Learning Asynchronous Common and Individual Functional Brain Network for AD Diagnosis

Xiang Tang[1], Xiaocai Zhang[2], Mengting Liu[1], and Jianjia Zhang[1(✉)]

[1] School of Biomedical Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen 518107, China
zhangjj225@mail.sysu.edu.cn

[2] Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore 138632, Singapore

Abstract. Construction and analysis of functional brain network (FBN) with rs-fMRI is a promising method to diagnose functional brain diseases. Traditional methods usually construct FBNs at the individual level for feature extraction and classification. There are several issues with these approaches. Firstly, due to the unpredictable interferences of noises and artifacts in rs-fMRI, these individual-level FBNs have large variability, leading to instability and unsatisfactory diagnosis accuracy. Secondly, the construction and analysis of FBNs are conducted in two successive steps without negotiation with or joint alignment for the target task. In this case, the two steps may not cooperate well. To address these issues, we propose to learn common and individual FBNs adaptively within the Transformer framework. The common FBN is shared, and it would regularize the FBN construction as prior knowledge, alleviating the variability and enabling the network to focus on these disease-specific individual functional connectivities (FCs). Both the common and individual FBNs are built by specially designed modules, whose parameters are jointly optimized with the rest of the network for FBN analysis in an end-to-end manner, improving the flexibility and discriminability of the model. Another limitation of the current methods is that the FCs are only measured with synchronous rs-fMRI signals of brain regions and ignore their possible asynchronous functional interactions. To better capture the actual FCs, the rs-fMRI signals are divided into short segments to enable modeling cross-spatiotemporal interactions. The superior performance of the proposed method is consistently demonstrated in early AD diagnosis tasks on ADNI2 and ADNI3 data sets.

Keywords: Functional brain network · Alzheimer's disease · rs-fMRI · Computer-aided diagnosis · Transformer

1 Introduction

Building and analyzing functional brain network (FBN) based on resting-state magnetic resonance imaging (rs-fMRI) have become a promising approach to

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14227, pp. 215–225, 2023.
https://doi.org/10.1007/978-3-031-43993-3_21

functional brain disease diagnosis, e.g., Alzheimer's disease (AD) and Parkinson's disease. Rs-fMRI probes neural activity by the fluctuations in the blood-oxygen-level-dependent (BOLD) signals. The strength of functional connectivities (FCs) between brain regions is measured by the correlation between pairwise BOLD signals. FBN represents the interaction patterns between brain regions during brain functioning and can be used to identify abnormal changes caused by brain diseases [9,28,32]. Such abnormal alterations can be used as diagnostic biomarkers of brain diseases [3,11,14,16,21,22].

Traditional FBNs are constructed at the individual level using Pearson's correlation [26]or sparse representation (SR) [12]. Although this approach is popular due to its simplicity and efficiency, it suffers from multiple limitations. Firstly, due to the unpredictable interferences of noises and artifacts, a potential issue is that the constructing of FBNs at the individual level inevitably leads to large variability in the topographical structure of the networks. So the subtle disease-specific FCs change in the FBNs of patients in comparison with the healthy are likely to be overwhelmed by such large variability [31]. Such variability results in unsatisfactory diagnosis accuracy and poor generality for the classifier. Another issue with the traditional FBN-based methods is that the construction and analysis of FBNs are conducted in two successive steps. Such a two-step approach may not be optimal considering the lack of communication between the two steps. Specifically, the FBNs constructed in the first step do not necessarily work well with the feature extraction methods in the second step since there are no interactions and there is no aligned objective towards solving the target task.

