Affinity Learning Based Brain Function Representation for Disease Diagnosis

Mengjun Liu[1,2], Zhiyun Song[1], Dongdong Chen[1], Xin Wang[1], Zixu Zhuang[1], Manman Fei[1], Lichi Zhang[1,3(✉)], and Qian Wang[2,4(✉)]

[1] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
lichizhang@sjtu.edu.cn
[2] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China
[3] National Engineering Research Center of Advanced Magnetic Resonance Technologies for Diagnosis and Therapy (NERC-AMRT), Shanghai Jiao Tong University, Shanghai, China
[4] Shanghai Clinical Research and Trial Center, Shanghai, China
qianwang@shanghaitech.edu.cn

Abstract. Resting-state functional magnetic resonance imaging (rs-fMRI) serves as a potent means to quantify brain functional connectivity (FC), which holds potential in diagnosing diseases. However, conventional FC measures may fall short in encapsulating the intricate functional dynamics of the brain; for instance, FC computed via Pearson correlation merely captures linear statistical dependencies among signals from different brain regions. In this study, we propose an affinity learning framework for modeling FC, leveraging a pre-training model to discern informative function representation among brain regions. Specifically, we employ randomly sampled patches and encode them to generate region embeddings, which are subsequently utilized by the proposed affinity learning module to deduce function representation between any pair of regions via an affinity encoder and a signal reconstruction decoder. Moreover, we integrate supervision from large language model (LLM) to incorporate prior brain function knowledge. We evaluate the efficacy of our framework across two datasets. The results from downstream brain disease diagnosis tasks underscore the effectiveness and generalizability of the acquired function representation. In summary, our approach furnishes a novel perspective on brain function representation in connectomics. Our code is available at https://github.com/mjliu2020/ALBFR.

Keywords: Affinity learning · Brain function representation · rs-fMRI · Brain disease diagnosis

1 Introduction

Resting-state functional magnetic resonance imaging (rs-fMRI) is a valuable tool in the diagnosis of brain diseases, particularly via measurement of brain functional connectivity (FC) [3, 17, 22]. The rs-fMRI records the spontaneous blood oxygen-level dependent (BOLD) fluctuations during rest [16]. The FC is typically defined as the statistical dependencies between two neurophysiological signals from different parts of the brain [2]. The deviations in FC indicate underlying dysfunction associated with diverse brain diseases [6].

In recent years, graph methods are prevalent to analyze brain FC for disease diagnosis [4,5,9]. Specifically, the brain can be conceptualized as a graph, wherein nodes correspond to distinct brain regions or regions of interest (ROIs), and edges denote the FC between these regions [8]. Therefore, the partition of brain regions and the measurement of FC become particularly important. Conventionally, the FC is assessed through Pearson correlation coefficient (PCC) [12]. However, the pre-defined linear correlation may be insufficient to characterize the complex connectivity patterns between brain regions.

With the development of deep learning techniques, many studies have employed deep networks to establish FC from BOLD signals. For instance, Zhang *et al.* [20] designed a non-linear attention mechanism to model complex function representation and successfully realized the diagnosis of brain diseases. However, it relies on training a supervised model to generate FC, which necessitates retraining from scratch when applied to new datasets. Zhang *et al.* [21] proposed an unsupervised graph structure learning method for capturing characteristics of functional brain network. However, they utilize PCC-based FC as supervision, resulting in suboptimal performance. Moreover, the aforementioned deep learning approaches rely on fixed brain region definitions, thereby limiting the universality of well-trained models across diverse brain parcellations. Recently, Liu *et al.* [10] has demonstrated that random sampling rather than atlas-fixed brain regions enables a more flexible brain representation.

To generate more informative FC, we propose an affinity learning based brain function representation framework. Benefiting from the inherent powerful learning capabilities of deep learning, the affinity learning model can encode significantly richer brain functions than PCC. Specifically, we adopt randomly sampled patches to improve the generality of our method in brain parcellation. These patches are encoded as region embeddings. Next, we present an affinity learning module for generating function representations in connectomics upon them. In addition, we introduce brain function knowledge from large language model (LLM) to guide the affinity learning. Ultimately, we employ the learned function representation to diagnose brain diseases. We validate our approach on ABIDE and ADNI datasets. The experimental results indicate that our proposed method not only outperforms the comparison methods but also provides a new way of function representation.

