmonized in our study. Therefore the classification performance of the proposed model may be weighted down a bit if the data acquisition protocols in new sites are very different from each other. We admit that the proposed MsRNN model is still a preliminary model which did not give each hidden state a specific weight. One complementary strategy which may enhance GRU's performance is “attention” mechanism that can learn the weight of each hidden state automatically [53]. Furthermore, interpretation of deep learning networks remains an emerging but key field of research, our future work will focus more on a better interpretation of deep learning results, which would provide us with more clues on identifying potential biomarkers.

In summary, to the best of our knowledge, this is the first attempt to enable RNN directly to work on time courses of fMRI components in schizophrenia classification. The model takes advantage of high-level spatiotemporal information of fMRI data, and the high classification performances indicate the advantages of the proposed model. Also, the proposed leave-one-IC-out strategy provides a potential solution for increasing the clinical interpretability of the deep learning-based methods. Our work promises great potentials of deep-chronnectome-learning and a broad utility on neuroimaging applications, *e.g.*, the extension to MEG, EEG learning.

Funding sources

This work was supported by the Natural Science Foundation of China (No. 61773380), the Strategic Priority Research Program of the Chinese Academy of Sciences (No.XDB32040100), Beijing Municipal Science and Technology Commission (Z181100001518005), the National Institute of Health (1R56MH117107, R01EB005846, R01MH094524, and P20GM103472) and the National Science Foundation (1539067). The authors report no financial interests or potential conflicts of interest. No funders played a role in the study. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Declaration of Competing Interest

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

Acknowledgments

The current research was designed by WZY and JS. Acquisition of data was performed by TZJ, MS, YC, LZF, ZYY, NMZ, KBX, LXL, JC, YCC, HG, PL, LL, PW, HNW, HLW, HY, JY, YFY, HXZ, DZ. The data was analyzed by WZY, SFL, VDC, JS. The draft was written by WZY, VDC and JS. All authors contribute to the discussion of the results and have approved the final manuscript to be published.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.08.023.

References

[1] Calhoun VD, Miller R, Pearlson G, Adali T. The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. Neuron 2014;84(2): 262–74.

[2] Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A method for making group inferences from functional MRI data using independent component analysis. Hum Brain Mapp 2001;14(3):140–51.

[3] Lynall M-E, Bassett DS, Kerwin R, McKenna PJ, Kitzbichler M, Müller U, et al. Functional connectivity and brain networks in schizophrenia. J Neurosci 2010;30(28): 9477–87.

[4] Sui J, Qi S, van Erp TGM, Bustillo J, Jiang R, Lin D, et al. Multimodal neuromarkers in schizophrenia via cognition-guided MRI fusion. Nat Commun 2018;9(1):3028.

[5] Rashid B, Arbabshirani MR, Damaraju E, Cetin MS, Miller R, Pearlson GD, et al. Classification of schizophrenia and bipolar patients using static and dynamic resting-state fMRI brain connectivity. Neuroimage 2016;134:645–57.

[6] Liu S, Wang H, Song M, Lv L, Cui Y, Liu Y, et al. Linked 4-way multimodal brain differences in schizophrenia in a large Chinese Han population. Schizophr Bull 2018; 45(2):436–49 [:sby045-sby].

[7] Yan W, Plis S, Calhoun VD, Liu S, Jiang R, Jiang TZ, et al. Discriminating schizophrenia from normal controls using resting state functional network connectivity: A deep neural network and layer-wise relevance propagation method. 2017 IEEE 27th international workshop on machine learning for signal processing (MLSP); 2017 [25-28 Sept. 2017].

[8] Yan W, Zhang H, Sui J, Shen D. Deep chronnectome learning via full bidirectional long short-term memory networks for MCI diagnosis. International conference on medical image computing and computer-assisted intervention. Springer; 2018.

[9] Zeng L-L, Wang H, Hu P, Yang B, Pu W, Shen H, et al. Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI. EBioMedicine 2018;30:74–85.

[10] Dvornek NC, Ventola P, Pelphrey KA, Duncan JS. Identifying autism from resting-state fMRI using long short-term memory networks. In: Wang Q, Shi Y, Suk H-I, Suzuki K, editors. Machine learning in medical imaging: 8th international workshop, MLMI 2017, held in conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, proceedings. Cham: Springer International Publishing; 2017. p. 362–70.

[11] Durstewitz D, Koppe G, Meyer-Lindenberg A. Deep neural networks in psychiatry. Mol Psychiatry 2019https://www.nature.com/articles/s41380-019-0365-9.

[12] Kim J, Calhoun VD, Shim E, Lee JH. 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 2015;124:127–46 Pt A.

[13] Davatzikos C, Sotiras A, Fan Y, Habes M, Erus G, Rathore S, et al. Precision diagnostics based on machine learning-derived imaging signatures. Magn Reson Imaging 2019https://www.sciencedirect.com/science/article/abs/pii/S0730725X18306301?via%3Dihub.

[14] Identification of temporal transition of functional states using recurrent neural networks from functional MRI. In: Li H, Fan Y, editors. Medical image computing and computer assisted intervention – MICCAI 2018. Cham: Springer International Publishing; 2018 2018.

[15] Hochreiter S, Schmidhuber J. Long short-term memory. J Neural Comput 1997;9(8): 1735–80.

[16] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling arXiv preprint arXiv:14123555 ; 2014.

[17] Dvornek NC, Ventola P, Pelphrey KA, Duncan JS. Identifying autism from resting-state fMRI using long short-term memory networks. Machine learning in medical imaging MLMI (Workshop), 10541. ; 2017. p. 362–70.

[18] Roy S, Kiral-Kornek I, Harrer S. ChronoNet: A deep recurrent neural network for abnormal EEG identification arXiv preprint arXiv:180200308 ; 2018.

[19] Brain decoding from functional mri using long short-term memory recurrent neural networks. In: Li H, Fan Y, editors. Medical image computing and computer assisted intervention – MICCAI 2018. Springer International Publishing: Cham; 2018 2018.

[20] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 2002;15(1):273–89.

[21] Fan L, Li H, Zhuo J, Zhang Y, Wang J, Chen L, et al. The human Brainnetome atlas: a new brain atlas based on connectional architecture. Cereb Cortex 2016;26(8): 3508–26.

[22] He Y, A Evans. Magnetic resonance imaging of healthy and diseased brain networks. Front Hum Neurosci 2014;vol. 8:890.

[23] Du Y, Fan Y. Group information guided ICA for fMRI data analysis. Neuroimage 2013; 69:157–97.

[24] Satterthwaite TD, Elliott MA, Gerraty RT, Ruparel K, Loughead J, Calkins ME, et al. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. Neuroimage 2013;64:240–56.