It has been demonstrated that the construction of the common FBN at population level as prior knowledge is beneficial to reduce the effects of the first issue of large variability aforementioned [24]. For example, G. Varoquaux [24] adds a common sparse structure across all subjects as the prior knowledge to construct group consistent individual FBN. Many methods based on Group Sparse Representation (GSR) [15,27] estimate the FBN for all subjects using Group Lasso with $l_{1,2}$-norm constraint to alleviate the variability. These results indicate that a typical whole FBN can be decomposed into a common and an individual FBN components. The common FBN component is shared by all subjects and can be used as prior knowledge regularization to increase the stability of FBN for a subject and facilitate the detection of disease-specific FCs. Inspired by this finding, we propose to learn common and individual FBNs adaptively within Transformer framework to increase the model's stability and efficacy. Unlike the traditional methods, both the common and individual FBNs in the proposed method are built by specially designed modules, whose parameters are jointly optimized within the FBN analysis modules of the network in an end-to-end manner, and this could improve the model's flexibility and discriminability.

Another equally important issue with the existing methods is that the current measures of FC strength only take the synchronous functional interaction into account and ignore the potential asynchronous interactions, as illustrated in Fig. 1(a). Specifically, the current FC measures, e.g., correlation or SR, implicitly assume that the activities of different brain regions start exactly at the same

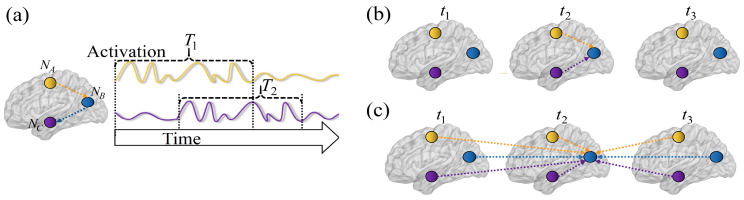

Fig. 1. Illustration of asynchronous functional connectivity. (a) Time-lagged response of N_C in T_2 for N_A' activity in T_1. Direct calculation of the correlation between their BOLD signals during T_1 cannot reflect their actual functional connectivity. (b) Traditional measures of synchronous correlation. (c) The proposed asynchronous cross-spatiotemporal correlation measure.

time, as shown in Fig. 1(b). However, this is not necessarily always the actual case in brain functioning, and certain brain regions may not be synchronized. The time-lagged effects could originate from two-fold: on the one hand, the information flow within the brain takes time, which would cause asynchronous rs-fMRI observations [7,19]; on the other hand, even when various brain regions get activated at the same time, their individual activities may have different patterns and durations, which will also lead to asynchronous time-lagged interactions [4,5]. In this case, only measuring the FCs synchronicity cannot fully capture the underlying functional interactions between brain regions, and this would degrade the model performance. To reflect the asynchronous FCs, we propose to divide the whole BOLD signal into short-term segments with the sliding windows [20]. Then their asynchronous interactions are measured in a specially designed attention module in the Transformer framework, as shown in Fig. 1(c).

To validate the effectiveness of the proposed method, we conducted experimental studies on rs-fMRI data sets of ADNI2 and ADNI3 for early AD diagnosis, i.e., classification of mild cognitive impairment (MCI) and normal control (NC). The experimental results show that the proposed method consistently achieves state-of-the-art performance.

2 Proposed Method

The overall network of our method is shown in Fig. 2(a). Firstly, the raw rs-fMRI BOLD signal $\boldsymbol{X}_{raw} \in R^{N\times T\times D_{raw}}$ is transformed into $\boldsymbol{X}_0 \in R^{N\times T\times D_0}$ by applying 1D temporal convolution for input mapping, where N denotes the number of brain nodes, T denotes frames, and D denotes the feature dimension. Then, before feeding into the L-th Transformer encoder, $\boldsymbol{X}_0$ needs to be position encoded for incorporating node index as spatial identity. The position encoding matrix $\boldsymbol{P}$ is defined as: $\boldsymbol{P}(pos,:,2dim) = sin(pos/10000^{2dim/D})$; $\boldsymbol{P}(pos,:,2dim+1) = cos(pos/10000^{2dim/D})$, where $pos \in [0,\cdots,N-1]$ is the node index, the symbol ':' refers to all indexes in the corresponding dimension, and $dim \in [0,\cdots,\frac{D}{2}-1]$ indicates the index of the feature dimension. $\boldsymbol{P}$ is added to $\boldsymbol{X}_0$, i.e., $\boldsymbol{X_0} = \boldsymbol{X}_0 + \boldsymbol{P}$. If we measure the interactions between brain

