Integrating meta-analysis in multi-modal brain studies with graph-based attention transformer

Hyoungshin Choi[1,2], Sunghun Kim[3], Jong-eun Lee[1,2], Bo-yong Park[2,3], and Hyunjin Park[1,2*]

[1] Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, South Korea
[2] Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, South Korea
[3] Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea
{gudtls17, hyunjinp}@skku.edu

Abstract. Multi-modal neuroimaging studies are essential for exploring various brain disorders; however, they are typically limited in sample size owing to the cost of image acquisition. Meta-analysis is an underutilized method that integrates the findings from multiple studies derived from large samples to assist individual studies. Neuroimaging studies are increasingly adopting transformer architecture for network analysis; however, they tend to overlook local brain networks. To address these gaps, we propose the Meta-analysis Enhanced Graph Attention TransFormer (MEGATF), a novel method for performing multimodal brain analysis built on a graph transformer framework aided with meta-analysis information derived from *NeuroSynth*. Our method adapts a graph neural network with a transformer attention mechanism that favors local networks and multimodal interactions using PET or cortical thickness. Our method achieved a state-of-the-art classification performance on mild cognitive impairment and attention-deficit/hyperactivity disorder datasets, distinguishing individuals with brain disorders from controls. Furthermore, it identified disease-affected brain regions and associated cognitive decoding that aligned with existing findings, thereby enhancing its interpretability. Our code is at https://github.com/gudtls17/MEGATF.

Keywords: Graph transformer · meta-analysis · multimodal analysis · disease classification.

1 Introduction

Neuroimaging helps explore neurodegenerative and psychiatric brain disorders [2,7]. Positron emission tomography (PET) can measure abnormal metabolic activity related to Alzheimer's disease (AD) and mild cognitive impairment (MCI) [19]. Magnetic resonance imaging (MRI), specifically structural MRI (sMRI) and functional MRI (fMRI), can detect structural and functional alterations, respectively in the brain related to AD, MCI, and attention-deficit/hyperactivity

disorder (ADHD) [20]. Each imaging modality makes a different contribution. Thus, it is crucial to integrate various modalities in a multimodal manner.

A meta-analysis integrates findings from multiple large-sample studies to help draw conclusions from a given study with fewer samples. Meta-analysis strengthens the overall conclusions and provides valuable supplementary information to address limitations of individual studies. This helps mitigate data shortage and focuses on critical brain regions, thereby enhancing the interpretability of the analysis [11,26]. Incorporating the information from meta-analysis into multimodal approaches can facilitate a deeper understanding of brain disorders by providing complementary insights. Notably, meta-analysis platforms based on fMRI offer cognitive decoding using standardized cognition-behavior terms [26], aiding the construction of cognition-related representations and interpretation [6,7]. That is, researchers can harness rich information on the cognition-behavior terms tied to brain activation maps. Thus, when integrated into multimodal approaches, meta-analysis has the potential to significantly advance the accuracy and reliability of multimodal analysis in brain disorder research. However, it remains underutilized.

The use of connectivity graphs in neuroimaging has facilitated the representation of interactions between brain regions, thereby advancing brain network analysis [17]. Although the transformer model has shown promising performance in conventional brain analysis [3,14], its application to brain networks remains limited because of the difficulty in modeling local interactions within a token (i.e., sub-modules in a brain region). To address this, research on graph transformers has been actively conducted by combining graph neural networks (GNNs) with transformer architectures [27]. However, existing graph transformer models tend to prioritize global network representations, while overlooking the significance of local brain networks [23,27]. Additionally, the node attributes of brain regions are well captured [10], but the edge attributes representing regional interactions are often neglected, especially in multimodal analysis. We aimed to bridge these gaps by incorporating diverse region-by-region interactions expressed in multiple modalities as edge features, enhancing the representation of local brain networks, and optimizing their integration within the transformer framework using an edge graph attention network [21]. Our study addresses the following two key challenges: **(1) The lack of meta-analysis information in multimodal brain analysis. (2) The limited use of edge features, often neglecting diverse regional brain interactions in multimodal inputs.**