[25] Zeng L-L, Wang D, Fox MD, Sabuncu M, Hu D, Ge M, et al. Neurobiological basis of head motion in brain imaging. Proc Natl Acad Sci 2014;111(16):6058.

[26] Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A 2005;102(27):9673.

[27] Yan C-G, Cheung B, Kelly C, Colcombe S, Craddock RC, Di Martino A, et al. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. Neuroimage 2013;76:183–201.

[28] Roy S, Kiral-Kornek I, S Harrer. ChronoNet: A deep recurrent neural network for abnormal EEG identification; 2018.

[29] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.

[30] Laurens van der Maaten GH. Visualizing Data using t-SNE. J Mach Learn Res 2008;9 (Nov):2579–605.

[31] Jo Y, Park S, Jung J, Yoon J, Joo H, Kim MH, et al. Holographic deep learning for rapid optical screening of anthrax spores. Sci Adv 2017;3(8):e1700606.

[32] Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46(1):389–422.

[33] Zhu J, Zou H, Rosset S, Hastie T. Multi-class AdaBoost. Stat Interface 2009;2(3): 349–60.

[34] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.

[35] Yu Q, Du Y, Chen J, He H, Sui J, Pearlson G, et al. Comparing brain graphs in which nodes are regions of interest or independent components: a simulation study. J Neurosci Methods 2017;291:61–8.

[36] Nestler EJ, Hyman SE, Malenka RC. Molecular basis of neuropharmacology: a foundation for clinical neuroscience. Conn: Appleton & Lange East Norwalk; 2001.

[37] Yager LM, Garcia AF, Wunsch AM, Ferguson SM. The ins and outs of the striatum: role in drug addiction. Neuroscience 2015;301:529–41.

[38] Ferré S, Lluís C, Justinova Z, Quiroz C, Orru M, Navarro G, et al. Adenosine-cannabinoid receptor interactions. Implications for striatal function. Br J Pharmacol 2010;160(3):443–53.

[39] Sui J, Pearlson GD, Du Y, Yu Q, Jones TR, Chen J, et al. In search of multimodal neuroimaging biomarkers of cognitive deficits in schizophrenia. Biol Psychiatry 2015;78 (11):794–804.

[40] Simpson EH, Kellendonk C, Kandel E. A possible role for the striatum in the pathogenesis of the cognitive symptoms of schizophrenia. Neuron 2010;65(5):585–96.

[41] Yoon JH, Minzenberg MJ, Raouf S, D'Esposito M, Carter CS. Impaired prefrontal-basal ganglia functional connectivity and substantia Nigra hyperactivity in schizophrenia. Biol Psychiatry 2013;74(2):122–9.

[42] Sarpal DK, Robinson DG, Lencz T, et al. Antipsychotic treatment and functional connectivity of the striatum in first-episode schizophrenia. JAMA Psychiat 2015;72(1):5–13.

[43] Andreasen NC, Pierson R. The role of the cerebellum in schizophrenia. Biol Psychiatry 2008;64(2):81–8.

[44] Kim D-J, Kent JS, Bolbecker AR, Sporns O, Cheng H, Newman SD, et al. Disrupted modular architecture of cerebellum in schizophrenia: a graph theoretic analysis, 40(6); 2014; 1216–26.

[45] Sheffield JM, Barch DM. Cognition and resting-state functional connectivity in schizophrenia. Neurosci Biobehav Rev 2016;61:108–20.

[46] Swain RA, Kerr AL, Thompson RF. The cerebellum: a neural system for the study of reinforcement learning. Front Behav Neurosci 2011;vol. 5:8.

[47] Su T-W, Lan T-H, Hsu T-W, Biswal BB, Tsai P-J, Lin W-C, et al. Reduced neuro-integration from the dorsolateral prefrontal cortex to the whole brain and executive dysfunction in schizophrenia patients and their relatives. Schizophr Res 2013;148(1):50–8.

[48] Repovs G, Csernansky JG, Barch DM. Brain network connectivity in individuals with schizophrenia and their siblings. Biol Psychiatry 2011;69(10):967–73.

[49] Anticevic A, Yang G, Savic A, Murray JD, Cole MW, Repovs G, et al. Mediodorsal and visual thalamic connectivity differ in schizophrenia and bipolar disorder with and without psychosis history. Schizophr Bull 2014;40(6):1227–43.

[50] Allen EA, Erhardt EB, Damaraju E, Gruner W, Segall JM, Silva RF, et al. A baseline for the multivariate comparison of resting-state networks. Front Syst Neurosci 2011;5:2.

[51] Du Y, Pearlson GD, Liu J, Sui J, Yu Q, He H, et al. A group ICA based framework for evaluating resting fMRI markers when disease categories are unclear: application to schizophrenia, bipolar, and schizoaffective disorders. Neuroimage 2015;122:272–80.

[52] Plis SM, Amin MF, Chekroud A, Hjelm D, Damaraju E, Lee HJ, et al. Reading the (functional) writing on the (structural) wall: multimodal fusion of brain structure and function via a deep neural network based translation approach reveals novel impairments in schizophrenia. Neuroimage 2018;181:734–47.

[53] Zhou P, Shi W, Tian J, Qi Z, Li B, Hao H, et al. Attention-based bidirectional long short-term memory networks for relation classification. Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers); 2016.

Medical Image Analysis 78 (2022) 102413

Contents lists available at ScienceDirect

Medical Image Analysis

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

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*

ARTICLE INFO

Article history:
Received 10 June 2021
Revised 27 October 2021
Accepted 1 March 2022
Available online 2 March 2022

Keywords:
Attention mechanism
Deep learning
Brain connectivity and activity
fMRI

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 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 $\sim$ 1100) and public autism datasets (ABIDE, n $\sim$ 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.

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. 1A). 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 disor-

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

https://doi.org/10.1016/j.media.2022.102413
1361-8415/© 2022 Elsevier B.V. All rights reserved.