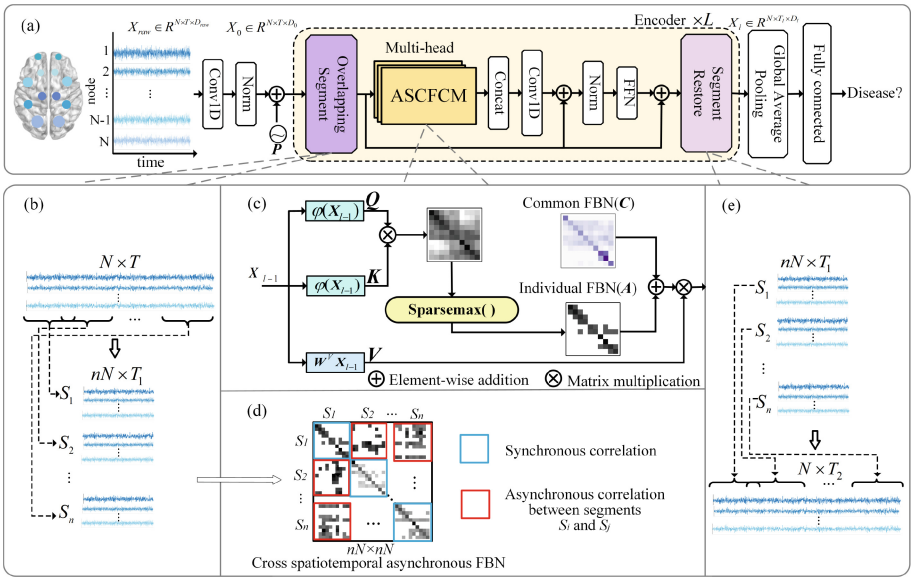

Fig. 2. Illustration of the proposed method. (a) Overall framework. (b) Dividing rs-fMRI signals into segments by sliding window. (c) Attention-based Sparse Common-and-individual FBN Construction Module (ASCFCM). (d) Asynchronous cross spatiotemporal FCs. (e) Restore the rs-fMRI segments.

regions directly, we can obtain synchronous FCs. In order to consider the time-lagged interaction between brain regions, we use a sliding window to segment the signals into short-time segments, the details are illustrated in Fig. 2(b) and introduced in detail later. Then, an improved version of the multi-head self-attention module is proposed to measure the dependencies between brain regions to obtain FBN. Conceptually, each encoder module can be formulated as a transformation function as:

$$\boldsymbol{X}_l = f_l(a_l(\boldsymbol{X}_{l-1}), \boldsymbol{X}_{l-1}), l \in 1, \cdots, L. \tag{1}$$

where $f_l(\cdot)$ denotes the transformation function of the l-th encoder layer with input from $\boldsymbol{X}_{l-1}$, which is the output of the (l-1)-th layer, and $a_l(\cdot)$ denotes the attention module to construct FBN. The details of $a_l(\cdot)$ are illustrated in Fig. 2(c) and introduced later. The final feature $\boldsymbol{X}_L$ of the L-th layer is fed into the fully connected layer for classification.

2.1 Attention-Based Sparse Common-and-Individual FBN Construction Module (ASCFCM)

In order to respect the sparsity property of the human brain and introduce population-level constraint, we redesign the kernel self-attention module in [30] and propose an attention-based sparse common-and-individual FBN construction module (ASCFCM), which will be introduced in detail below.