Our main contributions are summarized as follows. (1) We propose an affinity learning model for whole brain function representation in connectomics. (2) We integrate the knowledge of LLM to guide the affinity learning, and introduce this strategy for the first time in brain function representation. (3) We provide a pre-trained model to produce function representation, which improves the performance of brain disease diagnosis. Furthermore, the pre-trained model exhibits transferability across a spectrum of diseases and diverse brain parcellations.

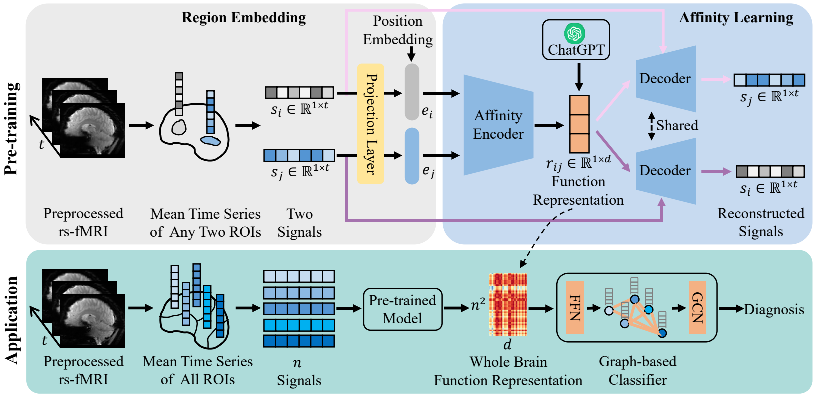


Fig. 1. The architecture of the proposed affinity learning based brain function representation method and its application in the diagnosis of brain diseases.

2 Methodology

The architecture of the proposed method is illustrated in Fig. 1. We randomly sample patches and encode them as region embeddings. Next, the affinity learning module generates the function representation through an encoder-decoder structure. Furthermore, we utilize the knowledge from LLM to guide the affinity learning, namely LLM-guided supervision. In applications, the pre-trained model is employed to generate function representation as a graph, which can effectively diagnose brain diseases, i.e., by graph-based classification.

2.1 Region Embedding

We design a region embedding strategy based on randomly sampling patches. It includes patch extraction and patch encoding, and finally obtains the region embedding for the subsequent affinity learning.

Patch Extraction. Random patch sampling is not constrained by the fixed brain parcellation of atlases. For each subject, 3D patches are randomly sampled in gray matter. The average signal of the gray matter within each patch serves as the functional signal $\boldsymbol{s} \in \mathbb{R}^{1\times t}$ for that region. The signal is normalized to zero mean and one standard deviation as done in [23]. We choose the center coordinate $\boldsymbol{p}$ of each region in the MNI space to depict its position in the brain. The size and number of patches are $9\times9\times9$ and 256 respectively, as used in [10]. In this way, each subject can provide $C_{256}^2 = 32640$ combinations of brain region signals for the pre-training model.

Patch Encoding. We design projection layer to learn low-dimensional BOLD signal features that reflect brain function in a more compact latent space. At the same time, we adopt learnable linear transformation matrix $\boldsymbol{W}$ to map the position coordinates of the patch to the position embeddings and integrate them with the signal features to the region embeddings $\boldsymbol{e}$:

$$\boldsymbol{e} = FFN(Conv(\boldsymbol{s})) + \boldsymbol{Wp}, \tag{1}$$

where $Conv$ represents one dimensional convolution, followed by a Feedforward Network (FFN) to obtain signal feature of size 64. The one dimensional convolution consists of four distinct convolution layers characterized by parameters as in [11]: filter size of (4, 4, 3, 1), stride of (2, 1, 2, 1), and output channel of (32, 64, 64, 10). The activation layer is rectified linear unit (ReLU) inserted between layers.

2.2 Affinity Learning Module

The affinity learning module consists of affinity encoder and signal reconstruction decoder. The former learns the signal similarity between any two regions, which describes brain function in connectomics. The latter reconstructs the target signal based on the learned similarity with reference to the other signal.