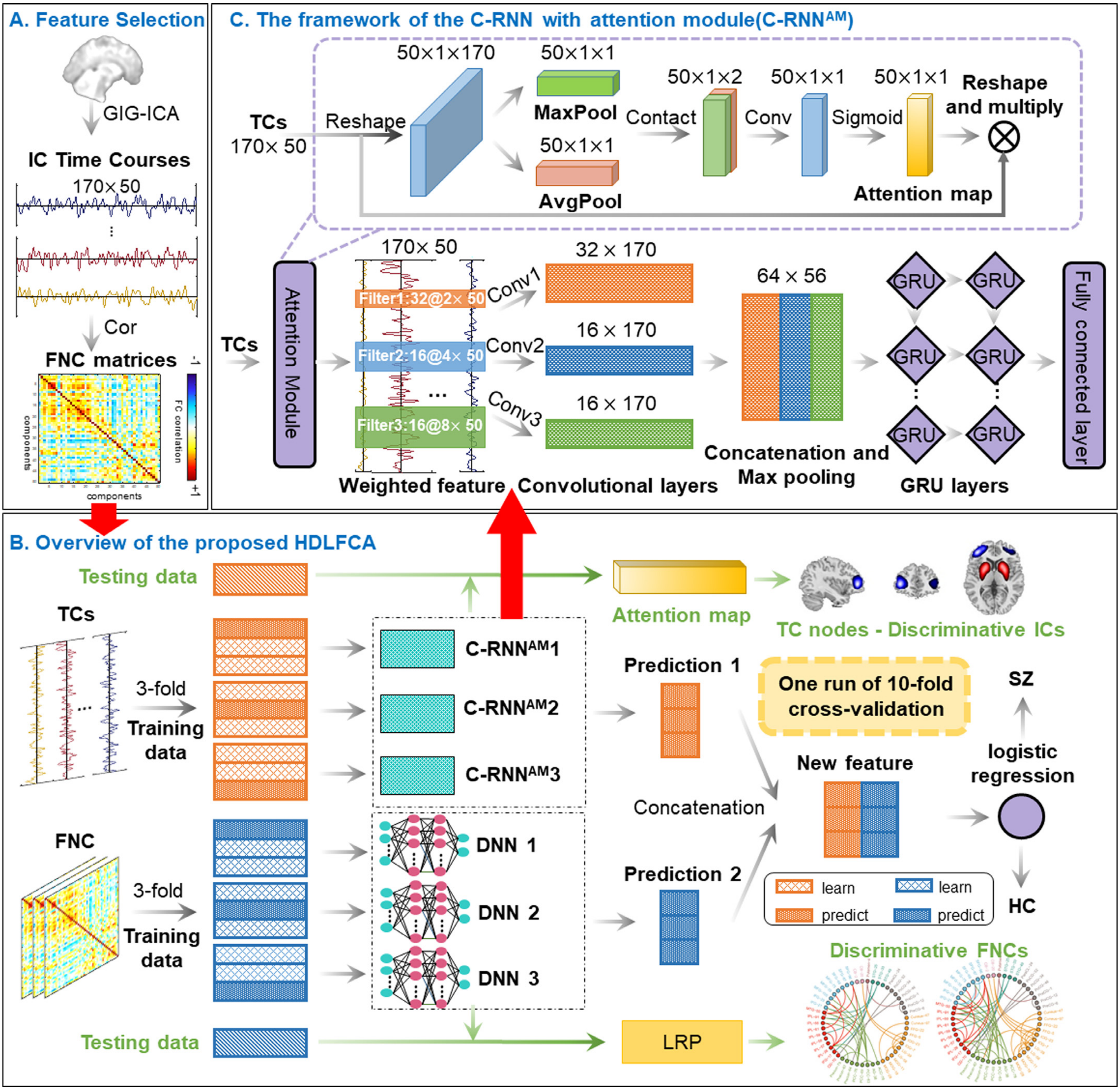

Fig. 1. The framework of the proposed HDLFCA in psychotic disorder classification. (A) Data preprocessing and Feature extraction. TCs was obtained by decomposing fMRI data using GIG-ICA, and FNCs was estimated from the TCs. (B) Overview of our proposed HDLFCA. C-RNN^{AM} and DNN were used to characterize 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.)

ders. 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.

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.

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 $M(X) \in R^{50\times 1\times 1}$. 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:

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

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 $P_2 \in R^{N\times 1}$. 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

$$R_{i\leftarrow j}^{(l,l+1)} = \left((1+\beta)\frac{z_{ij}^{+}}{z_j^{+}} - \beta\frac{z_{ij}^{-}}{z_j^{-}}\right)R_j^{(l+1)}$$

$$z_{ij} = x_i w_{ij}, z_j^{+} = \sum\nolimits_i z_{ij}^{+} + b_j^{+}, z_j^{-} = \sum\nolimits_i z_{ij}^{-} + b_j^{-}$$

where z_{ij}^{+} andz_{ij}^{-} denotes positive and negative activations respectively. b_j^{+} and b_j^{-} denote the positive and negative part of the bias item b_j.$R_j^{(l+1)}$ 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 iat 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 calcu-

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.