Revisiting Kernel Self-attention Mechanism: Here, we first review the kernel attention module proposed in [30] and then introduce how it is redesigned to fit the target task. In order to measure the FCs between N nodes, the l-th layer takes $\boldsymbol{X}_{l-1} \in R^{N\times T\times D_{l-1}}$ as input data and generates queries $\boldsymbol{Q}$, keys $\boldsymbol{K}$, and values $\boldsymbol{V}$ through the kernel embedding $\varphi(\cdot)$ and linear projection matrix $\boldsymbol{W}^V \in R^{D_{l-1}\times D_v}$, i.e., $\boldsymbol{Q} = \varphi(\boldsymbol{X}_{l-1}); \boldsymbol{K} = \varphi(\boldsymbol{X}_{l-1}); \boldsymbol{V} = \boldsymbol{X}_{l-1}\boldsymbol{W}^V$, where D_v is set as D_{l-1}/H, and H is the number of heads. The resulting attention matrix $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{K}^T$ is $N \times N$, and $\boldsymbol{A}_{i,j}$ is actually a Gaussian RBF kernel function [30] with the input features of the i-th and j-th brain regions from $\boldsymbol{X}_{l-1}$, denoted as x_i and x_j, i.e., $\boldsymbol{A}_{i,j} = \langle\varphi(x_i), \varphi(x_j)\rangle = \exp(-\beta\|x_i - x_j\|^2)$. The kernel parameter β can be adaptively learned during the model training process. One advantage of this method over the conventional methods is that it significantly reduces the number of parameters in calculating $\boldsymbol{Q}$ and $\boldsymbol{K}$. Another advantage is that it can effectively model the non-linearity between the features of brain nodes and can better model the complex interactions between brain regions. With the obtained attention matrix $\boldsymbol{A}$, the feature mapping function of the l-th layer $a_l(\cdot)$ can be defined as:

$$a_l(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{softmax}(\frac{\boldsymbol{A}}{\sqrt{TD_v}})\boldsymbol{V} \tag{2}$$

where $\text{softmax}(\frac{\boldsymbol{A}}{\sqrt{TD_v}})$ is a $N \times N$ matrix and it can be viewed as an adaptively built FBN, presenting the dependencies between N brain regions.

Sparsity: From Eq. (2), the original attention scores are transformed by $\text{softmax}(\cdot)$. Due to the non-negative nature, i.e., $\text{softmax}(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$, all the FCs between brain regions are positive, leading to a dense FBN. However, the brain is a sparse network [2], which is characterized by a limited number of FCs for most brain regions. In this case, the resulting FBN built with $\text{softmax}(\cdot)$ does not respect the nature of the human brain, and this may degrade the effectiveness of detecting disease-specific patterns. To address this issue, we propose to build a sparse FBN by using $\text{sparsemax}(\cdot)$ [17]. Let $\Delta^K := \{p \in R^K | 1^T p = 1, p \geq 0\}$ be the K-dimensional probability distribution simplex. The key idea of sparse transformation is mapping an input z into the nearest simplex from Δ^K by $\text{sparsemax}(z) := \underset{p\in\Delta^K}{\text{argmin}} \|p - z\|^2$. The optimization could produce a sparse distribution with proper regularization, as in [17], leading to a sparse FBN.

Population-Level Common Pattern: As mentioned previously, the above FBN has large variability since it is constructed with the rs-fMRI data of each subject. Previous research indicates that human brains have similar topologies, which can be modeled as prior knowledge separately to regularize the model, helping to alleviate the issue of large variability. Therefore, we parameterize a data-driven common FBN module as a regularization, as shown in Fig. 2(c). The resulting common FBN, denoted as $\boldsymbol{C}$, is shared by all the subjects and integrated with FBN at the individual level for joint diagnosis. The final $a_l(\cdot)$ in

Eq. 1 is defined as:

$$a_l(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = (\text{sparsemax}(\frac{\boldsymbol{A}}{\sqrt{TD_v}}) + \boldsymbol{C})\boldsymbol{V} \tag{3}$$

2.2 Cross Spatiotemporal Asynchronous FCs

To simulate the time-lagged asynchronous FCs, we segment the BOLD signals $\boldsymbol{X}_l \in R^{N \times T \times D_l}$ into $\boldsymbol{X}_l \in R^{nN \times T_1 \times D_l}$ by using a sliding window, as shown in Fig. 2(b). The length and number of segments, denoted as T_1 and n, respectively, are determined by the length and steps of the sliding window. Measuring the FCs between these new segments breaks the limitations of synchronous FCs, yielding cross-spatiotemporal interaction patterns, as shown in Fig. 2(d). These segments will be concatenated to restore the original number of nodes, i.e., $nN \rightarrow N$, to avoid exponential expansion of the network complexity (see Fig. 2(e)). By doing this, both the synchronous FCs in the traditional methods and the asynchronous ones targeted in this paper can be simultaneously modeled, providing more clues for disease diagnosis.

