Meta-analysis guided multi-task graph transformer network for diagnosis of neurological disease and cognitive deficits

Jing Xia[1], Yi Hao Chan[1], Jagath C. Rajapakse[1(✉)]

[1] School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore
jing_xia,yihao.chan,deepank002,asjagath}@ntu.edu.sg

Abstract. Neurological diseases, such as schizophrenia and attention deficit hyperactivity disorder (ADHD), alter functional connectivity (FC) and are often accompanied by cognitive deficits. Leveraging shared neural mechanisms underlying both neurological disease and cognitive deficits can enhance diagnostic accuracy. However, due to the complex neural mechanisms of these conditions, diagnosing them based on FC alone still presents challenges in terms of accuracy and biomarker reliability. To address these challenges, we designed a meta-analysis guided multi-task graph transformer network to simultaneously predict neurological disease and cognitive deficits and examine alterations in brain FC associated with these conditions. The framework employs a graph transformer method as the encoder and integrates a joint attention mechanism to capture shared disease–cognition features while utilizing saliency pooling to extract saliency weights for each task. To enhance the reliability of saliency weights, we incorporate meta-analysis guidance that aggregates data from 470 functional studies in the BrainMap database. Then, we establish reference probability maps for brain activations associated with neurological diseases and cognitive deficits using a Naive Bayes classifier. The saliency weights learned from saliency pooling are then constrained to align with these references using Pearson correlation. Experiments on the COBRE and ADHD-200 datasets indicate that our proposed method outperforms state-of-the-art multi-task learning models in classifying schizophrenia and ADHD, as well as in predicting their related cognitive deficits. Moreover, the biomarkers extracted from our models exhibit biologically meaningful patterns.

Keywords: Meta-analysis · Multi-task learning · Graph transformer · Functional connectivity · Neurological disease · Cognitive deficits.

1 Introduction

Neurological diseases, such as schizophrenia (SZ) and attention deficit hyperactivity disorder (ADHD), are frequently accompanied by cognitive deficits in domains including working memory, social cognition, language, and attention

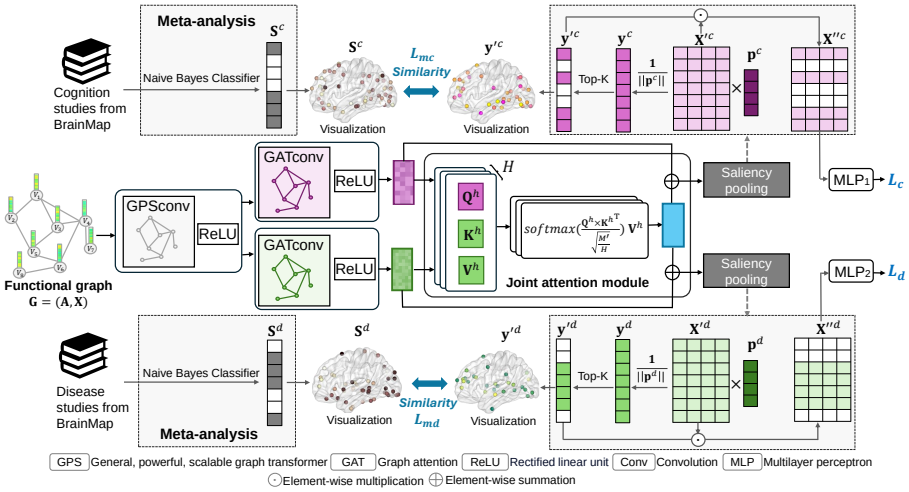

Fig. 1. MAG-MT employs a general, powerful, scalable graph transformer (GPS) convolution as a common encoder and graph attention (GAT) convolution as task-specific encoders. A joint attention module using multi-head cross-attention identifies shared features across tasks. Finally, a decoder—with a saliency pooling layer and an individual MLP for each prediction—generates task-specific outcomes, and the saliency weights are constrained by reference probability maps from meta-analysis.