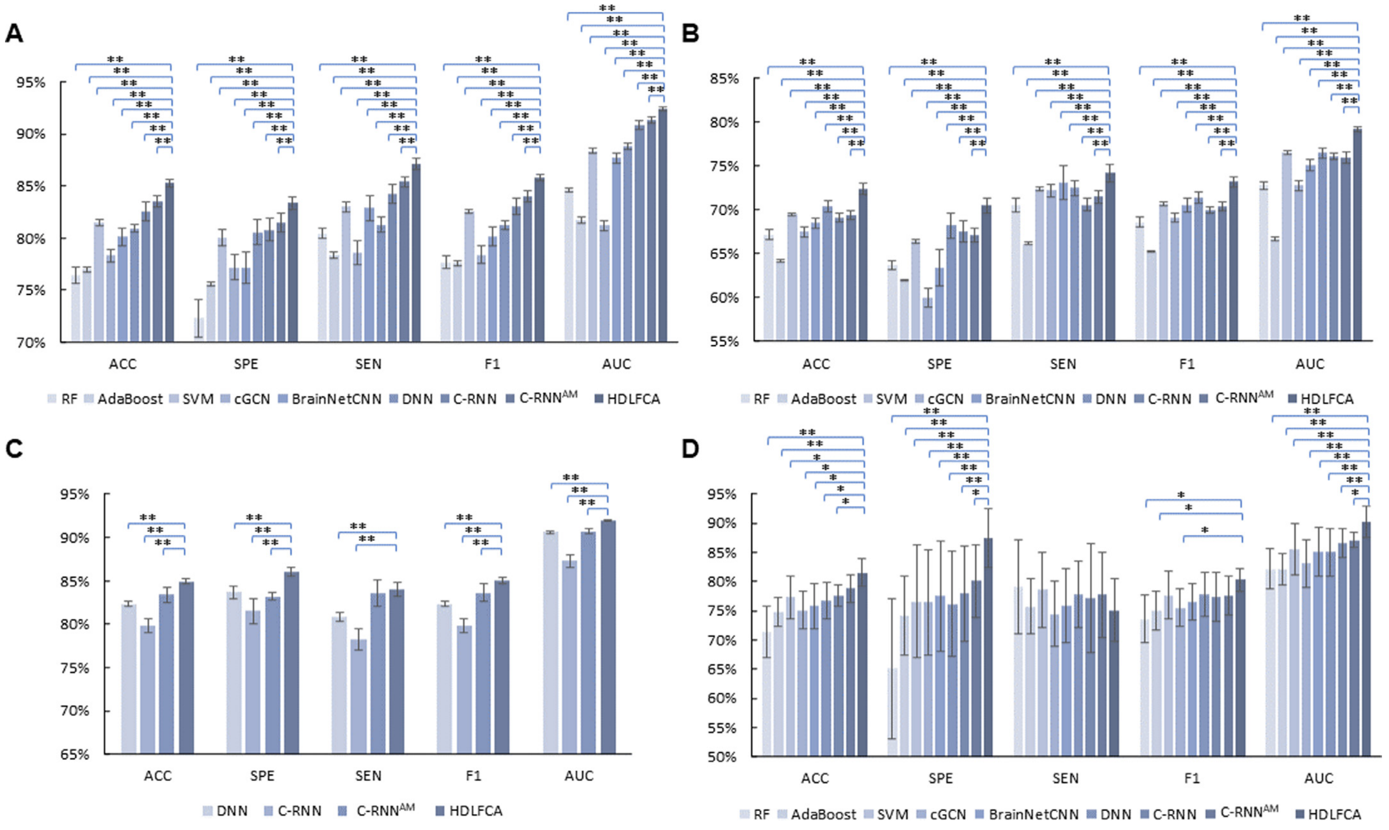

Fig. 2. The classification results of (A) multi-site pooling classification in in-house SZ datasets, (B) multi-site pooling classification in public ABIDE datasets, (C) multi-site pooling classification based on TCs or FNCs extracted by AAL atlas in in-house SZ datasets, and (D) leave-one-site-out classification 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.

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 value=0.05/0.01.

lated, 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. 2**C**. 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

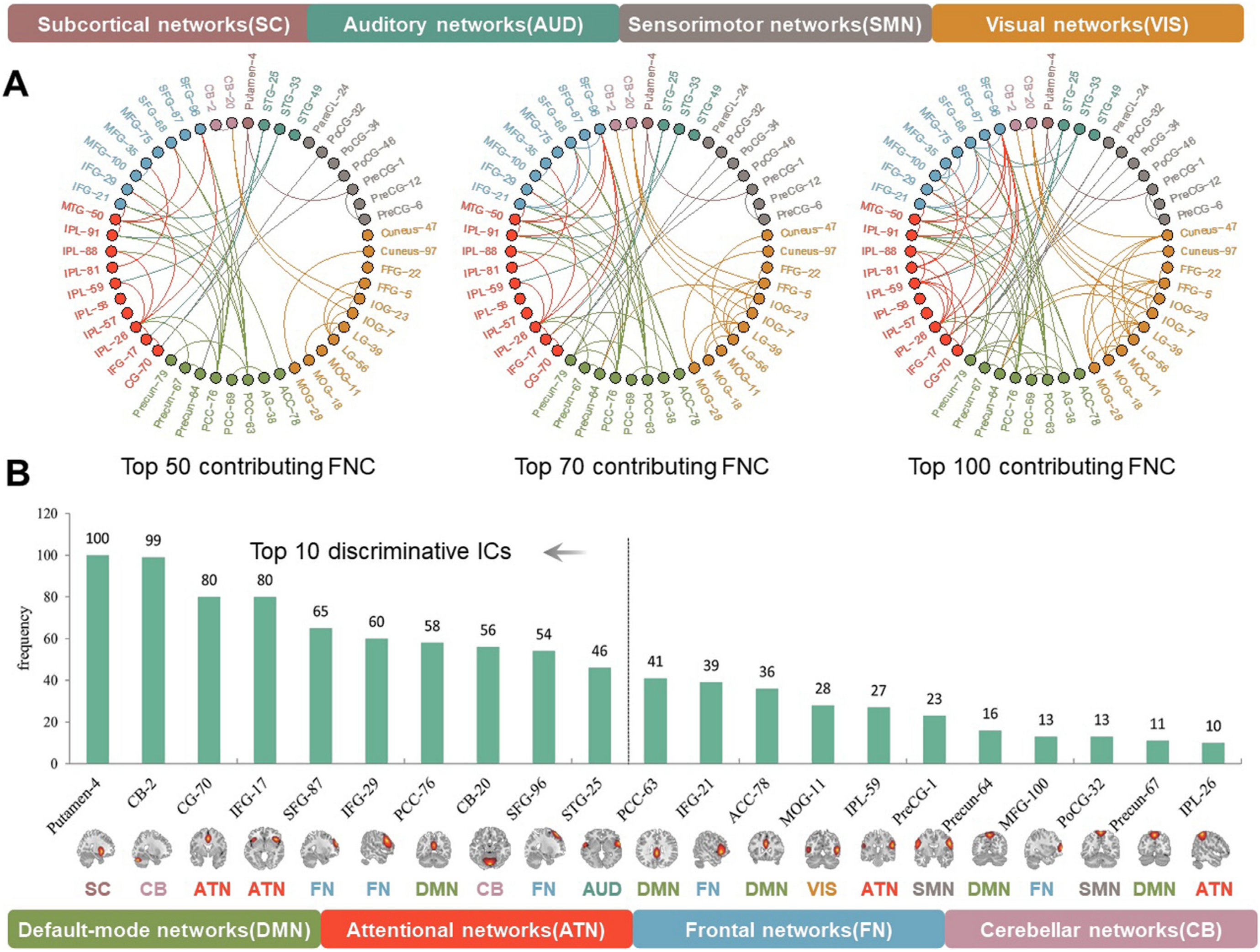

Fig. 3. The most HC-SZ discriminative features localization. (A) Illustration of the top 50, 70 and 100 contributing functional network connectivities identified by LRP. Connections between frontal network and default mode networks, frontal network and attention networks, and connections within visual 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 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.

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 sig-

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^{AM}	**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.

nal 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.

Declaration of Competing Interest

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

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.

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).

Supplementary materials

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

References

Allen, E.A., Erhardt, E.B., Damaraju, E., Gruner, W., Segall, J.M., Silva, R.F., 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, N.C., Pierson, R., 2008. The role of the cerebellum in schizophrenia. Biol. Psychiatry 64, 81–88.

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.

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

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

Calhoun, V.D., 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.

Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J., 2001. A method for making group inferences from functional MRI data using independent component analysis. Hum. Brain Mapp. 14, 140–151.

Cetin, M.S., Houck, J.M., Rashid, B., Agacoglu, O., Stephen, J.M., Sui, J., Canive, J., Mayer, A., Aine, C., Bustillo, J.R., 2016. Multimodal classification of schizophrenia patients with MEG and fMRI data using static and dynamic connectivity measures. Front. Neurosci. 10, 466.

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, K.W., Sabuncu, M.R., Kuceyeski, A., 2020. Sex classification using long-range temporal dependence of resting-state functionalMRItime series. Hum. Brain Mapp. 41, 3567–3579.

Dong, X., Lei, Y., Tian, S., Wang, T., Patel, P., Curran, W.J., Jani, A.B., 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.