3 Experiments

3.1 Data and Preprocessing

Both rs-fMRI data sets of ADNI2 and ADNI3 from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (https://adni.loni.usc.edu/) are used to evaluate the effectiveness of the proposed method with the most challenging task of differentiating MCI from NC. The subject number is 410 (147 MCI vs. 263 NC) for ADNI2 and 425 (168 MCI vs. 257 NC) for ADNI3. The ADNI2 data are acquired on a 3 Tesla (Philips) scanner with TR/TE set at 3000/30 ms and a flip angle of 80°. The ADNI3 data are acquired with a flip angle of 90°. The preprocessing is carried out using DPARSFA toolbox (http://rfmri.org/DPARSF) and SPM-12 (ehttps://www.fil.ion.ucl.ac.uk/spm/software/). We perform a standard approach to processing rs-fMRI by following [6], including discarding the first 10 volumes of each time series, slice timing, head motion correction, and Montreal Neurological Institute spatial normalization. Empirically, the AAL [23] atlas is used to extract all voxels of time series values in the 90 regions of interest (ROI).

3.2 Experimental Settings

We conduct a 5-fold cross validation in our evaluation to obtain stable results, and all the methods involved in the comparison use the same partitions of data for fairness. The performance of the model is measured by four metrics: Area Under the receiver operating characteristic Curve (AUC), accuracy (ACC), sensitivity (SEN), and specificity (SPE). The encoder layers of the proposed network L are set to 5. Asynchronous FCs are explored in the first two layers. The sliding

Table 1. The NC-vs-MCI classification performance by different methods on ADNI2 and ADNI3.

Method	ADNI2				ADNI3			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Compact Representation [29]	66.8 ± 1.7	98.5 ± 0.9	10.2 ± 4.8	54.3 ± 2.7	65.1 ± 4.1	52.3 ± 16.3	73.8 ± 16.2	63.0 ± 1.8
Kernel Transformer [30]	68.3 ± 1.3	98.1 ± 1.7	15.0 ± 6.0	56.5 ± 2.2	69.4 ± 3.7	47.9 ± 6.2	83.3 ± 5.5	65.6 ± 3.7
LCC [10]	69.7 ± 2.9	88.5 ± 10.1	36.1 ± 19.9	62.2 ± 5.9	71.0 ± 3.2	88.8 ± 6.3	43.7 ± 6.0	66.2 ± 3.2
Transformer [25]	70.3 ± 2.6	94.3 ± 2.8	29.7 ± 4.9	60.8 ± 2.7	72.9 ± 3.2	51.6 ± 10.0	86.8 ± 6.2	69.2 ± 3.8
BrainGNN [13]	71.2 ± 5.2	88.6 ± 6.3	41.3 ± 26.6	64.2 ± 9.5	68.4 ± 9.1	41.9 ± 32.6	85.6 ± 6.6	63.8 ± 13.2
STGCN [8]	74.2 ± 2.2	84.6 ± 8.8	48.2 ± 9.8	68.4 ± 3.0	71.3 ± 4.6	84.4 ± 7.4	50.8 ± 18.2	67.7 ± 6.6
BrainNetCNN [11]	78.5 ± 3.3	84.8 ± 5.5	67.3 ± 9.1	74.6 ± 5.8	72.4 ± 2.0	46.8 ± 8.9	89.1 ± 6.2	71.8 ± 2.4
Proposed network	83.2 ± 2.4	91.3 ± 4.4	68.8 ± 9.2	80.0 ± 5.3	78.1 ± 1.8	63.0 ± 7.3	87.9 ± 2.9	75.5 ± 5.6