[1, 2], suggesting a shared underlying mechanism. Functional connectivity (FC), which is derived from resting-state fMRI (rs-fMRI) by capturing neural activity fluctuations through blood-oxygen-level-dependent signals, is valuable for distinguishing patients and controls [3, 4] and for estimating various cognitive states [5, 6]. Given the shared mechanisms between neurological diseases and cognitive deficits, leveraging their unique and shared features based on FC may improve both neurological disease diagnosis and cognition prediction.

Multi-task learning effectively exploits common features across tasks to enhance overall performance [7–9]. For example, [8] developed a brain transformer network that employs a shared encoder to predict multiple cognitive scores. However, many existing methods do not fully leverage the complementary information available across tasks. To address this, [9] introduced an MLP-based mechanism to capture both shared and complementary features for various cognitive predictions. Nevertheless, due to the complexity of the neural mechanisms underlying multiple tasks, current approaches still lack a robust solution to identify reliable biomarkers for different tasks.

Meta-analysis aggregates data from numerous previous studies, such as those in the BrainMap database [10], and identifies brain activation regions associated with various conditions that serve as valuable biomarker references [11–13]. Specifically, statistical methods—such as the Naive Bayes classifier [14]—are employed to compute the probability of a region of interest (ROI) associated with each condition, thereby generating a reference map for each. Incorporating these

reference maps into a multi-task learning framework could provide useful and reliable prior knowledge for informative feature learning and guide model training for multiple tasks. However, current research has not integrated this prior knowledge into neurological disease diagnosis or cognitive prediction.

This study proposes a meta-analysis guided multi-task graph transformer network (MAG-MT) to simultaneously predict neurological disease and cognitive deficits, and to examine associated brain FC changes, as shown in Fig. 1. A graph transformer convolution serves as a common encoder, while separate graph attention convolutions extract disease-specific and cognition-specific node embeddings. A joint attention module, incorporating a multi-head cross-attention mechanism, identifies shared features that contribute to both neurological disease and cognitive predictions. The combined shared and task-specific features are fed into a decoder for final predictions. The decoder employs saliency pooling to select salient regions for each task and uses two separate multilayer perceptrons (MLPs) to simultaneously predict the respective outcomes. To guide model training, we incorporate meta-analysis-derived prior knowledge by generating reference probability maps from the BrainMap database and constraining the saliency weights learned from the saliency pooling to align with these reference maps via additional loss functions. Experiments demonstrate that MAG-MT outperforms state-of-the-art multi-task learning methods on both COBRE and ADHD-200 datasets, as well as in related cognitive prediction. Furthermore, MAG-MT yields interpretable saliency maps that reveal distinct regional contributions: in the COBRE dataset, the postcentral and anterior frontal cortices are most relevant for SZ classification, while the prefrontal and occipital regions are linked to cognitive deficits; for ADHD-200, the lateral temporal and middle frontal cortices are most relevant for ADHD classification, with the insula and inferior frontal cortex most related to cognitive deficits.

2 Materials and Methodology

2.1 Datasets and Image Preprocessing

We used the publicly available Center for Biomedical Research Excellence (COBRE) [15] dataset (135 participants: 75 healthy controls, 60 SZ patients) for SZ and its related cognitive deficits prediction, and the ADHD-200 [16] dataset (489 participants: 256 healthy controls, 233 ADHD patients) from three sites for ADHD and its related cognitive deficits prediction. Rs-fMRI images were preprocessed using fMRIPrep [18], and functional connectivity (FC) was constructed from 264 ROIs defined by the Power atlas [17]. Cognitive function was assessed by averaging scores for working memory, social ability, and verbal learning (range: 30–100) for SZ, and by full-scale IQ (range: 78–153, covering attention, working memory, and reasoning) for ADHD. To build the reference map, we selected 160 functional studies on SZ using the keyword "schizophrenia" and 160 studies on its cognitive deficits using the keywords "working memory", "social cognition", and "language" via the Sleuth toolbox [23] from the BrainMap database, and

similarly selected 75 studies on ADHD using the keyword "ADHD" and 75 studies on its cognitive deficits using the keywords "attention", "working memory", and "reasoning".

