2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)
April 13-16, 2021, Nice, France

AN ATTENTION-BASED HYBRID DEEP LEARNING FRAMEWORK INTEGRATING TEMPORAL COHERENCE AND DYNAMICS FOR DISCRIMINATING SCHIZOPHRENIA

Min Zhao[1,2,3], Weizheng Yan[1,2,3], Rongtao Xu[2,3], Dongmei Zhi[1,2,3], Rongtao Jiang[1,2,3], Tianzi Jiang[1,2,3], Vince D Calhoun[4], Jing Sui[1,2,3,4]*

[1]Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China; [2]NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China; [3]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China; [4]Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS) Center, Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA, USA

2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-1246-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/ISBI48211.2021.9433919

ABSTRACT

The heterogeneity of schizophrenia makes it difficult to discover reliable imaging biomarkers, and most existing fMRI-based classification methods fail to combine temporal coherence between brain regions and temporal dynamics of brain activity. Therefore, we proposed a unified Hybrid Deep Learning Framework that effectively integrates temporal Coherence and Dynamics (HDLFCD) to classify psychiatric disorders by combining C-RNN, DNN and SVM. An attention module was also introduced into the C-RNN model to improve classification accuracy and interpretability without increasing the computation complexity. An accuracy of 85% was achieved in a large multi-site fMRI dataset with 542 healthy controls and 558 schizophrenia patients, in which striatum, dorsolateral prefrontal cortex and cerebellum were identified as the most group-discriminative brain regions by the attention module. Note that the proposed framework is an end-to-end general module, which not only shows high superiority in combining multiple sources of information, but also can be easily applied to integrate other multimodal data.

Index Terms— Attention mechanism, Schizophrenia, Temporal coherence, fMRI, Temporal dynamics, ICA

1. INTRODUCTION

Schizophrenia (SZ) has been one of the leading mental disorders that cause huge global disease burden, strongly affecting public health and quality of life. However, to discover the reliable and reproducible objective biomarkers for SZ diagnosis is still challenging and has a long way to go. Functional magnetic resonance imaging (fMRI), a non-invasive imaging technique, has attracted growing attention as a promising tool to identify functional abnormalities and potential biomarkers.

Recently, deep learning (DL) methods have been widely applied to fMRI data for schizophrenia diagnosis. For example, based on functional (network) connectivity (FNC), Kim et al. used deep neural network (DNN) for SZ diagnosis and L1-norm regularization for feature selection [1], and Zeng et al. proposed a deep discriminant autoencoder network for multi-site diagnostic classification [2]; while Yan et al. used the time courses (TCs) directly to discriminate SZ by multi-scale RNN [3]. However, these existing methods adopted either FNC or TCs only. FNC reflects the temporal coherence of neuronal populations activation of spatially separated brain regions, while time courses contain temporal dynamics of brain activity. To leverage the complementary information between temporal coherence and temporal fluctuations in fMRI data, we are motivated to combine functional connectivity and time courses together via a deep learning-based framework to improve classification performance.

Moreover, interpretability is very important for medical classification to understand how the imaging-based diagnosis can make decisions. To this end, attention mechanism, derived from human perception, was developed to improve the DL model interpretability. That is, humans focus on the target area where is needed instead of the whole scene and combine information from different fixations to guide decision making [4]. Although several studies have used attention mechanism to detect discriminative region localization in medical classification, most of them are based on structural images, like sMRI and CT [5, 6]. Hence, we are inspired to incorporate an attention module (AM) to optimize feature representations and capture discriminative brain regions simultaneously based on fMRI data.

Consequently, we proposed a unified Hybrid Deep Learning Framework that effectively integrates temporal Coherence and Dynamics (HDLFCD) to improve fMRI-based classification performance. As shown in Figure 1,

* Correspondence goes to Prof. Jing Sui kittysj@gmail.com

independent components (ICs) and their corresponding time courses (TCs) can be obtained by decomposing fMRI data via group information guided independent component analysis (GIG-ICA), and functional network connectivity (FNCs) can be calculated based on the TCs. Then DNN and SVM were applied to learn functional dependency between brain regions based on FNCs respectively, while the Convolutional Recurrent Neural Network with attention module (C-RNN^{AM}) was applied to capture temporal dynamics from TCs. Then prediction output from 3 schemes were combined to build a new feature matrix to generate the final decision by logistic regression.