To overcome these limitations, we introduce a Meta-analysis Enhanced Graph Attention TransFormer (MEGATF), which is a novel method for multimodal brain analysis, built on a graph transformer framework, aided by meta-analysis information. Our key contributions are as follows: **(1) We incorporate meta-analysis information to extend multimodal brain analysis. (2) We design an attention mechanism using a GNN that incorporates local brain networks and multimodal interactions. (3) We demonstrate the superiority of our method and highlight the brain regions linked to various brain disorders (AD and ADHD) enhancing interpretability.**
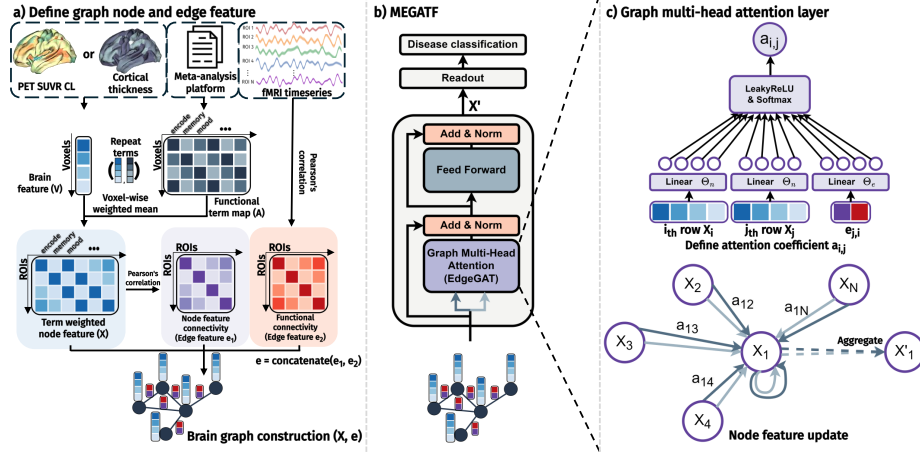
Fig. 1. Overview of the MEGATF model. (a) Brain features (SUVR CL or cortical thickness) from PET or sMRI are refined into cognition-behavior terms-weighted node features via meta-analysis. The connectivity between node features is calculated from the term-weighted node features and standard functional connectivity is calculated from functional MRI timeseries. Both are used to construct the edge features. (b) A standard transformer encoder layer is used for MEGATF and the edge graph attention network employs a multi-head attention mechanism. (c) Node and edge features were used to define attention coefficients (upper). Multi-head attention gathers neighbor information, aggregates features and linearly projects them to update node features. Arrow colors indicate independent attention computations (bottom).

2 Method

2.1 Overview of the method

We used preprocessed brain images of PET or sMRI and parcellated them into N regions of interest (ROIs) based on a given atlas. Subsequently, we construct the brain features, $V \in R^{N \times 1}$, such as a centiloid scale of the standardized uptake value ratio (SUVR CL) from PET or cortical thickness obtained from T1-weighted sMRI. We also obtain standardized voxel-level activation maps A_d for d-th distinct cognition-behavior terms (e.g., memory) from the meta-analysis platform of *Neurosynth* [5]. We further refine regional brain features by the weighted sum of the voxel-level term activation maps to obtain the graph node feature, $X \in R^{N \times D}$. Additionally, we construct a node feature connectivity matrix, $e^1 \in R^{N \times N}$ and functional connectivity matrix, $e^2 \in R^{N \times N}$ by calculating the Pearson's correlation coefficients between the term-weighted node features of pairs of ROIs and the fMRI timeseries of pairs of ROIs, respectively. The edge feature e is constructed by concatenating e^1 and e^2. The node feature X and edge feature e serve as inputs to the graph multi-head attention module of MEGATF to model the local brain networks. The output from MEGATF (X')

is passed through a readout layer followed by multi-layer perceptrons (MLPs) to perform a classification task (**Fig. 1**).

2.2   Constructing the graph structure

Generating brain features. We define brain features using SUVR CL from PET and cortical thickness from T1 sMRI. The SUVR CL is derived using the cerebellum as the reference region, by dividing the PET voxel values by the mean value of the cerebellum. Two types of PET images, AV45-PET and FBB-PET, are used for analysis. The measurements are standardized using a well-established formula to ensure comparability between the two types of PET images [15]. Cortical thickness is standardized using the mean and variance of the values from normal individuals, a method that has been successfully used in neurodegenerative disease research to emphasize disease-related variations [8].

Obtaining cognition-behavior terms activation maps from meta-analysis. We obtain activation maps from *Neurosynth*, which aggregates brain activation maps from 14,371 studies [26]. *Neurosynth* provides voxel-wise z-score maps for 1,334 cognition-behavior terms to assess the distribution of brain activation. A previous study narrowed down the 1,334 terms to 334 [9], and we further select D = 64 terms related to AD and ADHD.