2.2 Multi-task graph transformer network

Graph Representation Construction For each subject, a functional graph $\mathbf{G} = (\mathbf{A}, \mathbf{X})$ is constructed, with ROIs as nodes and connections as edges. $\mathbf{A} \in \mathbb{R}^{N \times N}$ denotes the adjacency matrix, which is obtained by thresholding the FC matrix to retain the top 10% of edges [19] for each node, thereby preserving connection sparsity. $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\} \in \mathbb{R}^{N \times M}$ contains the features for the N nodes, where each $\mathbf{x}_n$ is an M-dimensional feature of node n, defined as its connections with all nodes.

Common and Specific Encoders The common encoder captures features shared by brain neurological disease and cognition predictions. It consists of one layer of general, powerful, scalable graph transformer (GPS) convolution [20], which combines traditional transformer and spatial graph convolution techniques to effectively learn node representations, followed by a rectified linear unit (ReLU) activation. Task-specific encoders for neurological disease and cognition extract relevant features using one layer of graph attention (GAT) convolution [21] followed by a ReLU activation. GAT convolution, with its low parameter count, is well-suited for multi-task settings. The disease-specific node embeddings are generated as:

$$\mathbf{X}^d = ReLU\Big[GAT_1\Big(ReLU\big(GPS(\mathbf{X})\big)\Big)\Big] \in \mathbb{R}^{N \times M'}, \tag{1}$$

where M' is the length of node embeddings. Cognitive-specific node embeddings $\mathbf{X}^c \in \mathbb{R}^{N \times M'}$ are obtained similarly.

Joint Attention Module The joint attention module uses multi-head cross-attention to identify shared node embeddings for neurological disease and cognition. Given cognitive-specific embeddings $\mathbf{X}^c$ and neurological-specific embeddings $\mathbf{X}^d$, for each attention head h, we compute,

$$\mathbf{Q}^h = \mathbf{X}^c \mathbf{W}^{Q,h}, \quad \mathbf{K}^h = \mathbf{X}^d \mathbf{W}^{K,h}, \quad \mathbf{V}^h = \mathbf{X}^d \mathbf{W}^{V,h}, h \in \{1, ..., H\}, \tag{2}$$

with learnable parameters $\mathbf{W}^{Q,h}$, $\mathbf{W}^{K,h}$, and $\mathbf{W}^{V,h}$. $\mathbf{Q}^h$, $\mathbf{K}^h$ and $\mathbf{V}^h \in \mathbb{R}^{N \times \frac{M'}{H}}$ represent the query, key and value matrices in h-th head, respectively. Each head's output is: $\mathbf{E}^h = softmax(\frac{\mathbf{Q}^h(\mathbf{K}^h)^T}{\sqrt{M'/H}})\mathbf{V}^h$, and H is the number of attention heads. The joint feature $\mathbf{X}^{'}$ is then formed by concatenating all heads and applying a linear projection: $\mathbf{X}^{'} = concat(\mathbf{E}^1, \mathbf{E}^2, ..., \mathbf{E}^H)\mathbf{W}$, where $\mathbf{W} \in \mathbb{R}^{M^{'} \times M^{'}}$ is learnable. The joint attention module thus learns shared features between neurological disease and cognitive functioning predictions. Finally, to effectively integrate task-specific features and shared features, the joint feature is added element-wise to the original embeddings: $\mathbf{X}^{'c} = \mathbf{X}^c + \mathbf{X}^{'}; \mathbf{X}^{'d} = \mathbf{X}^d + \mathbf{X}^{'}$.