Du, W., Ma, S., Fu, G.-S., Calhoun, V.D., Adalı, 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, E.A., He, H., Sui, J., Wu, L., Calhoun, V.D., 2016. Artifact removal in the context of group ICA: A comparison of single-subject and group approaches. Hum. Brain Mapp. 37, 1005–1025.

Du, Y., Fan, Y., 2013. Group information guided ICA for fMRI data analysis. Neuroimage 69, 157–197.

Du, Y., Fu, Z., Calhoun, V.D., 2018. Classification and prediction of brain disorders using functional connectivity: promising but challenging. Front. Neurosci. 12, 525.

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

Du, Y., Pearlson, G.D., Lin, D., Sui, J., Chen, J., Salman, M., Tamminga, C.A., Ivleva, E.I., Sweeney, J.A., Keshavan, M.S., 2017. Identifying dynamic functional connectivity biomarkers using GIG-ICA: Application to schizophrenia, schizoaffective disorder, and psychotic bipolar disorder. Hum. Brain Mapp. 38, 2683–2708.

Fletcher, P., McKenna, P.J., Friston, K.J., Frith, C.D., Dolan, R.J., 1999. Abnormal cingulate modulation of fronto-temporal connectivity in schizophrenia. Neuroimage 9, 342.

Fletcher, P.C., Frith, C.D., Grasby, P.M., Friston, K.J., Dolan, R.J., 1996. Local and distributed effects of apomorphine on fronto-temporal function in acute unmedicated schizophrenia. J. Neurosci. Methods 16, 7062.

Guclu, U., van Gerven, M.A.J., 2017. Modeling the dynamics of human brain activity with recurrent neural networks. Front. Comput. Neurosci. 11.

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

Jafri, M.J., Pearlson, G.D., Stevens, M., Calhoun, V.D., 2008. A method for functional network connectivity among spatially independent resting-state components in schizophrenia. Neuroimage 39, 1666–1681.

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, J.S., Bolbecker, A.R., Sporns, O., Cheng, H., Newman, S.D., Puce, A., O'Donnell, B.F., Hetrick, W.P., 2014. Disrupted modular architecture of cerebellum in schizophrenia: a graph theoretic analysis. Schizophr. Bull. 40, 1216–1226.

Kim, J., Calhoun, V.D., 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.

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

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

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

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, M.R., Yeo, B.T.T., 2019. Resting brain dynamics at different timescales capture distinct aspects of human behavior. Nat. Commun. 10.

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, V.D., Jiang, T., Sui, J., 2019. Linked 4-way multimodal brain differences in Schizophrenia in a large Chinese Han population. Schizophr. Bull. 45, 436–449.

Luo, N., Sui, J., Abrol, A., Chen, J., Turner, J.A., 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.

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

McCutcheon, R.A., Abi-Dargham, A., Howes, O.D., 2019. Schizophrenia, dopamine and the striatum: from biology to symptoms. Trends Neurosci. 42, 205–220.

Rashid, B., Arbabshirani, M.R., Damaraju, E., Cetin, M.S., Miller, R., Pearlson, G.D., Calhoun, V.D., 2016. Classification of schizophrenia and bipolar patients using static and dynamic resting-state fMRI brain connectivity. Neuroimage 134, 645–657.

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, T.D., Koutsouleris, N., Erus, G., Doshi, J., Wolf, D.H., Fan, Y., Gur, R.E., Gur, R.C., Meisenzahl, E.M., 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.

Sarpal, D.K., Argyelan, M., Robinson, D.G., Szeszko, P.R., Karlsgodt, K.H., John, M., Weissman, N., Gallego, J.A., Kane, J.M., Lencz, T., 2016. Baseline striatal functional connectivity as a predictor of response to antipsychotic drug treatment. Am. J. Psychiatry 173, 69–77.

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

Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Reiss, A.L., Greicius, M.D., 2007. Dissociable intrinsic connectivity networks for salience processing and executive control. J. Neurosci. 27, 2349–2356.

Skåtun, K.C., Kaufmann, T., Doan, N.T., Alnæs, D., Córdova-Palomera, A., Jönsson, E.G., 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.

Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., 2009. Correspondence of the brain's functional architecture during activation and rest. Proc. Natl. Acad. Sci. 106, 13040–13045.

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.

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, X.P., 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, L.M., Garcia, A.F., Wunsch, A.M., Ferguson, S.M., 2015. The ins and outs of the striatum: role in drug addiction. Neuroscience 301, 529–541.

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.

Yan, W., Plis, S., Calhoun, V.D., 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, G.D., Sui, J., Calhoun, V.D., 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.

Zhi, D., Calhoun, V.D., 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.

Deep Chronnectome Learning via Full Bidirectional Long Short-Term Memory Networks for MCI Diagnosis

Weizheng Yan[1,2,3], Han Zhang[3], Jing Sui[1,2], and Dinggang Shen[3(✉)]

[1] Brainnetome Center and National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2] University of Chinese Academy of Sciences, Beijing, China
[3] Department of Radiology and BRIC, University of North Carolina of Chapel Hill,
Chapel Hill, NC, USA
dgshen@med.unc.edu

Abstract. Brain functional connectivity (FC) extracted from resting-state fMRI (RS-fMRI) has become a popular approach for disease diagnosis, where discriminating subjects with mild cognitive impairment (MCI) from normal controls (NC) is still one of the most challenging problems. Dynamic functional connectivity (dFC), consisting of time-varying spatiotemporal dynamics, may characterize "chronnectome" diagnostic information for improving MCI classification. However, most of the current dFC studies are based on detecting discrete major "brain status" via spatial clustering, which ignores rich spatiotemporal dynamics contained in such chronnectome. We propose *Deep Chronnectome Learning* for exhaustively mining the comprehensive information, especially the hidden higher-level features, i.e., the dFC time series that may add critical diagnostic power for MCI classification. To this end, we devise a new Fully-connected *bidirectional* Long Short-Term Memory (LSTM) network (Full-BiLSTM) to effectively learn the periodic brain status changes using both past and future information for each brief time segment and then fuse them to form the final output. We have applied our method to a rigorously built large-scale multi-site database (i.e., with 164 data from NCs and 330 from MCIs, which can be further augmented by 25 folds). Our method outperforms other state-of-the-art approaches with an accuracy of 73.6% under solid cross-validations. We also made extensive comparisons among multiple variants of LSTM models. The results suggest high feasibility of our method with promising value also for other brain disorder diagnoses.