Generating term-weighted node features. To enrich the information on the brain regions, we integrated the meta-analysis information with the brain features. Based on the given atlas, the entire brain is composed of N ROIs, where each region has many voxels. Conventionally, the i-th ROI in the brain feature V or the term activation map A_d for the d-th cognition-behavior terms is defined as the mean value of the voxels in the ROI. However, the term activation map is a voxel-wise map with finer granularity. Considering the spatially varying weights of the activation map, we define the weighted brain feature for the i-th ROI and the d-th cognitive term X_{id} using the voxel-wise weighted mean of the cognitive map with the brain feature matrix V, where K is the number of voxels in a given ROI. Finally, we aggregate each term and region obtaining $X \in R^{N \times D}$ (**Fig. 1(a)**).

$$X_{id} = \frac{\sum_{k=1}^{K}(\text{voxel}_{i,k}^{A_d} \cdot \text{voxel}_{i,k}^{V})}{\sum_{k=1}^{K}\text{voxel}_{i,k}^{A_d}} \quad i \in 1, 2, ..., N, \quad d \in 1, 2, ..., D \qquad (1)$$

Generating edge features. We define the interactions between a pair of ROIs in different modalities as follows: First, we measure the relationship between term-weighted node features. The ij-th element of the term-weighted node feature connectivity (NFC) matrix e^1 is defined by calculating the Pearson's correlation coefficient between the i-th row of X (X_i) and the j-th row (X_j), where $i, j \in 1, 2, ..., N$. Second, to obtain spatial perspective, we measure the relationship

between the fMRI timeseries of a pair of ROIs, which is noted as functional connectivity (FC). Given that fMRI timeseries $\in R^{N \times t}$, where t represents the number of timepoints across N ROIs, the ij-th element of the FC matrix e^2 is defined by calculating Pearson's correlation coefficient between timeseries$_i$ and timeseries$_j$, where $i, j \in 1, 2, ..., N$. Finally, we concatenate the term-weighted NFC matrix e^1 and FC matrix e^2 to define the edge feature $e \in R^{N \times N \times 2}$ (**Fig. 1(a)**).

2.3 Graph-based attention transformer

Graph multi-head attention. We propose a graph-based multi-head attention mechanism to the standard transformer encoder layer (**Fig. 1(b)**), applying self-attention to graph structures and updating node representations by attending to neighbors. Each node computes its hidden representation by dynamically weighing its neighbors through a self-attention strategy [24]. The edges between node pairs contribute to this process, assigning specific weights to neighboring connections and thereby enhancing information propagation. In brain network analysis, the nodes represent brain regions and the edges denote interactions between them. The graph multi-head attention mechanism is implemented using the edge graph attention layer (EdgeGAT) [21], where the term-weighted node feature X and edge feature e serve as inputs. Specifically, EdgeGAT$(X, e) = \Theta_s \cdot X_i + \sum_{Head=1}^{h} \sum_{j=1}^{N} a_{j,i}^h (\Theta_n^h \cdot X_j^h + \Theta_e^h \cdot e_{j,i})$, where $\Theta_s, \Theta_n, \Theta_e$ are learnable weight matrices for node features (self and neighbor) and edge features and a is attention coefficient between the nodes X_j and X_i. Attention coefficients are defined as $a_{j,i} = \text{softmax}(\text{LeakyReLU}(\alpha^T[\Theta_n \cdot X_i \| \Theta_n \cdot X_j \| \Theta_e \cdot e_{j,i}]))$, where α is a learnable vector (**Fig. 1(c)**). Finally, the output of MEGATF, X' is calculated.

Readout layer and loss function. To efficiently handle the graph structure, we employ the OCRead layer, which is specifically designed for graph data [14]. The outputs from the OCRead layer are subsequently passed through the MLPs to perform the classification task, with the cross-entropy loss used for prediction.