window of each layer has a length of 32 and a step size of 16 while the number of heads H in each layer is set as 2. The model input $\boldsymbol{X}_{raw} \in R^{90\times128\times1}$ is transformed into $\boldsymbol{X}_0 \in R^{90\times128\times6}$ by a mapping with a kernel size of 3×1. The output feature dimension D_l of each layer is set as 12. The drop rate of the attention score is 0.2, and the batch size is 16. We train 250 epochs with stochastic gradient descent (SGD), whose learning rate is 0.1 and momentum is 0.9. The code is publicly available at https://github.com/seuzjj/ACIFBN.

3.3 Experimental Results

The average performance of the proposed method and the competing methods is summarized in Table 1, and their ROC curves are shown in Fig. 3(a). As seen, our proposed method achieves the highest accuracy compared with other classical methods, i.e., 83.2% on ADNI2 and 78.1% on ADNI3, obtaining an improvement of 4.7% and 5.7%, respectively, over the current state-of-the art methods. Our result increases by 14.9% on ADNI2 over the base model–Kernel Transformer. These results are encouraging, indicating the significant efficacy and stability of our method.

Table 2. Different model components affect classification performance. The backbone structure is Kernel Transformer, and except for the settings available in the table, all other settings are the same.

FC pattern		activation		asynchrony	ACC of ADNI2	ACC of ADNI3
individual	common	softmax	sparsemax			
✓		✓			68.3 ± 1.3	69.4 ± 3.7
	✓				79.2 ± 2.9	76.7 ± 5.1
✓	✓	✓			79.5 ± 4.0	77.1 ± 3.5
✓	✓		✓		82.3 ± 1.1	76.7 ± 1.3
✓	✓		✓	✓	83.2 ± 2.4	78.1 ± 1.8

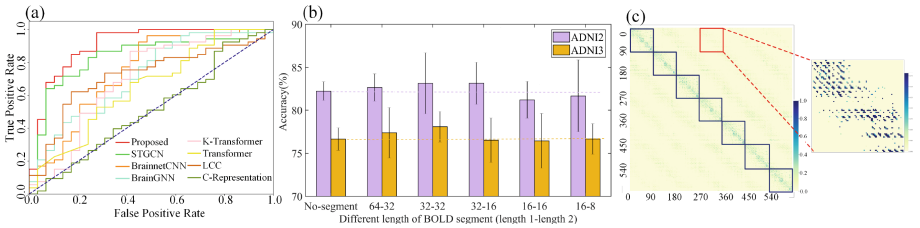

Fig. 3. Visualization of experimental results. (a) ROC on ADNI2. (b) The effect of different T_1. (c) Asynchronous FCs.

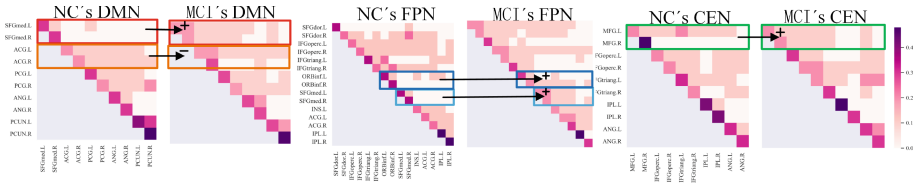

Fig. 4. Difference of DMN, FPN, CEN between NC and MCI.

3.4 Ablation Study

In order to further verify the effectiveness of the proposed method and the contribution of each component, we conduct ablation studies and the results are shown in Table 2. As seen, introducing common FBN improves the accuracy significantly from 68.3% to 79.2%, and this is probably attributed to the efficacy of it as prior knowledge regularizer. On top of it, considering the sparsity property of FBN further increases the accuracy to 82.3%. Modeling asynchronous FCs, denoted as 'asynchrony', further boost the model's accuracy, reaching 83.2%.