1 Introduction

Alzheimer's Disease (AD) is an irreversible neurodegenerative disease leading to progressive cognitive and memory deficits. Early diagnosis of its preclinical

W. Yan and H. Zhang—Contribute equally to this paper.

A. F. Frangi et al. (Eds.): MICCAI 2018, LNCS 11072, pp. 249–257, 2018.
https://doi.org/10.1007/978-3-030-00931-1_29

stage, mild cognitive impairment (MCI), is of critical value as timely treatment could be the most effective during this stage. Resting-state functional MRI (RS-fMRI) provides an opportunity to assess brain function non-invasively and has been successfully exploited to identify MCI [1]. To capture the time-varying information brain networks, dynamic functional connectivity (dFC) was proposed to characterize the time-resolved connectome, i.e., chronnectome, mostly using sliding-window correlation approach [2,4]. While promising, many current studies have not deeply exploited the rich spatiotemporal information of the chronnectome and utilized it in classification. For example, many studies focused on group comparison by detecting a set of discrete major brain status via clustering time-resolved FC matrices and further calculating their occurrence and dwelling time [4]. Inspired by the new finding that the brain dynamics are hierarchically organized in time (i.e., certain networks are more likely to occur preceding and/or following others [5]), we propose to learn diagnostic features in an end-to-end deep learning framework to better classify MCI.

Recurrent neural networks (RNNs) is a powerful neural sequence learning model for time series analysis. LSTMs are improved RNNs that can effectively solve the "gradient exploding/vanishing" problem by controlling information flow with several gates [6]. It has recently been demonstrated to be able to handle large-scale learning in speech recognition and language translation tasks [7]. However, there is still a significant gap between brain chronnectome modeling and common time series analysis. Directly applying LSTM to dFC-based MCI diagnosis is non-trivial: *(1)* Brain is extraordinary complex whose dynamics could be substantially different from natural language interpretation. *(2)* The background noise is usually more intense in the brain dFC signals, compared to audio/video signals, making it very difficult to capture. *(3)* The brain may continuously use contextual information for guiding higher-level cognitive functions rather than produce an output at the end of the time series with a strict direction. Therefore, a general LSTM could not be suitable for brain chronnectome-based classification. To solve this problem, we propose a new deep learning framework that changes the traditional LSTM in two aspects. *First*, we create Full-LSTM that connects the outputs of all cells to a "fusion" layer to capture a common time-invariant status-switching pattern, based on which the MCI can be diagnosed. *Second*, to excavate the contextual information hidden in the dFC, we further use a bidirectional LSTM (BiLSTM) to access long-range context in both directions [8]. We hereby come out with an end-to-end chronnectome-based classification model, namely *Full-BiLSTM*. The performance of our proposed method has been compared with state-of-the-art methods on ADNI-2 database. As the first "Deep Chronnectome Learning" study, we comprehensively compared the performance of three variants of LSTMs and reported the effect of different hyperparameters. The results support our hypothesis and significantly improved MCI diagnosis.

2 Methods

2.1 Computing dFC via a Sliding Window Method

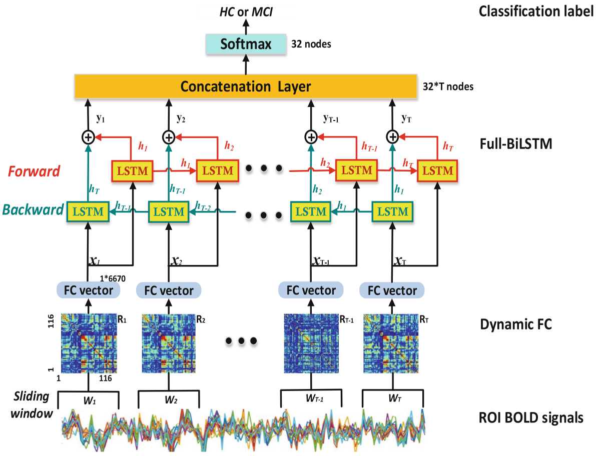

Fig. 1. Overview of the Full-BiLSTM for MCI classification.

For each subject, the whole-brain time-varying connectivity matrices are computed based on $M(M = 116)$ ROIs from the automated anatomical labeling (AAL) template using a sliding window approach [3,4]. As shown in Fig. 1, the averaged BOLD time-series S_i in ROI i are first computed. Then, the window $\{W_t\}$ are generated and applied to $S = \{S_i\}$, where T is the total number of sliding windows. Next, for each W_t, an FC matrix R_t of size $M * M$ that includes FC strengths between all pairs of S_{it} are calculated. Thus, for each subject, a set of $R_t (t = 1, 2, \ldots, T)$ are obtained, representing the subjects' whole-brain dFC. Due to the symmetry of each R_t, all FC strengths in R_t among M ROIs corresponding to a window t are converted to a vector x_t with $M(M-1)/2$ elements. Therefore, all the dFC time series from the k_{th} subject can be represented by a matrix $X^k = [x_1^k, x_2^k, \ldots, x_t^k]$ with a size of $T * \{M(M-1)/2\}$ and used as input to Full-BiLSTM classification model.

2.2 Fully-Connected Bidirectional LSTM (Full-BiLSTM)

Long Short-Term Memory (LSTM). LSTMs incorporates recurrently connected units, each of which receives an input h_{t-1} from its previous unit as well as the current input x_t for the current time point t. Each unit has its memory updating the previous memory c_{t-1} with the current input modulation. The network takes three inputs: x_t, h_{t-1}, and c_{t-1}, and has two outputs: h_t (the output of the current cell state) and c_t (the current cell state). Three gates separately controls input, forget, output. The unit can be expressed as:

$$Input\ Gate\colon\ i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{1}$$

$$Forget\ Gate\colon\ f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \tag{2}$$

$$Output\ Gate\colon\ o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{3}$$

$$\mathit{Input\ Modulation}\colon\ g_t = \phi(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{4}$$

$$\mathit{Memory\ Cell\ Update}\colon\ c_t = i_t \odot g_t + f_t \odot c_{t-1} \tag{5}$$

$$\mathit{Output}\colon\ h_t = o_t \odot \tanh(c_t) \tag{6}$$

Specifically, the input gate i_t controls how much influence the inputs x_t and h_{t-1} exerts to the current memory cell (Eq. 1). The forget gate f_t controls how much influence the previous memory cell c_{t-1} exerts to the current memory cell c_t (Eq. 2). Output gate controls how much influence the current cell c_t has on the hidden state cell h_t (Eq. 3). The memory cell unit c_t is a summation of two components: the previous memory cell unit c_{t-1}, which is modulated by f_t and g_t (Eq. 4), and a weighted combination of the current input and the previous hidden state, modulated by the input gate i_t (Eq. 5). Likewise, cell state is filtered with the output gate $o(t)$ for a hidden state updating (Eq. 6), which is the final output from an LSTM cell. With the inputting dFC time series, $W_{x\cdot}$ matrices (containing weights applied to the current input) and $W_{h\cdot}$ matrices (representing weights applied to the previous hidden state) can be learned, $b_{\cdot}$ vectors are biases for each layer, σ is sigmoid, ϕ is tanh function, and $\odot$ denotes element-wise multiplication.