Decoder For each task, the decoder consists of a saliency pooling operation and an MLP model including two fully connected layers, dropout, and ReLU activation function for final prediction. The saliency pooling is adapted from top-K pooling [22], preserving the most informative node embeddings while discarding less relevant ones. Taking the cognition decoder as an example, a trainable projection vector $\mathbf{p}^c \in \mathbb{R}^{M' \times 1}$ is employed to compute a weight for each node, and the propagation rule of saliency pooling is: $\mathbf{y}^c = \mathbf{X}'^c \mathbf{p}^c / \|\mathbf{p}^c\|$, where $\mathbf{y}^c \in \mathbb{R}^{N \times 1}$ is the weights indicating the retention potential of all nodes. The u nodes with highest weights (i.e., $idx^c = rank(\mathbf{y}^c, u)$) are selected, where idx^c represents the indices of u selected nodes. The weights for the u nodes are passed through a sigmoid function, while weights for all other nodes are set to zero:

$$\mathbf{y}_i^{c'} = \begin{cases} sigmoid(\mathbf{y}_i^c), & i \in idx^c \\ 0, & i \notin idx^c \end{cases} \tag{3}$$

where $\mathbf{y}^{c'} \in \mathbb{R}^{N \times 1}$ represents the saliency weights of all nodes. The final cognition-specific node embeddings are computed as: $\mathbf{X}''^c = \mathbf{X}'^c \odot \mathbf{y}^{c'}$, where $\odot$ denotes element-wise multiplication. This process ensures that saliency weights reflect the importance of nodes. The final node embeddings for neurological disease $\mathbf{X}''^d$ are obtained similarly, and MLPs are used for prediction.

2.3 Meta-analysis Constraint

Generation of Reference Probability Maps The BrainMap database provides foci MNI coordinates in brain regions associated with neurological disease or cognition in each FC study. For each study, we mapped the coordinates onto 264 ROIs, thereby identifying the activated and non-activated ROIs relevant to conditions. A Naive Bayes classifier was used to compute the probability of each ROI contributing to neurological disease and cognition across all studies [14], generating the reference probability maps $\mathbf{S}^d$ and $\mathbf{S}^c \in \mathbb{R}^{N \times 1}$, where each entry represents the probability of an ROI being associated with neurological disease or cognition, respectively. For consistency with saliency weights from pooling, we retained the highest u probabilities in $\mathbf{S}^d$ and $\mathbf{S}^c$, setting all others to zero.

Incorporation of Reference Probability Maps To integrate meta-analysis guidance, we introduce two additional constraints, L_{mc} and L_{md}, that encourage the saliency weights to match the reference probability maps from meta-analysis. Specifically, these constraints are defined as:

$$L_{mc} = \sum\nolimits_{j=1}^{B} (1 - corr(\mathbf{S}_j^c, \mathbf{y}_j'^c)); \quad L_{md} = \sum\nolimits_{j=1}^{B} (1 - corr(\mathbf{S}_j^d, \mathbf{y}_j'^d)). \tag{4}$$

corr indicates the Pearson correlation coefficient, and B is the number of subjects. For neurological disease classification, we use cross-entropy loss (L_d); for cognition prediction, we use root mean square error (RMSE, L_c). The overall loss is defined as: $L = L_c + L_d + \lambda(L_{mc} + L_{md})$, with λ controlling the weight of the meta-analysis constraint.

Table 1. Mean (standard deviation) of performance measures on COBRE (top) and ADHD-200 (bottom) based on FC.