To verify the effects of the settings during dividing rs-fMRI signals into segments, we set the window lengths as different values in the two asynchronous FCs modeling layers, seen in Fig. 3(b). The step of windows is half of the length. It can be consistently observed on ADNI2 and ADNI3 that the model performance first improves with the decreasing segment lengths, and then start to decrease. This implies that modeling asynchronous FCs in a wide range is beneficial, however, too short segments would harm the performance due to increased model complexity or weak of temporal information within short segments.

4 Visualization and Conclusion

The learned individual FBNs carry diagnostic clues, which are visualized in Fig. 4 for MCI and NC comparison. As seen, increased connectivities could be identified in the superior frontal gyrus (SFGmed) of the default mode network (DMN) and central executive network (CEN), the middle frontal gyrus (MFG) of the frontal parietal network (FPN), and the part inferior frontal gyrus (ORBinf)

of CEN, while decreased connectivities could be found in the anterior cingulate and paracingulate gyri (ACG) of the DMN. These results are consistent with the literature [1,18]. The effects of modeling asynchronous FCs are shown in Fig. 3(c). As seen, the dominant FCs are mainly synchronous from the diagonal blocks (inside the black block), while certain asynchronous FCs (inside the red block) are also well captured. This is reasonable since most brain regions interact synchronously, while a portion of them would have asynchronous time-lagged functional responses.

In sum, this paper proposes to adaptively construct and analyze asynchronous common and individual functional brain networks within a specially designed Transformer-based network, which is more flexible, discriminative and capable of capturing both synchronous and asynchronous FCs. The experimental results on two ADNI data sets well demonstrate the effectiveness of the proposed method.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (grant number 62101611), Guangdong Basic and Applied Basic Research Foundation (grant number 2022A1515011375, 2023A1515012278) and Shenzhen Science and Technology Program (grant number JCYJ20220530145411027, JCYJ20220818102414031).

References

1. Ahmadi, H., Fatemizadeh, E., Motie-Nasrabadi, A.: Identifying brain functional connectivity alterations during different stages of Alzheimer's disease. Int. J. Neurosci. **132**(10), 1005–1013 (2022)
2. Bassett, D.S., Bullmore, E.T.: Small-world brain networks revisited. Neuroscientist **23**(5), 499–516 (2017)
3. Chen, H., Zhang, Y., Zhang, L., Qiao, L., Shen, D.: Estimating brain functional networks based on adaptively-weighted fMRI signals for MCI identification. Front. Aging Neurosci. **12**, 595322 (2021)
4. Deshpande, G., Santhanam, P., Hu, X.: Instantaneous and causal connectivity in resting state brain networks derived from functional MRI data. Neuroimage **54**(2), 1043–1052 (2011)
5. Deshpande, G., Sathian, K., Hu, X.: Assessing and compensating for zero-lag correlation effects in time-lagged Granger causality analysis of fMRI. IEEE Trans. Biomed. Eng. **57**(6), 1446–1456 (2010)
6. Esteban, O., et al.: fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods **16**(1), 111–116 (2019)
7. Friston, K., Moran, R., Seth, A.K.: Analysing connectivity with Granger causality and dynamic causal modelling. Curr. Opin. Neurobiol. **23**(2), 172–178 (2013)
8. Gadgil, S., Zhao, Q., Pfefferbaum, A., Sullivan, E.V., Adeli, E., Pohl, K.M.: Spatio-temporal graph convolution for resting-state fMRI analysis. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12267, pp. 528–538. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59728-3_52
9. Ghanbari, M., et al.: Alterations of dynamic redundancy of functional brain subnetworks in Alzheimer's disease and major depression disorders. NeuroImage Clin. **33**, 102917 (2022)