Affinity Encoder. First, a shared learnable transformation matrix $\boldsymbol{W}$ is applied to a pair of region embeddings $(\boldsymbol{e_i}, \boldsymbol{e_j})$ to obtain sufficient expressive power. Next, we calculate the similarity as follows:

$$\boldsymbol{r_{ij}} = tanh(FFN(\boldsymbol{We_i} \| \boldsymbol{We_j})), \tag{2}$$

where $\|$ represents the concatenation operation in the feature dimension. The similarity $\boldsymbol{r_{ij}} \in \mathbb{R}^{1 \times d}$ is the output of the affinity encoder, which serves as function representation between regions. The symbol d denotes the dimension of learned function representation.

Signal Reconstruction Decoder. The decoder reconstructs the target signal $\boldsymbol{s_{rec}}$ based on the learned function representation $\boldsymbol{r_{ij}}$ and the reference signal $\boldsymbol{s_{ref}}$. Initially, the learned function representation is aligned with the reference signal in dimension via a Feedforward Network, then added to the reference signal, and subsequently utilized for target signal reconstruction:

$$\boldsymbol{s_{rec}} = FFN(FFN(\boldsymbol{r_{ij}}) + \boldsymbol{s_{ref}}). \tag{3}$$

It is noteworthy that in the model training, a pair of signals are mutually referenced to reconstruct the other signal.

The affinity learning model is subjected to self-supervised optimization, aiming to minimize the Mean Square Error (MSE) between the original fMRI signals and their corresponding reconstructions, denoted as L_{rec}.

2.3 LLM-guided Supervision

To enhance the elucidation, we introduce LLM-guided supervision to refine the learned function representations. Presently, ChatGPT is acknowledged as an LLM endowed with profound knowledge reservoirs [1]. A compendium of brain regions linked to the Precentral_L can be obtained by querying ChatGPT with the *prompt*: "Which brain regions are affiliated with the Precentral_L in the Automated Anatomical Labeling (AAL) atlas?" Subsequently, we establish the correlation between the Precentral_L and the brain regions identified by ChatGPT as associated with Precentral_L, denoted mathematically as 1. In contrast, the association between Precentral_L and the remainder of the brain is designated as 0. Following a systematic execution of the aforementioned procedures for all brain regions delineated in the AAL atlas, we derive a LLM-guided supervision matrix $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ $(m = 116)$.

During the optimization of the affinity learning model, the learned function representation passes through an FFN and computes an MSE loss with the LLM-guided supervision matrix $\boldsymbol{U}$:

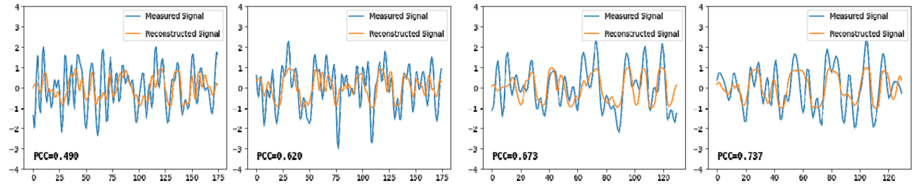
$$L_{llm} = sim(FFN(\boldsymbol{r_{ij}}), \boldsymbol{U_{ij}}), \tag{4}$$

where $sim()$ is an MSE measure. The $\boldsymbol{r_{ij}}$ represents the learned function representation between region i and j. Notably, the randomly sampled patch is designated as a specific brain region in AAL according to the maximum overlap first principle. We incorporate prior knowledge to guide the affinity learning, thereby ensuring outcomes reflect brain function.

The ultimate objective function is formulated as $L = L_{rec} + \alpha L_{llm}$, where α is the trade-off hyperparameter.

2.4 Brain Disease Diagnosis

We introduce a brain disease diagnosis task to assess the acquired function representation. The diagnosis process involves two stages. Initially, we conduct pre-training of an affinity learning model. Subsequently, in the second stage, we freeze the pre-trained model and derive the function representation for whole brain. As illustrated in Fig. 1, the representation is compressed by an FFN to yield an $n \times n$ adjacency matrix, where each row of the matrix constitutes the node features, and the matrix elements denote the weights of graph edges. The graph is then fed into a classifier comprising three graph convolutional network (GCN) layers with ReLU activation functions, respectively. We minimize the cross-entropy loss between predictions and labels to optimize the classifier. This strategy ensures that the acquired function representation remains agnostic to specific tasks, allowing for a fair evaluation of its effectiveness and generalizability.