3 Experiments

3.1 Datasets and experimental settings

Dataset. We evaluated our method using the publicly available AD dataset from the AD Neuroimaging Initiative (ADNI) [25], and the ADHD dataset from the Healthy Brain Network (HBN) [1]. We selected individuals who had both PET and fMRI data from the ADNI 1, 2, 3, and GO phases, T1-weighted sMRI, and fMRI data from the HBN. MRI data were preprocessed using fMRIprep [12], and PET images were preprocessed following conventional steps [13]. Brain regions were parcellated into 200 regions using the Schafer atlas [22]. We performed three classification tasks related to diagnosis or prognosis. Task 1 was the classification of individuals with MCI and cognitively normal (CN) individuals (ADNI: 117

Table 1. Diagnosis prediction results for MCI vs. CN (task1) (Mean±standard deviation [SD]). **Bold** indicates the best performance. Each model's top result is <u>underlined</u>.

Task 1					
Model	**Input types**	**Accuracy**	**AUROC**	**Sensitivity**	**Specificity**
BrainNetGNN (MIA21) [17]	SUVR CL	<u>68.8±2.4</u>	73.4±5.3	<u>72.2±8.1</u>	65.9±5.9
	FC	60.0±6.9	67.2±9.3	57.4±6.4	62.2±10.3
	Term weighted	64.4±6.1	<u>76.1±4.0</u>	43.5±18.2	<u>82.2±9.5</u>
BNT (NeurIPS22) [14]	SUVR CL	72.0±4.0	81.5±2.2	67.0±10.5	76.3±6.5
	FC	72.0±7.3	<u>85.9±5.0</u>	51.3±22.9	<u>89.6±8.3</u>
	Term weighted	<u>75.2±7.7</u>	85.2±6.4	<u>72.2±9.8</u>	77.8±14.2
Com-BrainTF (MICCAI23) [3]	SUVR CL	69.2±8.1	81.3±2.7	60.9±32.2	76.3±13.0
	FC	<u>75.6±4.1</u>	85.3±4.1	60.0±15.9	<u>88.9±8.4</u>
	Term weighted	74.0±3.8	<u>85.5±4.7</u>	<u>68.7±14.9</u>	78.5±15.8
Graphormer (NeurIPS21) [27]	SUVR CL	73.2±3.5	81.6±2.5	64.4±6.4	80.7±7.6
	FC	75.2±7.4	85.4±6.0	63.5±16.6	<u>85.2±12.2</u>
	Term weighted	<u>77.6±3.2</u>	<u>85.7±4.5</u>	<u>72.2±6.5</u>	82.2±9.5
Ours	SUVR CL	73.6±4.1	82.3±3.8	67.0±3.5	79.3±6.5
	FC	70.4±11.7	85.9±5.5	53.9±33.3	**84.4±8.9**
	Term weighted	**78.0±3.6**	**86.6±4.0**	**71.3±2.1**	83.7±5.0

MCI and 132 CN), task 2 was AD conversion prediction (i.e., ADNI: 39 stable MCI [sMCI] and 47 MCI progressing to AD [pMCI]), and task 3 was classification of individuals with ADHD and CN (HBN: 86 ADHD and 68 CN). Additional 30 CN individuals from the HBN were used to standardize cortical thickness.

Implementation details. All models were implemented in PyTorch and trained using NVIDIA RTX 4070 TI (12GB). The transformer layer had two attention heads with Adam optimization (learning rate=10^{-4}, weight decay=10^{-4}). The batch size was set to 32 and a stratified five-fold cross-validation was applied over 50 epochs, selecting the highest area under the receiver operator curve (AUROC) model per fold and reporting the average of five folds.

3.2   Experimental results

Comparison with state-of-the-art methods. The quantitative diagnostic and prognostic results are presented in **Tables 1 and 2**. Our model achieved the highest accuracy and AUROC, compared with the other models, for all three tasks, with marginal improvements in sensitivity and specificity. Term-weighted node features tended to outperform other inputs like SUVR CL or cortical thickness, and FC. Brain features were converted into a matrix using Gaussian similarity function, $\exp(-\frac{\|v_i - v_j\|^2}{2\sigma^2})$, where v is the specific feature of a given brain region. Unlike term-weighted node features, brain features and FC used identical values for both the node and edge features. These results emphasize the benefits of meta-analysis information and graph multi-head attention.