Methods	Cognition prediction			Schizophrenia classification		
	Correlation	RMSE	MAE	Accuracy	Sensitivity	Specificity
GAT (MT)	0.43(0.02)	11.69(0.74)	9.59(0.66)	0.75(0.05)	0.76(0.05)	0.75(0.05)
GPS (MT)	0.45(0.05)	11.38(0.86)	9.45(0.71)	0.77(0.05)	0.78(0.06)	0.77(0.05)
ML-Att	0.46(0.02)	10.95(0.88)	9.08(0.63)	0.79(0.05)	0.79(0.06)	0.79(0.05)
MT-brain	0.47(0.03)	10.87(0.63)	8.92(0.57)	0.78(0.06)	0.79(0.07)	0.78(0.06)
Ours	**0.51(0.02)**	**10.24(0.70)**	**8.26(0.61)**	**0.81(0.05)**	**0.82(0.05)**	**0.81(0.04)**
Methods	Cognition prediction			ADHD classification		
	Correlation	RMSE	MAE	Accuracy	Sensitivity	Specificity
GAT (MT)	0.38(0.03)	13.93(0.86)	11.80(0.97)	0.66(0.04)	0.66(0.04)	0.66(0.04)
GPS (MT)	0.41(0.05)	13.21(0.83)	11.25(0.89)	0.67(0.06)	0.68(0.05)	0.67(0.08)
ML-Att	0.43(0.04)	12.76(0.76)	11.08(0.60)	0.69(0.07)	0.69(0.07)	0.69(0.08)
MT-brain	0.44(0.03)	12.65(0.89)	10.69(0.74)	0.69(0.07)	0.70(0.08)	0.69(0.06)
Ours	**0.47(0.04)**	**12.31(0.84)**	**10.33(0.91)**	**0.71(0.04)**	**0.72(0.05)**	**0.71(0.05)**

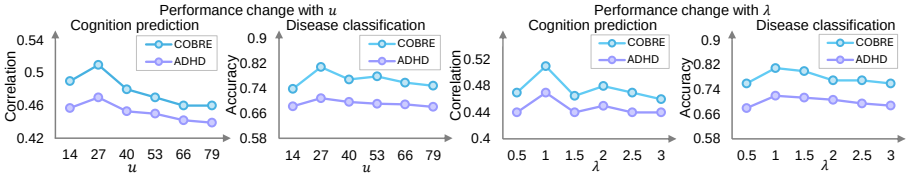

Fig. 2. The results of cognition and neurological disease prediction changes with different u in saliency pooling and λ on COBRE and ADHD-200 datasets.

2.4 Implementation

All models were implemented on an NVIDIA A100 GPU. We computed the mean and standard deviation over ten runs. To reduce overfitting, we incorporated dropout layers (dropout rate = 0.5) in the MLP layer, applied L2 regularization (weight = 0.001), and employed early stopping. The Adam optimizer was used with a batch size of 10. The learning rate and number of epochs were set to 0.001 and 40 for the COBRE dataset, and 0.0001 and 40 for the ADHD-200 dataset, respectively. We conducted 5-fold cross-validation, and calculated the mean accuracy and standard deviation across 10 runs. Pearson's correlation coefficient, RMSE, and mean absolute error (MAE) were used to quantify the differences between the true and predicted cognitive values, while accuracy, sensitivity, and specificity were used to assess the classification performance. The code is available at: https://github.com/qiezi005/meta-guided.

3 Experiments and Results

Competing Methods and Hyperparameter Setting We compared our method against GAT, GPS, and two state-of-the-art multi-task learning methods

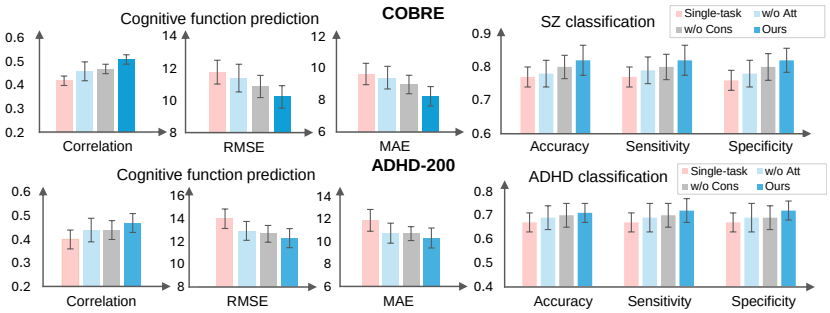

Fig. 3. Results of the ablation study of MAG-MT, comparing with three ablated variants for COBRE and ADHD-200 datasets. The performance of our model is significantly better than the ablated variants, confirmed by the Student's t-test ($p<0.05$).