(a) Signal Reconstruction on ABDIE Dataset

(b) Signal Reconstruction on ADNI Dataset

Fig. 2. Examples of reconstructed and measured signals, together with the Pearson correlation between them on two datasets.

3 Experimental Results

3.1 Data Preprocessing and Experimental Settings

We evaluate our method on two brain diseases, including the autism spectrum disorder (ASD) and the mild cognitive impairment (MCI). For the ASD diagnosis task, the dataset is from the publicly available Autism Brain Imaging Data Exchange I (ABIDE I). We employ the largest site (NYU), which includes preprocessed rs-fMRI of 170 subjects (73 ASDs and 97 NCs). We download the preprocessed data with configurations as using DPARSF toolbox [18], band-pass filtering (0.01-0.1Hz), and without global signal regression. For the MCI diagnosis task, we utilize the publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, focusing on data from the six sites with the largest cohort sizes. A total of 333 rs-fMRI samples are selected from 148 distinct subjects, comprising 60 MCIs and 88 NCs. The rs-fMRI data undergoes standardized preprocessing, which is available online.

Our framework is implemented with PyTorch on a single GPU for both tasks. The data is split into training, validation, and test set according to 3:1:1 for pre-training. The average performance of the subject in the test set is reported. The PCC is adopted to evaluate the performance of the pre-training task [15]. The classification performance is evaluated using four metrics: accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC).

3.2 Signal Reconstruction

The pre-training task involves signal reconstruction in brain regions. The mean PCC between reconstructed and measured signals in the test set reaches 0.539 and 0.690 respectively on the ABIDE and ADNI datasets. The reconstructed signals, showcasing varying performance, are illustrated in Fig. 2. It is evident that the model adeptly predicts numerous characteristics of the original signal.

3.3 Classification Results

To verify the effectiveness of our proposed approach, we compare with the following methods in the diagnosis of ASD and MCI. The FC calculated by

Table 1. Diagnosis results of ASD and MCI. The term L_{llm} is for the LLM-guided supervision.

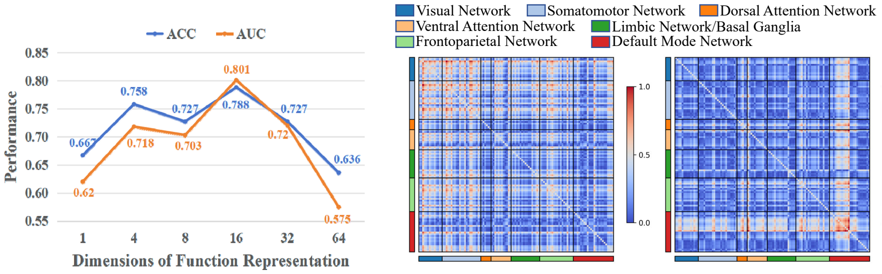
Methods	ASD vs. HC				MCI vs. HC			
	ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE
PCC-FC + MLP	0.727	0.699	0.842	0.707	0.672	0.667	0.649	0.679
BrainNetCNN [7]	0.697	0.699	0.842	0.671	0.705	0.690	0.730	0.698
BrainGNN [9]	0.697	0.643	**0.947**	0.652	0.738	0.660	0.838	0.711
KD-Transformer [20]	0.758	0.703	0.895	0.733	0.738	**0.752**	0.892	0.696
BrainUSL [21]	0.581	0.525	0.901	0.148	0.707	0.691	0.806	0.577
Ours (w/o L_{llm})	0.758	0.771	**0.947**	0.724	0.705	0.662	**0.946**	0.640
Ours	**0.788**	**0.801**	0.895	**0.769**	**0.754**	0.747	0.892	**0.717**
Ours (ASD→MCI)	-	-	-	-	**0.754**	0.669	0.919	0.710
Ours (MCI→ASD)	0.758	0.711	**0.947**	0.724	-	-	-	-

conventional PCC is followed by multilayer perceptron (MLP) as a classifier. BrainNetCNN [7] is convolutional neural networks for brain network analysis. BrainGNN [9] is a graph neural network (GNN) framework tailored for fMRI data and brain disorder diagnosis. KD-Transformer [20] is a Transformer-based FC modeling and analysis framework for brain disease diagnosis. BrainUSL [21] is an unsupervised graph structure learning framework for brain network analysis.