Table 2. Prediction results for sMCI vs. pMCI (task2) and ADHD vs. CN (task3) (Mean±SD). The brain feature used in task2 is SUVR CL and task3 uses cortical thickness. **Bold** indicates the best performance. Each model's top result is <u>underlined</u>.

| Model	Input types	Task 2		Task 3	
		Accuracy	AUROC	Accuracy	AUROC
BrainNetGNN (MIA21) [17]	Brain feature	65.6±6.5	73.8±10.4	59.4±6.3	59.9±11.6
	FC	58.9±9.0	72.5±13.4	<u>61.3±7.1</u>	59.5±9.1
	Term weighted	<u>77.8±9.3</u>	<u>91.5±8.5</u>	56.8±3.3	<u>60.0±3.3</u>
BNT (NeurIPS22) [14]	Brain feature	80.0±6.7	87.8±7.5	57.4±2.4	57.4±5.8
	FC	76.7±11.3	84.0±5.8	<u>60.6±4.3</u>	62.8±12.0
	Term weighted	<u>85.6±7.5</u>	<u>91.8±8.4</u>	60.0±3.9	<u>63.6±2.7</u>
Com-BrainTF (MICCAI23) [3]	Brain feature	78.9±8.2	87.0±6.1	55.5±2.4	58.4±4.6
	FC	80.0±9.0	85.8±6.8	55.5±5.2	61.6±10.1
	Term weighted	<u>84.4±6.5</u>	<u>92.8±5.7</u>	<u>56.8±2.6</u>	61.8±6.8
Graphormer (NeurIPS21) [27]	Brain feature	71.1±8.9	87.2±6.9	<u>58.1±4.1</u>	59.2±7.6
	FC	77.8±8.6	88.5±5.6	58.1±4.6	55.7±8.6
	Term weighted	<u>85.6±6.7</u>	<u>91.3±8.3</u>	57.4±4.7	<u>61.1±4.6</u>
Ours	Brain feature	81.1±8.3	86.2±7.2	56.8±5.6	60.8±4.7
	FC	76.7±6.5	86.3±8.5	61.3±4.1	61.6±8.0
	Term weighted	**87.8±9.6**	**93.5±7.8**	**61.9±9.0**	**63.8±7.0**
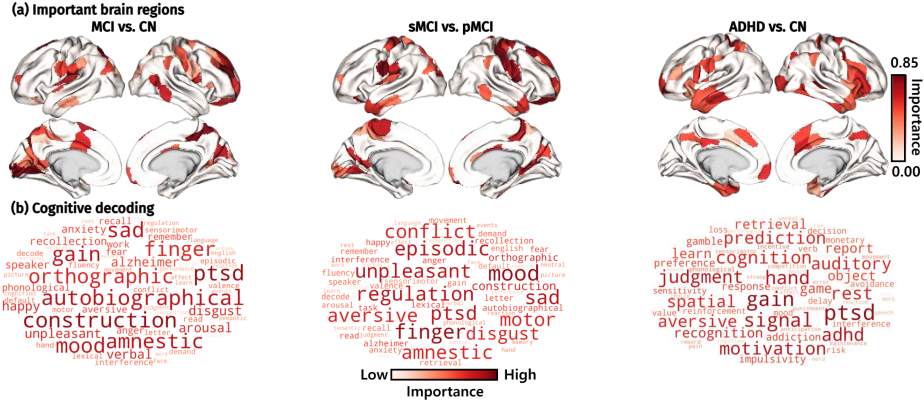
Fig. 2. Interpretation of the important brain regions and terms. (a) shows the top 25% important brain regions in the three tasks. (b) shows the important terms.

Interpretation of the ROI and term-level importance. To assess the key ROIs and cognition-behavior terms in the prediction, we visualized the important regions and terms using the final layer features. **Fig. 2(a)** shows the top 25% important regions. In tasks 1 and 2, cingulate gyrus, occipital cortex, temporal gyrus, preforntal, limbic, and motor cortex were identified related to AD prediction and prognosis [4,5,28]. In ADHD prediction (task 3), frontoparietal, temporal gyrus, limbic, and cingulate regions were important [16,18]. **Fig. 2(b)** shows the top terms. In task 1, memory, cognition, and emotion-related terms

Table 3. Effect of different combinations of edge features on the prediction performance in ablation studies. **Bold** indicates the best performance.