10. Kaiser, M.: A tutorial in connectome analysis: topological and spatial features of brain networks. Neuroimage **57**(3), 892–907 (2011)
11. Kawahara, J., et al.: BrainNetCNN: convolutional neural networks for brain networks; towards predicting neurodevelopment. Neuroimage **146**, 1038–1049 (2017)
12. Lee, H., Lee, D.S., Kang, H., Kim, B.N., Chung, M.K.: Sparse brain network recovery under compressed sensing. IEEE Trans. Med. Imaging **30**(5), 1154–1165 (2011)
13. Li, X., et al.: Braingnn: interpretable brain graph neural network for fMRI analysis. Med. Image Anal. **74**, 102233 (2021)
14. Li, Y., Liu, J., Tang, Z., Lei, B.: Deep spatial-temporal feature fusion from adaptive dynamic functional connectivity for MCI identification. IEEE Trans. Med. Imaging **39**(9), 2818–2830 (2020)
15. Li, Y., Yu, Z.L., Bi, N., Xu, Y., Gu, Z., Amari, S.I.: Sparse representation for brain signal processing: a tutorial on methods and applications. IEEE Signal Process. Mag. **31**(3), 96–106 (2014)
16. Liu, M., Zhang, H., Shi, F., Shen, D.: Building dynamic hierarchical brain networks and capturing transient meta-states for early mild cognitive impairment diagnosis. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12907, pp. 574–583. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87234-2_54
17. Martins, A., Astudillo, R.: From softmax to sparsemax: a sparse model of attention and multi-label classification. In: International Conference on Machine Learning, pp. 1614–1623. PMLR (2016)
18. Metmer, H., Lu, J., Zhao, Q., Li, W., Lu, H.: Evaluating functional connectivity of executive control network and frontoparietal network in Alzheimer's disease (2013)
19. Mitra, A., Snyder, A.Z., Hacker, C.D., Raichle, M.E.: Lag structure in resting-state fMRI. J. Neurophysiol. **111**(11), 2374–2391 (2014)
20. Qiu, H., Hou, B., Ren, B., Zhang, X.: Spatio-temporal tuples transformer for skeleton-based action recognition. arXiv preprint arXiv:2201.02849 (2022)
21. Riaz, A., et al.: FCNet: a convolutional neural network for calculating functional connectivity from functional MRI. In: Wu, G., Laurienti, P., Bonilha, L., Munsell, B.C. (eds.) CNI 2017. LNCS, vol. 10511, pp. 70–78. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67159-8_9
22. Shi, Y., et al.: ASMFS: adaptive-similarity-based multi-modality feature selection for classification of Alzheimer's disease. Pattern Recogn. **126**, 108566 (2022)
23. Tzourio-Mazoyer, N., et al.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage **15**(1), 273–289 (2002)
24. Varoquaux, G., Gramfort, A., Poline, J.B., Thirion, B.: Brain covariance selection: better individual functional connectivity models using population prior. In: Advances in Neural Information Processing Systems, vol. 23 (2010)
25. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
26. Wang, J., et al.: Disrupted functional brain connectome in individuals at risk for Alzheimer's disease. Biol. Psychiat. **73**(5), 472–481 (2013)
27. Wee, C.Y., Yap, P.T., Zhang, D., Wang, L., Shen, D.: Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. Brain Struct. Funct. **219**, 641–656 (2014)
28. Zhang, H.Y., et al.: Detection of PCC functional connectivity characteristics in resting-state fMRI in mild Alzheimer's disease. Behav. Brain Res. **197**(1), 103–108 (2009)

29. Zhang, J., Zhou, L., Wang, L., Li, W.: Functional brain network classification with compact representation of SICE matrices. IEEE Trans. Biomed. Eng. **62**(6), 1623–1634 (2015)
30. Zhang, J., Zhou, L., Wang, L., Liu, M., Shen, D.: Diffusion kernel attention network for brain disorder classification. IEEE Trans. Med. Imaging **41**(10), 2814–2827 (2022)
31. Zhang, Y., et al.: Strength and similarity guided group-level brain functional network construction for MCI diagnosis. Pattern Recogn. **88**, 421–430 (2019)
32. Zhu, X., Cortes, C.R., Mathur, K., Tomasi, D., Momenan, R.: Model-free functional connectivity and impulsivity correlates of alcohol dependence: a resting-state study. Addict. Biol. **22**(1), 206–217 (2017)