Bidirectional LSTM (BiLSTM). BiLSTM is an effective solution that gets access to both preceding and succeeding information (i.e., context) by involving two separate hidden layers with opposite information flow directions [9]. For a brief description, we denote a process of an LSTM cell as H. BiLSTM first computes the forward hidden $\overrightarrow{h}$ and the backward hidden sequence $\overleftarrow{h}$ separately (Eqs. 7–8), and then combines $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ to generate the final output y_t (Eq. 9). The $W_{x\cdot}$ and $W_{h\cdot}$ matrices in (Eqs. 7–8) are the same as those in (Eqs. 1–4). The $W_{\overrightarrow{h}y}$ (representing weights applied to the forward hidden state) and $W_{\overleftarrow{h}y}$ (representing weights applied to the backward hidden state) are learned with the inputting dFC time series. $b_{\cdot}$ vectors are biases for each layer.

$$\mathit{Forward\ LSTM}\colon\ \overrightarrow{h}_t = H(W_{x\overrightarrow{h}}x_t + W_{\overrightarrow{hh}}\overrightarrow{h}_{t-1} + b_{\overrightarrow{h}}) \tag{7}$$

$$\mathit{Backward\ LSTM}\colon\ \overleftarrow{h}_t = H(W_{x\overleftarrow{h}}x_t + W_{\overleftarrow{hh}}\overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}) \tag{8}$$

$$\mathit{Combined\ Output}\colon\ y_t = H(W_{\overrightarrow{h}y}\overrightarrow{h}_t + W_{\overleftarrow{h}y}\overleftarrow{h}_t + b_y) \tag{9}$$

Full-BiLSTM. The traditional BiLSTM classification model usually uses the final state y_T for classification [8]. However, this is insufficient for chronnectome-based diagnosis, because brain may continuously use contextual information to facilitate higher-level cognition and guide status transition, rather than producing a single output at the end of the scanning period. Therefore, the outputs of every repeating cell could be of equally important use and should be concatenated into a dense layer $Y = [y_1, \ldots y_t, \ldots, y_T]$ (see “Concatenation Layer” in Fig. 1).). With this layer, we may abstract a common and time-invariant dynamic transition pattern from all the BiLSTM cells which may represent a constant "trait" information of each subject, instead of the continuously varying brief brain status. While the latter could be of great use in previous status-based studies such as those used *Hidden Markov Chain* for status transition probability modeling in group-level comparison studies [5], it will inevitably lose the precious temporal information which could capture more subtle individual differences for the more challenging disease diagnosis studies. In our framework for MCI diagnosis, the dense layer Y is followed with softmax layer to get the final classification result.

3 Experiments and Results

3.1 Data Preprocessing

In this study, we use the publicly available Alzheimer's Disease Neuroimaging Initiative dataset (ADNI) to test our method. As shown in Table 1, 143 age- and gender-matched subjects (48 NCs with 164 RS-fMRI scans, and 95 MCIs with 330 RS-fMRI scans) were selected from ADNI-2 database. The goal of ADNI-2 study is to validate the use of various biomarkers including RS-MRI to find the best way to diagnose AD at pre-dementia stage. Each RS-fMRI scan was acquired using 3.0T Philips scanners at different medical centers. All the data were carefully reviewed by the quality control team in Mayo Clinic. ADNI is to date the largest, multi-site, rigorously controlled early AD diagnosis data. The RS-fMRI data were preprocessed following the standard procedure [1].

3.2 Dynamic Functional Connectivity Matrix

In this experiment, the window length was 90s (30 volumes) as suggested by previous dFC studies [4]. The window slides in a step of 2 volumes (6s), resulting in 54 segments of BOLD signals. For each subject and each scan, 54 FC matrices were obtained, reflecting the chronnectome. The upper half of the matrix containing 6670 unique dFC links were used and then reshaped into X^k with the size of $54 * 6670$.

3.3 Data Augmentation

Training deep learning models requires a large number of samples. Fortunately, only part of the dFC time series might be sufficient for discriminating MCIs from NCs because the FC dynamics could happen in a very brief period [5]. This allows us to conduct data augmentation to increase the sample size. Specifically, for each X^k, a continuous submatrix of length 30 were cropped as a new sample. By using a sliding window strategy

Table 1. Demographic information.

	NC	MCI
Number of scans	164	330
Age(mean(±std, yrs))	75.4 ± 6.2	72.0 ± 7.5
Gender(M/F)	72/92	178/152

with a stride of 1, the original X^k can be augmented for $54 - 30 + 1 = 25$ times (augmented by a factor of 25). The label of the augmented data from the same subject was kept the same. Of note, all augmented sequences belonging to the same subject were used solely in the training, or validation, or testing phase. In the testing phase, the predicted labels for all the augmented data from the same subject was derived with majority voting to determine the final label for this subject.

3.4 Full-BLSTM Parameters and Training Strategy

The Full-BiLSTM model was trained and evaluated using Keras. Data was split into 80% for training and 20% for testing (5-fold cross-validation). 10% of samples from training data were further selected for validation to monitor the training procedure. Training was stopped when the validation loss stopped decreasing for 20 epochs or when the maximum epochs had been executed. The testing data was applied to the trained model to evaluate the performance. The model was trained for minimizing the weighted cross-entropy loss function using stochastic gradient descent (SGD) optimizer. The learning rate (lr) was started from 0.001 and decayed over each update as follow: $lr_t = lr_{t-1}/(1 + decay_{rate} * epochs)$. The $decay_{rate}$ was 10^{-6}, and the maximum epochs was 200. The batch size was 32. The weights and biases were initialized randomly. To improve the generalization performance of the model and overcome the overfitting problem, we used a dropout method ($dropout = 0.5$) and $l_1 norm$ regularization ($l1 = 0.0005$).

3.5 Method Comparison

As dFC is novel in this field, the disease diagnosis works using dFC are quite limited. We compared our approach against various classifiers commonly used. The majority of the dFC studies focus on brain statuses detected by clustering, or the temporal variability of dFC series. Therefore, in the competing methods, we also use these two types of the dFC features for MCI classification. In summary, we compared our method with the classification models using: (1) static FC (sFC); (2) dFC-based brain statuses [4]; and (3) dFC variability [1], as detailed below.

sFC. The traditional FC method used in most of the FC studies are based on Pearson's correlation of full-length BOLD signals. After building sFC matrix, an SVM classifier is trained based on the sFC strengths.

Status-Based. Group-level chronnectome status is identified by using k-means clustering with all of the dFC matrices in the training data. The occurrence frequency of each status is computed to as features. Then, an SVM classifier is constructed based on the frequency features of all status.

Variability-Based. Based on the dFC matrices, the quadratic mean value is computed for each dFC. A total of 6670 features are generated for each subject representing the fluctuation of the signals. The features are further reduced using two-sample t-test. An SVM classifier is constructed based on the dFC variability features.