based on brain FC. For multi-tasking (MT) structure, we constructed **GAT (MT)** by incorporating a shared GAT convolution layer along with two task-specific GAT convolution layers for simultaneous regression and classification; a similar construction was applied to **GPS (MT)**. **MT-brain** [8] employs a brain network transformer as a common encoder for multiple cognition prediction, while **ML-Att** [9] incorporates a joint mechanism for multi-task FC analysis. The number of attention heads in GPS and MT-brain was set to 4, matching our method. The convolution kernel in ML-Att was configured according to its original publication. For our model, the dimension of node embeddings was tuned across the range [32, 64, 96, 128, 256]. For SZ classification, the highest accuracy was achieved with a dimension of 64, while for ADHD prediction, the optimal performance was obtained with a dimension of 128. H was set to 4. λ was set to 1, and u in saliency pooling was set to 27, based on the best performance in Fig. 2 on both datasets. Each of MLP_1 and MLP_2 consisted of two fully connected layers. For regression tasks, the layers contained 256 and 1 hidden node, respectively, while for classification tasks, they included 256 and 2 hidden nodes.

Table 1 shows the comparison results on the COBRE and ADHD-200 datasets, listing the mean and standard deviation of ten runs. ML-Att and MT-brain can achieve better performance than GPS (MT) and GAT (MT). Moreover, our model significantly outperformed other methods on both tasks, as confirmed by a Students' t-test ($p < 0.05$). Compared with ML-Att and MT-brain, our method significantly improved accuracy by 2.5% to 3.8% on SZ classification and 8.5% to 10.9% on its related cognition prediction, respectively. Similarly, our method achieved improvements of 2.9% for ADHD classification and 6.8% to 9.3% for cognition prediction. We attribute these improvements to the utilization of the joint attention module and meta-analysis constraints to guide the model.

Ablation Study To evaluate the effectiveness of the joint attention module and meta-analysis constraints, we compared the proposed model against three ab-

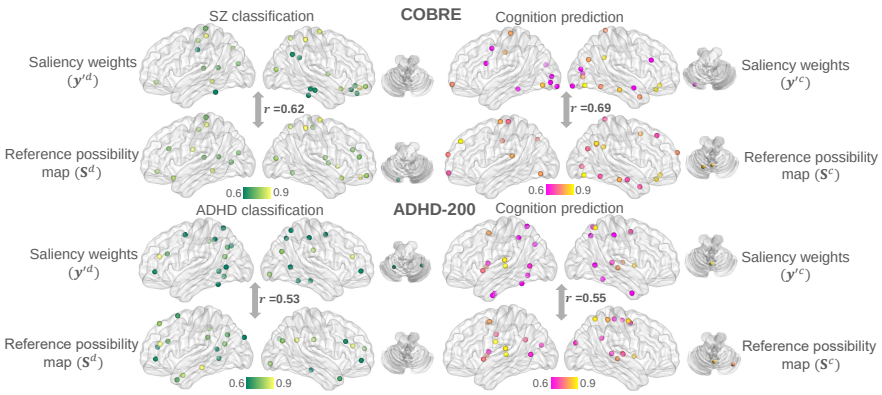

Fig. 4. The top 27 discriminative ROIs identified by saliency pooling were associated with neurological disease and cognitive deficits on both datasets.

lated variants: 1) GPS on the single task; 2) a model without joint cross-attention (w/o Att); 3) a model without the meta-analysis constraints (w/o Cons). Fig. 3 shows that incorporating the joint attention module significantly improves performance, highlighting the importance of capturing shared features. The addition of the meta-analysis constraints (ours) gets significant improvements compared with other ablated variants. We infer that the meta-analysis constraints guide the model to avoid learning shortcuts that may otherwise lead to overfitting.