Comparison with SOTA. As shown in Table 1, our results are superior to the comparison methods. For the diagnosis of ASD, our method achieves the best ACC of 0.788, the best AUC of 0.801, and the best SPE of 0.769. It suggests that our proposed method can sufficiently capture brain information for function representation to boost brain disease diagnosis. For the diagnosis of MCI, similar conclusions can be drawn from Table 1. Specifically, our method achieves the best ACC of 0.754, and the best SPE of 0.717. Overall, the results in Table 1 prove that our proposed method is effective and general.

Ablation Study. We conduct ablation study on the proposed LLM-guided supervision to illustrate its effectiveness. As shown in Table 1, compared with no LLM-guided supervision, our method improves ACC by 0.030 and 0.049 on ASD and MCI diagnosis, respectively (Row 8, Row 9). It indicates that the incorporation of prior knowledge enhances the learned affinities, aligning them more closely with the complex brain function, thereby boosting performance.

Transfer Learning Study. We pre-train on one dataset and apply the pre-trained model on another dataset to obtain function representation for brain disease diagnosis. The results are shown in the third part of Table 1. We achieve competitive results in comparison to other methods. It illustrates the generality of our proposed model, that is, a well-trained affinity learning model can be directly applied to unseen datasets without fine-tuning.



(a) Parameter Sensitivity

(b) Discrepancy Maps of Two Channels

Fig. 3. Parameter sensitivity analysis and brain network analysis: (a) The performance of various dimensions of the learned function representation on ASD diagnosis; (b) Discrepancy maps between ASD and NC groups of two channels from learned function representation.

3.4 Analysis and Discussion

In this section we perform a parameter sensitivity analysis for the dimension of learned function representation, which plays a key role in representation capability. In addition, the learned function representation should reflect the brain function, which is investigated by brain network analysis.

Parameter Sensitivity Analysis. An important hyperparameter of the affinity learning module is the dimension of the learned function representation, so we investigate the effect of it on performance. We conduct experiments under dimension settings of 1, 4, 8, 16, 32, 64, respectively, and the results are shown in Fig. 3 (a). We observe that the performance shows a trend of first improvement and then decline. It may be because small dimensions are insufficient to represent complex brain function and too large dimensions contain noisy interference. Therefore, we choose 16 as the dimension of the learned function representation.

Brain Network Analysis. To explore the significance of the learned function representation, we map each channel of the learned function representation with Yeo 7 brain networks [19], which are influential network identification scheme derived from rs-fMRI. Specifically, we use AAL atlas to parcellate brain regions for ease of comparison. We compute the disparity in learned function representations between the NC and ASD groups, subsequently correlating them with the Yeo 7 brain networks on a per-channel basis using PCC measures. The two channels with the largest PCC values (0.623 and 0.628) are shown in Fig. 3 (b), and all channels are detailed in supplementary materials. We observe that the most significant differences between the ASD and NC groups are in the somatomotor network and the default mode network, which is consistent with previous literatures [13, 14]. It reveals that our proposed approach learns significant function representation for disease diagnosis.

4 Conclusion

In this paper, we introduce an affinity learning framework aimed at acquiring informative brain function representation in connectomics. We evaluate the effectiveness of our proposed approach across two distinct datasets, yielding promising performance. Future endeavors will involve validating our approach on additional datasets and investigating the association between acquired brain function representation and cognitive tasks.