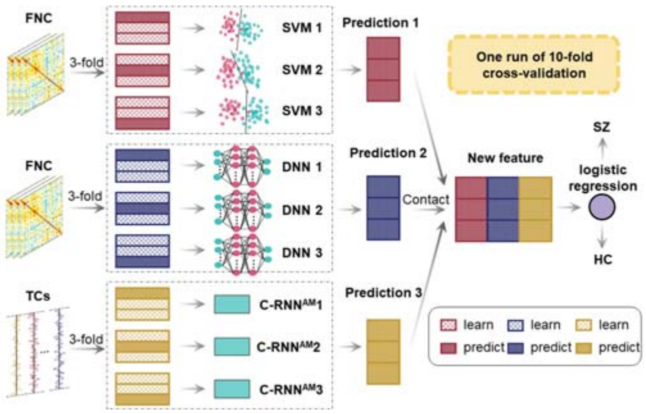

Fig. 1. Overview of Hybrid Deep Learning Framework that effectively integrates temporal Coherence and Dynamics

The proposed model has been applied on multi-site fMRI dataset including 1100 participants (542 HC and 558 SZ), and also compared with 5 alternative classification models including 3 traditional methods and 2 DL methods by 10-fold cross-validation for ten times.

2. METHODS

2.1. Data description and preprocessing

In this study, participants (558 patients with schizophrenia and 542 healthy controls) were recruited from 7 sites. Detail demographic information was listed in Table 1.

Table 1. Demographic information of datasets

	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

All resting-state fMRI data were preprocessed using the SPM software package (http://www.fil.ion.ucl.ac.uk/spm/). The processing pipeline included: 1) slice timing correction; 2) motion correction; 3) normalization into the standard Montreal Neurological Institute (MNI) space, resliced to 3×3×3 mm. The 100 stable group independent components (ICs) were first selected by GIG-ICA in the GIFT software (http://trendscenter.org/software/gift). Then, 50 ICs were selected as intrinsic connectivity networks that showed higher low-frequency spectral power and their peak activation fell on the grey matter with minimal overlap with white matter, ventricles, and edge regions. The following post-processing steps were performed on the TCs of selected 50 ICs: linear, quadratic and cubic detrending, regressing out six realignment parameters and temporal derivatives, despiking, and low-pass filtering (<0.15 Hz). In addition, age and gender were also regressed out. Then the TCs and static FNC matrices were used as the inputs of the C-RNN model, DNN and SVM respectively.

2.2. Hybrid Deep Learning Framework that effectively integrates temporal Coherence and Dynamics (HDLFCD)

2.2.1. Overview

As shown in Fig. 3, the HDLFCD used different models to describe heterogeneous input features to exploit complementary information among TCs and FNC. Due to abundant sequential temporal fluctuations in BOLD time series, we used the C-RNN^{AM} to capture temporal dynamic dependencies. As for FNC, DNN and SVM were applied to learn functional interaction patterns between brain regions. Their class probabilities were then taken as new features to train a meta-learner, whose output is the final prediction. Since we have used complex nonlinear transformations, logic regression was chosen as a meta-learner to combine the three models with different features. Cross-validation was conducted for first-level learners to avoid overfitting.

2.2.2. Attention module (AM)

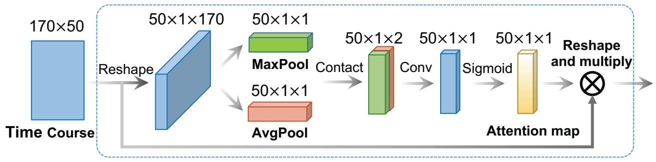

Fig. 2. Scheme of the attention module

The attention module aims to pay more attention to important brain regions and suppress unnecessary ones. Given the previously processed TCs as $X \in R^{T \times C}$, where T and C is the number of time points and components, respectively. Here, there are 170 time points and 50 components in total. First, TCs were reshaped to a C×1×T matrix. To aggregate temporal information fully, we adopted average-pooling and max-pooling operations to learn temporal statistics[7], resulting in two different temporal context descriptors: F^{avg} and F^{max}, and then concatenated them. A convolution layer was applied to produce a region attention map $M \in R^{C \times 1 \times 1}$ with a filter of 4×1 kernel size and same padding after sigmoid activation. The attention map provides the importance of components. The attention map was reshaped into the same size as X and then multiplied to original TCs, which helps to guide the network to focus on more important information instead of the full feature. 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.