Discriminative ROIs for Prediction As shown in Fig. 4, for SZ, the postcentral gyrus, temporal–parietal junction, right anterior frontal cortex, and cerebellum strongly contribute to neurological disease classification, while the frontal, occipital, and temporal gyri are primarily associated with cognitive deficits. Notably, abnormalities in the right anterior frontal lobe are linked to negative SZ symptoms (e.g., flattened affect, alogia, abolition) [24], and temporal–parietal junction changes correlate with impaired social perception, theory of mind, and auditory hallucinations [25, 26]. These regions significantly correlate with the reference probability maps from BrainMap ($r = 0.62, p < 0.001$ for SZ, and $r = 0.69, p < 0.001$ for cognition), validating the meta-analysis constraint.

For ADHD, the most discriminative regions for classification are in the middle frontal gyrus, superior parietal, and lateral temporal cortices, which often exhibit reduced activity or structural abnormalities underlying attention deficits. Key regions for cognitive deficit prediction include the insula, inferior frontal, and lateral temporal cortices; here, the insula integrates internal states and emotions [27], while the inferior frontal cortex supports response inhibition and decision-making [28, 29]. These regions also show significant correlations with the reference probability maps ($r = 0.53, p < 0.001$ for ADHD, and $r = 0.55, p < 0.001$ for cognition), indicating that the model can learn general biomarkers consistent with findings from hundreds of studies.

4 Conclusion

We propose using brain reference probability maps generated from meta-analyses to guide a multi-task graph transformer network that simultaneously predicts neurological diseases and cognitive performance. The resulting salient maps highlight biologically meaningful biomarkers associated with ADHD, schizophrenia (SZ), and confounded cognitive deficits. In future work, we aim to extend our method to multi-disease classification applications.

Ackwonledgement. Research supported by AcRF Tier-2 grant MOE T2EP20121-0003 and Tier-1 grant RG15/24 of Ministry of Education, Singapore.

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

References

1. Sheffield, J.M. and Barch, D.M.: Cognition and resting-state functional connectivity in schizophrenia. Neuroscience & Biobehavioral Reviews **61**, 108-120 (2016).
2. Castellanos, F.X., Sonuga-Barke, et al.: Characterizing cognition in ADHD: beyond executive dysfunction. Trends in cognitive sciences **10**(3), 117-123 (2006).
3. Passiatore, R., Antonucci, L.A., et al.: Changes in patterns of age-related network connectivity are associated with risk for schizophrenia. Proceedings of the National Academy of Sciences **120**(32), e2221533120 (2023).
4. Tomasi, D. and Volkow, N.D.: Abnormal functional connectivity in children with attention-deficit/hyperactivity disorder. Biological psychiatry, **71**(5), pp.443-450 (2012).
5. Menon, V., Palaniyappan, L. et al.: Integrative brain network and salience models of psychopathology and cognitive dysfunction in schizophrenia. Biological Psychiatry, **94**(2), pp.108-120 (2023).
6. Zhang, R., Murray, S.B., et al.: Functional connectivity and complexity analyses of resting-state fMRI in pre-adolescents demonstrating the behavioral symptoms of ADHD. Psychiatry Research, **334**, 115794 (2024).
7. Huang, Z.A., Hu, Y., et al.: Federated multi-task learning for joint diagnosis of multiple mental disorders on MRI scans. IEEE Transactions on Biomedical Engineering, **70**(4), 1137-1149 (2022).
8. Kan, X., Cui, H., et al.: Multi-task Learning for Brain Network Analysis in the ABCD study. In IEEE-EMBS International Conference on Biomedical and Health Informatics (2024).
9. Xia, J., Chen, N., et al.: Multi-level and joint attention networks on brain functional connectivity for cross-cognitive prediction. Medical image analysis, **90**, p.102921 (2023).
10. Laird, A.R., Riedel, M.C., et al.: Neural architecture underlying classification of face perception paradigms. Neuroimage, **119**, 70-80 (2015).
11. Cash, R.F., Müller, V.I., et al.: Altered brain activity in unipolar depression unveiled using connectomics. Nature Mental Health, **1**(3), 174-185 (2023).