Acknowledgments. This work was partially supported by National Natural Science Foundation of China (62131015) and Shanghai Municipal Central Guided Local Science and Technology Development Fund (YDZX20233100001001).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alshami, A., Elsayed, M., Ali, E., Eltoukhy, A.E., Zayed, T.: Harnessing the power of chatgpt for automating systematic review process: Methodology, case study, limitations, and future directions. Systems **11**(7), 351 (2023)
2. Bijsterbosch, J., Smith, S.M., Beckmann, C.: An introduction to resting state fMRI functional connectivity. Oxford University Press (2017)
3. Chen, D., Liu, M., Shen, Z., Zhao, X., Wang, Q., Zhang, L.: Learnable subdivision graph neural network for functional brain network analysis and interpretable cognitive disorder diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 56–66. Springer (2023)
4. Chen, D., Zhang, L.: Fe-stgnn: Spatio-temporal graph neural network with functional and effective connectivity fusion for mci diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 67–76. Springer (2023)
5. Cui, H., Dai, W., Zhu, Y., Kan, X., Gu, A.A.C., Lukemire, J., Zhan, L., He, L., Guo, Y., Yang, C.: Braingb: A benchmark for brain network analysis with graph neural networks. IEEE Transactions on Medical Imaging **42**(2), 493–506 (2022)
6. Fox, M.D., Greicius, M.: Clinical applications of resting state functional connectivity. Frontiers in Systems Neuroscience **4**, 1443 (2010)
7. Kawahara, J., Brown, C.J., Miller, S.P., Booth, B.G., Chau, V., Grunau, R.E., Zwicker, J.G., Hamarneh, G.: Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage **146**, 1038–1049 (2017)
8. Khosla, M., Jamison, K., Ngo, G.H., Kuceyeski, A., Sabuncu, M.R.: Machine learning in resting-state fmri analysis. Magnetic Resonance Imaging **64**, 101–121 (2019)
9. Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Staib, L.H., Ventola, P., Duncan, J.S.: Braingnn: Interpretable brain graph neural network for fmri analysis. Medical Image Analysis **74**, 102233 (2021)
10. Liu, M., Zhang, H., Liu, M., Chen, D., Zhuang, Z., Wang, X., Zhang, L., Peng, D., Wang, Q.: Randomizing human brain function representation for brain disease diagnosis. IEEE Transactions on Medical Imaging **43**(7), 2537–2546 (2024)

11. Mahmood, U., Fu, Z., Calhoun, V.D., Plis, S.: A deep learning model for data-driven discovery of functional connectivity. Algorithms **14**(3), 75 (2021)
12. Mohanty, R., Sethares, W.A., Nair, V.A., Prabhakaran, V.: Rethinking measures of functional connectivity via feature extraction. Scientific Reports **10**(1), 1298 (2020)
13. Nunes, A.S., Peatfield, N., Vakorin, V., Doesburg, S.M.: Idiosyncratic organization of cortical networks in autism spectrum disorder. Neuroimage **190**, 182–190 (2019)
14. Padmanabhan, A., Lynch, C.J., Schaer, M., Menon, V.: The default mode network in autism. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging **2**(6), 476–486 (2017)
15. Salas, J.A., Bayrak, R.G., Huo, Y., Chang, C.: Reconstruction of respiratory variation signals from fmri data. Neuroimage **225**, 117459 (2021)
16. Vigneau-Roy, N., Bernier, M., Descoteaux, M., Whittingstall, K.: Regional variations in vascular density correlate with resting-state and task-evoked blood oxygen level-dependent signal amplitude. Human Brain Mapping **35**(5), 1906–1920 (2014)
17. Wang, Z., Jie, B., Feng, C., Wang, T., Bian, W., Ding, X., Zhou, W., Liu, M.: Distribution-guided network thresholding for functional connectivity analysis in fmri-based brain disorder identification. IEEE Journal of Biomedical and Health Informatics **26**(4), 1602–1613 (2021)
18. Yan, C., Zang, Y.: Dparsf: a matlab toolbox for" pipeline" data analysis of resting-state fmri. Frontiers in Systems Neuroscience **4**, 1377 (2010)
19. Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al.: The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology (2011)
20. Zhang, J., Zhou, L., Wang, L., Liu, M., Shen, D.: Diffusion kernel attention network for brain disorder classification. IEEE Transactions on Medical Imaging **41**(10), 2814–2827 (2022)
21. Zhang, P., Wen, G., Cao, P., Yang, J., Zhang, J., Zhang, X., Zhu, X., Zaiane, O.R., Wang, F.: Brainusl: U nsupervised graph s tructure l earning for functional brain network analysis. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 205–214. Springer (2023)
22. Zhao, K., Fonzo, G.A., Xie, H., Oathes, D.J., Keller, C.J., Carlisle, N.B., Etkin, A., Garza-Villarreal, E.A., Zhang, Y.: Discriminative functional connectivity signature of cocaine use disorder links to rtms treatment response. Nature Mental Health pp. 1–13 (2024)
23. Zhao, L., Wu, Z., Dai, H., Liu, Z., Hu, X., Zhang, T., Zhu, D., Liu, T.: A generic framework for embedding human brain function with temporally correlated autoencoder. Medical Image Analysis **89**, 102892 (2023)