Methods	Task 1		Task 2		Task 3	
	Accuracy	AUROC	Accuracy	AUROC	Accuracy	AUROC
w/o edge feature	73.6±7.3	86.1±4.4	82.2±6.5	93.3±3.8	59.4±4.4	61.3±4.8
One edge feature (FC)	73.6±3.9	84.2±5.8	83.3±7.0	92.8±4.1	56.1±2.6	60.1±9.5
One edge feature (NFC)	72.8±5.6	84.4±3.5	83.3±11.7	93.2±4.9	58.7±2.4	59.0±5.4
Our method	**78.0±3.6**	**86.6±4.0**	**87.8±9.6**	**93.5±7.8**	**61.9±9.0**	**63.8±7.0**
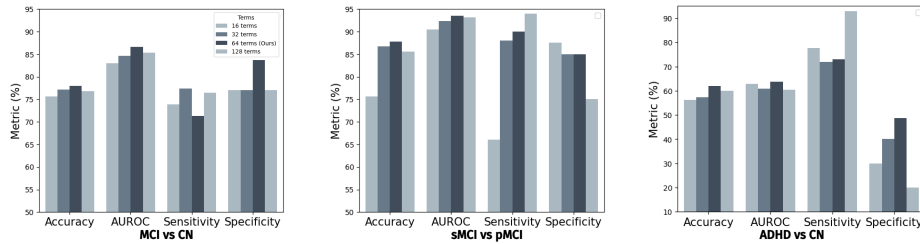
Fig. 3. Effect of the number of cognitive terms on the prediction. Comparison of results based on the number of terms used to create the node features.

like "amnestic", "ptsd", and "sad" were crucial, known to be affected by AD [4,5]. Similar process-related terms appeared in task 2. Task 3 revealed terms linked to emotion regulation and decision-making, aligning with ADHD studies [18].

Ablation study. We performed ablation studies about the effects of edge features and number of terms. **Table 3** highlights the performance variations based on the edge feature type, showing that removing edge features or using only FC or term-weighted feature connectivity individually resulted in worse performance, compared with our method. That is, combining both types of edge features enhanced the prediction performance. In summary, we observed that both FC derived from regional interactions in fMRI and term-weighted feature connectivity derived from the meta-analysis contributed to improvements in the prediction performance. **Fig. 3** shows the prediction performance based on the number of terms used. The accuracy and AUROC increased monotonically up to 64 terms and then decreased, demonstrating the effectiveness of using 64 terms.

4   Conclusion

We proposed MEGATF, a meta-analysis-guided method for multimodal brain analysis. It achieved superior performance in classifying individuals with brain disorders while identifying explainable and clinically relevant brain regions. By leveraging meta-analysis information, MEGATF linked brain feature maps to AD

or ADHD related cognitive functions. A key strength is using graph-based multi-head attention to capture local brain network interactions. Given its success in AD and ADHD, MEGATF has potential to be applied to other brain disorders.

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

References

1. Alexander, L.M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., Vega-Potler, N., Langer, N., Alexander, A., Kovacs, M., et al.: An open resource for transdiagnostic research in pediatric mental health and learning disorders. Scientific data **4**(1), 1–26 (2017)
2. Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D.: Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. Neuroimage **145**, 137–165 (2017)
3. Bannadabhavi, A., Lee, S., Deng, W., Ying, R., Li, X.: Community-aware transformer for autism prediction in fmri connectome. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 287–297. Springer (2023)
4. Brewer, A.A., Barton, B.: Visual cortex in aging and alzheimer's disease: changes in visual field maps and population receptive fields. Frontiers in psychology **5**, 74 (2014)
5. Cerami, C., Della Rosa, P.A., Magnani, G., Santangelo, R., Marcone, A., Cappa, S.F., Perani, D.: Brain metabolic maps in mild cognitive impairment predict heterogeneity of progression to dementia. NeuroImage: Clinical **7**, 187–194 (2015)
6. Choi, H., Byeon, K., Lee, J.e., Hong, S.J., Park, B.y., Park, H.: Identifying subgroups of eating behavior traits unrelated to obesity using functional connectivity and feature representation learning. Human Brain Mapping **45**(1), e26581 (2024)
7. Choi, H., Byeon, K., Park, B.y., Lee, J.e., Valk, S.L., Bernhardt, B., Di Martino, A., Milham, M., Hong, S.J., Park, H.: Diagnosis-informed connectivity subtyping discovers subgroups of autism with reproducible symptom profiles. NeuroImage **256**, 119212 (2022)
8. Chung, J., Yoo, K., Lee, P., Kim, C.M., Roh, J.H., Park, J.E., Kim, S.J., Seo, S.W., Shin, J.H., Seong, J.K., et al.: Normalization of cortical thickness measurements across different t1 magnetic resonance imaging protocols by novel w-score standardization. Neuroimage **159**, 224–235 (2017)
9. Collins, E., Chishti, O., Obaid, S., McGrath, H., King, A., Shen, X., Arora, J., Papademetris, X., Constable, R.T., Spencer, D.D., et al.: Mapping the structure-function relationship along macroscale gradients in the human brain. Nature Communications **15**(1), 7063 (2024)