Table 2. Performance of different methods in MCI/NC classification.

Method	ACC(std)%	SEN(std)%	SPE(std)%	f1(std)%	AUC(std)%
Static FC + SVM	61.5(10.0)	74.0(9.2)	41.7(14.0)	70.9(8.2)	64.2(10.8)
dFC-variability	54.8(12.9)	54.4(12.3)	56.8(19.1)	60.5(12.3)	49.0(17.0)
dFC-status	61.3(10.0)	70.8(12.2)	47.2(13.6)	69.9(8.6)	61.9(15.9)
Full-LSTM32	71.9(5.9)	72.3(7.9)	70.5(15.1)	76.2(5.3)	75.9(5.8)
Full-BiLSTM32-Stack	69.0(5.0)	66.7(4.7)	73.0(9.2)	73.1(3.5)	79.2(2.7)
BiLSTM32-Last	71.0(10.3)	76.8(9.6)	60.9(12.8)	76.7(8.8)	75.9(6.0)
Full-BiLSTM32	73.6(3.7)	73.9(10.1)	73.5(7.3)	77.6(4.4)	79.8(6.9)

Notes: Blue-colored methods are the traditional methods; Methods in italic are LSTM-based methods; Our method is in bold italic; Red italic indicates the model without bi-directional LSTM or without Full-LSTM

The performance comparison results are summarized in Table 2 and Fig. 2 showing the ROI curves of all methods. Because of sample imbalance, the area under the ROC curve (AUC) was used as the main metric for comparing the performance of all the methods. Our method achieved 79.8% in AUC and significantly outperformed the traditional sFC and dFC methods. The dFC variability method achieved the lowest result, which could be caused by the severe noise in dFC time series. In contrast, our method could learn the intrinsic brain status transition, thus is more robust to such noise.

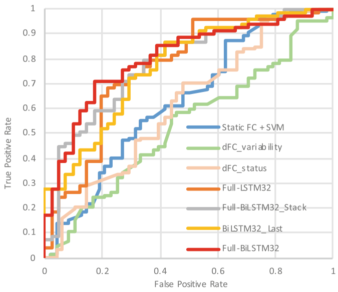

Fig. 2. ROC curves of different methods.

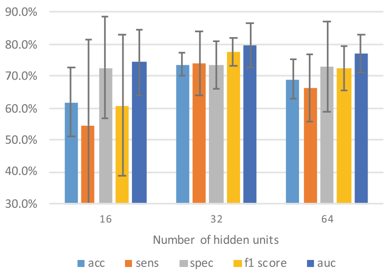

Fig. 3. Effect of different hidden units

To validate the advantage of Full-BiLSTM, we tested three other LSTM-based architectures. The BiLSTM_Last model uses the output of the last

BiLSTM cell for classification, as used in the traditional sequence processing studies. The Full-LSTM uses the same architecture as our method, but with uni-directional LSTM cells. To investigate whether a deeper BiLSTM layer could increase the performance, the third model is built using stacked Full-BiLSTM (two layers). All these three models use the same parameters as our Full-BiLSTM method. As shown in (Fig. 2), our model still outperformed all these three LSTM-based competing models. Specifically, we observed that (1) BiLSTM outperforms uni-directional LSTM; (2) Full-BiLSTM performs better than BiLSTM_Last; (3) A deeper model does not improve the final performance. In addition, we also compared the performance with and without data augmentation, and found that the accuracy was decreased by 2% without data augmentation. Furthermore, the number of hidden nodes in LSTM may directly affect the learning capacity of an LSTM network. Therefore, we compared the performance of Full-BiLSTM models with a varying number of hidden units, i.e., 16, 32, 64. As shown in Fig. 3, the Full-BiLSTM model with 16 hidden nodes has decreased performance and increased performance variability, compared to the Full-BiLSTM model with 32 hidden nodes. It is likely that 16 hidden units are too limited to store the sequential information of the dFC process. The model with 64 hidden nodes also has suboptimal performance, which could be attributed to overfitting.

The results together indicate that data augmentation and the choice of network structure are crucial for training an effective dFC-based classification model. Most notably, this is the first attempt to use a deep learning framework for individualized disease diagnosis based on dFC. Our results indicate that a sequence model can take advantage of more series information from dFC than the conventional methods. It is also worth noting that our model can be applied to other brain disorder diagnoses.

4 Conclusions

In this study, we proposed a new deep learning framework, a Full-BiLSTM model, for brain disease diagnosis using dynamic functional connectivity. To the best of our knowledge, this is the first attempt to propose the "deep chronnetome learning" framework and to prove its feasibility and superiority in a challenging MCI diagnosis task by using time-varying functional information. Comprehensive comparisons among different architectures of the LSTM model were conducted, and the insightful discussions on the influence of the hyperparameters were provided. In summary, the proposed model can not only effectively capture the trait-related brain dynamic changes from the spatiotemporally complex chronnectome, but also can be applied to improve classification of other brain disorders, which shows great promise to be used as a powerful tool to detect potential biomarkers in the community.

Acknowledgment. This work was supported in part by NIH grants (AG041721, AG049371, AG042599, and AG053867), NSFC grants (81471367 and 61773380), and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDBS01040102).

References

1. Chen, X., Zhang, H., Zhang, L., Shen, C., Lee, S., Shen, D.: Extraction of dynamic functional connectivity from brain grey matter and white matter for MCI classification. Hum. Brain Mapp. **38**(10), 5019–5034 (2017)
2. Calhoun, V.D., Miller, R., Pearlson, G., Adali, T.: The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. Neuron **84**(2), 262–274 (2014)
3. Rashid, B., Damaraju, E., Pearlson, G.D., Calhoun, V.D.: Dynamic connectivity states estimated from resting fMRI identify differences among Schizophrenia, bipolar disorder, and healthy control subjects. Front. Hum. Neurosci. **8**, 897 (2014)
4. Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., Calhoun, V.D.: Tracking whole-brain connectivity dynamics in the resting state. Cereb. Cortex **24**(3), 663–676 (2014)
5. Vidaurre, D., Smith, S.M., Woolrich, M.W.: Brain network dynamics are hierarchically organized in time. Proc. Natl. Acad. Sci. **114**(48), 12827–12832 (2017)
6. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
7. Sak, H., Senior, A., Beaufays, F.: Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: INTERSPEECH 2014, pp. 338–342 (2014)
8. Fan, B., Xie, L., Yang, S., Wang, L., Soong, F.A.: A deep bidirectional LSTM approach for video-realistic talking head. Multimed. Tools Appl. **75**(9), 5287–5309 (2016)
9. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional LSTM networks. IEEE Int. Joint Conf. Neural Netw. **4**, 2047–2052 (2005). https://doi.org/10.1109/IJCNN.2005.1556215