2.2.3. Convolutional Recurrent Neural Network with attention module (C-RNN^{AM})

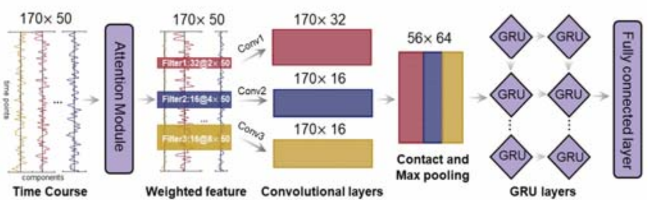

Fig. 3. The architecture of C-RNN with attention module

As shown in Fig. 3, the C-RNN network consists of an attention module, multiple 1D convolutional (Conv1D) layers, one contact and max pooling layer, two gated recurrent unit (GRU) layers and a fully connected layer. Conv1D layers with different scales were proposed to extract local information for multiple time scales of brain activity. The size of three convolutional filters are 32×2×50 (number of filters × time scales × ICs), 16×4 ×50, 16×8 ×50 with the same padding, respectively. The GRU layers were stacked to capture long-term dynamic dependencies. It is worth noting that the GRU layers were densely connected, which can provide short-cut paths during back-propagation. We averaged the GRU outputs to make full use of brain activity throughout the scan.

2.3. Experimental setup

Based on fMRI data from 7 sites, ten-fold cross-validation was conducted to evaluate classification performance. We first compared our methods with single feature-based methods and typical classifiers to verify the effectiveness of integrating temporal coherence and dynamics information in our framework. Ablation experiments on C-RNN architecture were further performed to evaluate the effectiveness of the proposed AM. All experiments were repeated ten times.

In this study, the DNN stacks one input layer, two hidden layers (32 and 16 hidden notes), and one output layer. L2 norm regularization and dropout strategy were also applied to avoid overfitting. All methods were implicated in the Keras (https://keras.io/) and sklearn (https://scikit-learn.org/stable/). We trained the models using the Adam optimizer with an initial learning rate of 0.001 and decay with 0.01. Three metrics including accuracy (ACC), specificity (SPE), and sensitivity (SEN) were used to evaluate the performance of methods.

3. RESULTS AND DISCUSSION

3.1. Comparison with single feature-based methods

The classification performance of our method along with single feature-based methods was reported in Table 2. The results demonstrated that: (1) Our method outperforms regular SVM, DNN and C-RNN^{AM} with single temporal coherence or dynamic features. Specifically, HDLFCD achieved an improvement of 2.6%, 4.3%, and 1.4% respectively in ACC. This demonstrates the necessity and effectiveness and of integrating temporal dynamics of brain activity and functional dependency. (2) Compared with SVM+DNN method only using single FNC, SVM+C-RNN^{AM} and DNN+C-RNN^{AM} using both FNC and TCs achieved higher classification performance, showing the superiority in leveraging the complementary information between FNC and TCs. (3) The proposed method was superior to three popular classical classifiers, i.e., Random Forest, AdaBoost and SVM. (4) We achieved 85% accuracy in distinguishing SZ from HCs. The competitive result is comparable to, if not better than, the recent studies on large multi-site functional MRI datasets [1-3, 8].

Table 2. Performance Comparison in multi-site pooling classification

Methods	Feature	ACC	SPE	SEN
RF	FNC	77.1(0.3)**	74.0(0.6)**	80.1(0.5)**
AdaBoost	FNC	75.7(0.1)**	75.4(0.1)**	76.1(0.1)**
SVM	FNC	82.2(0.3)**	80.5(0.6)**	83.9(0.5)**
DNN	FNC	80.5(0.3)**	79.6(1.2)**	81.3(0.7)**
C-RNN^{AM}	TCs	83.4(0.6)**	81.5(0.9)**	85.3(1.0)**
S+D	FNC	82.8(0.2)**	80.6(0.4)**	84.8(0.3)**
S+C	FNC+TCs	84.5(0.2)*	82.0(0.7)	86.8(0.4)
D+C	FNC+TCs	84.7(0.6)	82.4(1.1)	86.9(0.6)
HDLFCD	**FNC+TCs**	**84.9(0.2)**	**82.7(0.8)**	**87.0(0.7)**

S+D: SVM+DNN; S+C: SVM+C-RNN^{AM}; D+C: DNN+C-RNN^{AM}; */** shows the methods are significantly worse than the proposed methods with P value=0.05/0.01.

3.2. Comparison of methods with/without attention

To evaluate the effectiveness of the proposed attention module, C-RNN without AM and C-RNN^{AM} was respectively performed for SZ diagnosis. The parameter reflects model complexity. As shown in Table 2, the C-RNN^{AM} was generally superior to the C-RNN in all metrics with an improvement about 1%. Experimental results proved that C-RNN^{AM} not only identified the discriminative functional networks but also improved the classification performance. The average-pooling reserved global temporal statistics and max-pooling focused on the most salient part, which were both useful to detect significant components. Besides, only eight parameters were added in C-RNN^{AM}, which means the computational complexity was not increased nearly.

Table 3. Performance comparison of methods with/without attention based on C-RNN framework.