10. 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)

11. Dockès, J., Poldrack, R.A., Primet, R., Gözükan, H., Yarkoni, T., Suchanek, F., Thirion, B., Varoquaux, G.: Neuroquery, comprehensive meta-analysis of human brain mapping. elife **9**, e53385 (2020)

12. Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., et al.: fmriprep: a robust preprocessing pipeline for functional mri. Nature methods **16**(1), 111–116 (2019)

13. Jagust, W.J., Landau, S.M., Koeppe, R.A., Reiman, E.M., Chen, K., Mathis, C.A., Price, J.C., Foster, N.L., Wang, A.Y.: The alzheimer's disease neuroimaging initiative 2 pet core: 2015. Alzheimer's & Dementia **11**(7), 757–771 (2015)

14. Kan, X., Dai, W., Cui, H., Zhang, Z., Guo, Y., Yang, C.: Brain network transformer. Advances in Neural Information Processing Systems **35**, 25586–25599 (2022)

15. Klunk, W.E., Koeppe, R.A., Price, J.C., Benzinger, T.L., Devous Sr, M.D., Jagust, W.J., Johnson, K.A., Mathis, C.A., Minhas, D., Pontecorvo, M.J., et al.: The centiloid project: standardizing quantitative amyloid plaque estimation by pet. Alzheimer's & dementia **11**(1), 1–15 (2015)

16. Krain, A.L., Castellanos, F.X.: Brain development and adhd. Clinical psychology review **26**(4), 433–444 (2006)

17. 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)

18. Lin, H.Y., Tseng, W.Y.I., Lai, M.C., Matsuo, K., Gau, S.S.F.: Altered resting-state frontoparietal control network in children with attention-deficit/hyperactivity disorder. Journal of the International Neuropsychological Society **21**(4), 271–284 (2015)

19. Marcus, C., Mena, E., Subramaniam, R.M.: Brain pet in the diagnosis of alzheimer's disease. Clinical nuclear medicine **39**(10), e413–e426 (2014)

20. McAlonan, G.M., Cheung, V., Cheung, C., Chua, S.E., Murphy, D.G., Suckling, J., Tai, K.S., Yip, L.K., Leung, P., Ho, T.P.: Mapping brain structure in attention deficit-hyperactivity disorder: a voxel-based mri study of regional grey and white matter volume. Psychiatry Research: Neuroimaging **154**(2), 171–180 (2007)

21. Monninger, T., Schmidt, J., Rupprecht, J., Raba, D., Jordan, J., Frank, D., Staab, S., Dietmayer, K.: Scene: Reasoning about traffic scenes using heterogeneous graph neural networks. IEEE Robotics and Automation Letters **8**(3), 1531–1538 (2023)

22. Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.: Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. Cerebral cortex **28**(9), 3095–3114 (2018)

23. Sepulcre, J., Liu, H., Talukdar, T., Martincorena, I., Yeo, B.T., Buckner, R.L.: The organization of local and distant functional connectivity in the human brain. PLoS computational biology **6**(6), e1000808 (2010)

24. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: International Conference on Learning Representations (2018), https://openreview.net/forum?id=rJXMpikCZ

25. Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E., et al.: The alzheimer's disease neuroimaging initiative: a review of papers published since its inception. Alzheimer's & Dementia **9**(5), e111–e194 (2013)

26. Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D.: Large-scale automated synthesis of human functional neuroimaging data. Nature methods **8**(8), 665–670 (2011)
27. Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., Shen, Y., Liu, T.Y.: Do transformers really perform badly for graph representation? Advances in neural information processing systems **34**, 28877–28888 (2021)
28. Yu, E., Liao, Z., Mao, D., Zhang, Q., Ji, G., Li, Y., Ding, Z.: Directed functional connectivity of posterior cingulate cortex and whole brain in alzheimer's disease and mild cognitive impairment. Current Alzheimer Research **14**(6), 628–635 (2017)