12. Meinke, C., Lueken, U., et al.: Predicting treatment outcome based on resting-state functional connectivity in internalizing mental disorders: a systematic review and meta-analysis. Neuroscience & Biobehavioral Reviews, 105640 (2024).
13. Zhu, J., Qiu, A.: Interindividual variability in functional connectivity discovers differential development of cognition and transdiagnostic dimensions of psychopathology in youth. NeuroImage, **260**, 119482 (2022).
14. Yarkoni, T., Poldrack, R.A., et al.: Large-scale automated synthesis of human functional neuroimaging data. Nature methods, **8**(8), 665-670 (2021).
15. Aine, C.J., Bockholt, H.J., et al.: Multimodal neuroimaging in schizophrenia: description and dissemination. Neuroinformatics, **15**, 343-364, 2017.
16. ADHD-200 consortium.: The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. Frontiers in systems neuroscience, **6**, 62 (2012).
17. Power, J.D., Cohen, A.L., et al.: Functional network organization of the human brain. Neuron, **72**(4), 665-678 (2011).
18. Esteban, O., Markiewicz, C.J., et al.: fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature methods, **16**(1), 111-116 (2019).
19. Li, X., Zhou, et al.: BrainGNN: Interpretable brain graph neural network for fmri analysis. Medical Image Analysis, **74**, 102233 (2021).
20. Rampášek, L., Galkin, M., et al.: Recipe for a general, powerful, scalable graph transformer. Advances in Neural Information Processing Systems, **35**, 14501-14515 (2022).
21. Velickovic, P., Cucurull, G., et al.: Graph attention networks. arXiv preprint arXiv:1710.10903.
22. Knyazev, B., Taylor, G.W., et al.: Understanding attention and generalization in graph neural networks. Advances in neural information processing systems, 32 (2019).
23. Laird, A.R., Eickhoff, S.B., et al.: The BrainMap strategy for standardization, sharing, and meta-analysis of neuroimaging data. BMC research notes, **4**, 1-9 (2011).
24. Kirschner, M., Schmidt, A., et al.: Orbitofrontal-striatal structural alterations linked to negative symptoms at different stages of the schizophrenia spectrum. Schizophrenia bulletin, **47**(3), 849-863 (2021).
25. Patel, G.H., Arkin, S.C., et al.: Failure to engage the temporoparietal junction/posterior superior temporal sulcus predicts impaired naturalistic social cognition in schizophrenia. Brain, **144**(6), 1898-1910 (2021).
26. Mondino, M., Jardri, R., et al.: Effects of fronto-temporal transcranial direct current stimulation on auditory verbal hallucinations and resting-state functional connectivity of the left temporo-parietal junction in patients with schizophrenia. Schizophrenia bulletin, **42**(2), 318-326 (2016).
27. Wang, M., Yu, J., et al.: Neural correlates of executive function and attention in children with ADHD: An ALE meta-analysis of task-based functional connectivity studies. Psychiatry Research, **345**, 116338 (2025).
28. Lee, C.Y., Goh, J.O.S., et al.: Differential neural processing of value during decision-making in adults with attention-deficit/hyperactivity disorder and healthy controls. Journal of Psychiatry and Neuroscience, **48**(2), E115-E124 (2023).
29. Mukherjee, P., Vilgis, V., et al.: Associations of irritability with functional connectivity of amygdala and nucleus accumbens in adolescents and young adults with ADHD. Journal of Attention Disorders, **26**(7), 1040-1050 (2022).