Methods	Parameter	ACC	SPE	SEN
C-RNN	35682	82.6(0.4)*	81.0(0.9)	84.2(0.8)*
C-RNN^{AM}	**35690**	**83.4(0.6)**	**81.5(0.9)**	**85.3(1.0)**

*/** shows the methods are significantly worse than the proposed methods with P value=0.05/0.01.

3.3. Attention map for biomarker exploration

The AM was introduced to focus on SZ-associated components and suppress useless ones, resulting in an attention map. The greater the weight of the attention map, the more important the component was. The spatial maps of top four components were shown in Figure 4.

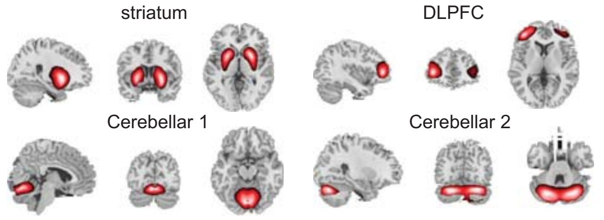

Fig. 4. Spatial map of top 4 selected HC-SZ discriminative components

The results showed that the brain regions identified by AM were mainly concentrated in the striatum, cerebellum and dorsolateral prefrontal cortex (DLPFC). The dorsal striatum, has been proved to play a vital role in the pathophysiology of schizophrenia. The dopaminergic hyperfunction in the striatum contributes to cognitive deficits in SZ, and the popular antipsychotics usually blocks the dopamine D2 receptors in the striatum to achieve a good treatment response [9]. The other two highlighted components were located in the cerebellum. Many studies showed significant evidence for cerebellar abnormalities in SZ during cognition tasks [10]. The DLPFC has been shown to be important for working memory (WM), a key part of cognition. Numerous neuroimaging studies have reported aberrant DLPFC activation during WM performance.

Overall, the most discriminative brain regions were consistent with and extended previous studies, implying that the proposed AM can effectively extract useful information and ignore irrelevant components.

4. CONCLUSION

In this study, we proposed a framework that characterizes temporal coherence between brain regions and temporal dynamics of brain activity jointly to distinguish SZ from HCs. Experimental results showed the superiority of combining multiple features. To the best of our knowledge, this is the first attempt to introduce an attention mechanism in a C-RNN based framework to identify discriminative imaging biomarkers and improve the classification performance by learning which regions to emphasize or suppress, without any increase in model complexity. It should be noted that the AM was an end-to-end module and trainable along with other modules, which can be integrated into other architectures.

5. COMPLIANCE WITH ETHICAL STANDARDS

This work has been approved by relevant ethical committees. All participants signed written informed consents.

6. ACKNOWLEDGMENTS

This work was supported by The National Key Research and Development Program of China (2017YFB1002502), China Natural Science Foundation (82022035, 61773380), Beijing municipal science and technology commission (Z181100001518005), and NIH grants (R01MH117107, R01MH118695).

7. REFERENCES

[1] J. Kim, V. D. Calhoun, E. Shim, and J.-H. J. N. Lee, "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," vol. 124, pp. 127-146, 2016.

[2] L.-L. Zeng *et al.*, "Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI," *EBioMedicine,* vol. 30, pp. 74-85, 2018.

[3] W. Yan *et al.*, "Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site FMRI data," vol. 47, pp. 543-552, 2019.

[4] V. Mnih, N. Heess, and A. Graves, "Recurrent models of visual attention," in *Advances in neural information processing systems*, 2014, pp. 2204-2212.

[5] D. Jin *et al.*, "Generalizable, Reproducible, and Neuroscientifically Interpretable Imaging Biomarkers for Alzheimer's Disease," *Advanced Science,* p. 2000675, 2020.

[6] C. Lian, M. Liu, Y. Pan, and D. J. I. T. o. C. Shen, "Attention-Guided Hybrid Network for Dementia Diagnosis With Structural MR Images," *IEEE Transactions on Cybernetics,* 2020.

[7] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3-19.

[8] M. S. Cetin *et al.*, "Multimodal classification of schizophrenia patients with MEG and fMRI data using static and dynamic connectivity measures," *Frontiers in neuroscience,* vol. 10, p. 466, 2016.

[9] E. H. Simpson, C. Kellendonk, and E. J. N. Kandel, "A possible role for the striatum in the pathogenesis of the cognitive symptoms of schizophrenia," *Neuron,* vol. 65, no. 5, pp. 585-596, 2010.

[10] N. C. Andreasen and R. J. B. p. Pierson, "The role of the cerebellum in schizophrenia," *Biological psychiatry,* vol. 64, no. 2, pp. 81-88, 2008.