



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Joint Imaging–ROI Representation Learning via Cross-View Contrastive Alignment for Brain Disorder Classification

Wei Liang¹[0009–0007–5618–8961] and Lifang He¹,✉[0000–0001–7810–9071]

Lehigh University, Bethlehem, PA 18015, USA
{wel425, lih319}@lehigh.edu

Abstract. Brain imaging classification is commonly approached from two perspectives: modeling the full image volume to capture global anatomical context, or constructing ROI-based graphs to encode localized and topological interactions. Although both representations have demonstrated independent efficacy, their relative contributions and potential complementarity remain insufficiently understood. Existing fusion approaches are typically task-specific and do not enable controlled evaluation of each representation under consistent training settings. To address this gap, we propose a unified cross-view contrastive framework for joint imaging–ROI representation learning. Our method learns subject-level global (imaging) and local (ROI-graph) embeddings and aligns them in a shared latent space using a bidirectional contrastive objective, encouraging representations from the same subject to converge while separating those from different subjects. This alignment produces comparable embeddings suitable for downstream fusion and enables systematic evaluation of imaging-only, ROI-only, and joint configurations within a unified training protocol. Extensive experiments on the ADHD-200 and ABIDE datasets demonstrate that joint learning consistently improves classification performance over either branch alone across multiple backbone choices. Moreover, interpretability analyses reveal that imaging-based and ROI-based branches emphasize distinct yet complementary discriminative patterns, explaining the observed performance gains. These findings provide principled evidence that explicitly integrating global volumetric and ROI-level representations is a promising direction for neuroimaging-based brain disorder classification. The source code is available at <https://github.com/vveii14/circa-brain>.

Keywords: Neuroimaging · Brain Disorder Classification · Graph Neural Networks · Contrastive Learning · Imaging–Graph Fusion

1 Introduction

Neuroimaging has become an essential tool for studying neurological and neurodevelopmental disorders, enabling data-driven diagnosis and characterization of disease-related brain alterations [9,29]. With advances in deep learning, brain

imaging classification methods have increasingly focused on learning discriminative representations directly from volumetric data [1]. Despite substantial progress, a fundamental question remains open: **how should brain imaging data be represented for optimal classification performance?**

Existing approaches largely follow two distinct paradigms. The first models the entire 3D imaging volume using convolutional or transformer-based architectures to capture global anatomical context [16,18,25]. These methods leverage holistic spatial information but may overlook fine-grained inter-regional relationships.

The second paradigm constructs region-of-interest (ROI) graphs, where nodes correspond to predefined brain regions and edges encode structural or functional relationships [6,20,24,26]. ROI-based graph modeling emphasizes localized topology and inter-regional interactions that are often clinically meaningful. While both paradigms have demonstrated effectiveness in various brain disorder classification tasks [2,9], they capture fundamentally different aspects of brain organization.

Although imaging-based and ROI-based representations have been studied extensively in isolation [15,20,25], their relative contributions and potential complementarity remain poorly characterized. Prior fusion approaches typically integrate modalities within customized architectures tailored to specific tasks [8,21]. Such designs often confound architectural differences with representational benefits, making it difficult to determine whether performance improvements stem from genuine representational synergy or from variations in model complexity and training strategies. A principled and controlled evaluation framework is therefore needed to clarify how global and ROI-level representations contribute individually and jointly to classification performance.

To address this limitation, we introduce a unified framework for joint imaging-ROI representation learning via cross-view contrastive alignment. Our approach employs two modular encoders to extract subject-level global imaging embeddings and local ROI-graph embeddings. A bidirectional cross-view contrastive objective aligns these heterogeneous representations in a shared latent space, encouraging consistency across views while preserving discriminative capacity within each branch. This alignment yields comparable embeddings suitable for downstream fusion, enabling systematic comparison of imaging-only, ROI-only, and joint configurations under identical training conditions.

Extensive experiments on the ADHD-200 and ABIDE datasets demonstrate that joint learning consistently outperforms single-branch baselines across multiple encoder choices. Beyond quantitative gains, interpretability analyses reveal that the two branches focus on distinct yet complementary neuroanatomical patterns, providing insight into their synergistic effects. Collectively, our findings establish that explicitly aligning and integrating global volumetric and ROI-level graph representations offers a principled and effective strategy for brain disorder classification. Our contributions are threefold:

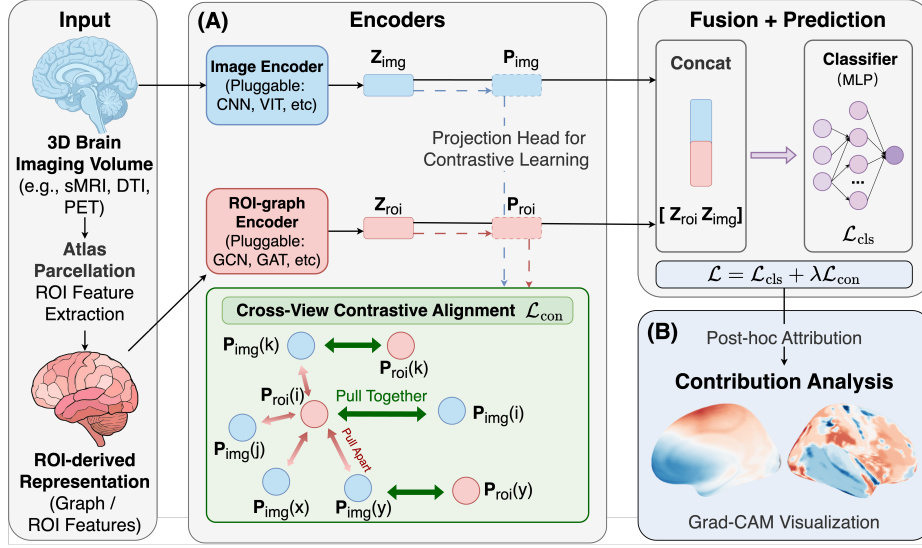


Fig. 1: **Overview of the proposed joint imaging–ROI representation learning framework.** (A) Joint representation learning via cross-view contrastive alignment. Subject-level global imaging and local ROI-graph embeddings are extracted by modular encoders and aligned in a shared latent space to enable downstream fusion and classification. (B) Post-hoc contribution analysis. Branch-specific attribution maps quantify complementary discriminative patterns captured by the imaging and ROI-graph representations.

- We propose a unified cross-view contrastive framework for joint modeling of volumetric imaging and ROI-based graph representations under consistent training settings.
- We provide a systematic and controlled evaluation of imaging-only, ROI-only, and joint configurations, clarifying their individual and complementary contributions.
- Through experiments and interpretability analyses, we demonstrate that aligned imaging–ROI fusion yields consistent and complementary benefits for brain disorder classification.

2 Methodology

We propose a unified framework for joint imaging–ROI representation learning through cross-view contrastive alignment. As illustrated in Fig. 1, the framework consists of three components: (i) extraction of global imaging and local ROI-graph representations, (ii) cross-view contrastive alignment in a shared latent space, and (iii) downstream fusion and classification.

2.1 Imaging and ROI-Graph Representation Learning

Given a subject-level imaging volume $\mathbf{x}_i$, we construct two complementary representations: a global volumetric representation and a local ROI-graph representation.

The imaging embedding is obtained via an encoder $f_{\text{img}}(\cdot)$:

$$\mathbf{z}_{\text{img}}^{(i)} = f_{\text{img}}(\mathbf{x}_i), \quad \text{where } \mathbf{z}_{\text{img}}^{(i)} \in \mathbb{R}^{d_{\text{img}}}. \quad (1)$$

To model inter-regional relationships, we construct an ROI graph $\mathcal{G}(\mathbf{x}_i)$ using AAL atlas parcellation [27]. Node features correspond to mean voxel intensities within each ROI, and edges are defined by pairwise Pearson correlations between regions, yielding a subject-specific adjacency matrix. The ROI embedding is then computed via a graph encoder $f_{\text{roi}}(\cdot)$:

$$\mathbf{z}_{\text{roi}}^{(i)} = f_{\text{roi}}(\mathcal{G}(\mathbf{x}_i)), \quad \text{where } \mathbf{z}_{\text{roi}}^{(i)} \in \mathbb{R}^{d_{\text{roi}}}. \quad (2)$$

In practice, we instantiate f_{img} using 3DSC-TF [25], a hybrid CNN–Transformer architecture designed for 3D brain imaging, and f_{roi} using NeuroGraph [26], a graph neural network framework developed for brain connectome analysis. These models serve as strong and representative backbones for volumetric and ROI-based learning, respectively. Importantly, our framework is modular and agnostic to specific encoder choices. Alternative backbone choices are evaluated in Sec. 3.

2.2 Cross-View Contrastive Alignment

To align heterogeneous representations from imaging and ROI views, we adopt a bidirectional InfoNCE objective for cross-view contrastive learning [5]. Given subject-level imaging and ROI embeddings $\mathbf{z}_{\text{img}}^{(i)}$ and $\mathbf{z}_{\text{roi}}^{(i)}$, two projection heads $g_{\text{img}}(\cdot)$ and $g_{\text{roi}}(\cdot)$ map them into a shared latent space:

$$\mathbf{p}_{\text{img}}^{(i)} = g_{\text{img}}(\mathbf{z}_{\text{img}}^{(i)}), \quad \mathbf{p}_{\text{roi}}^{(i)} = g_{\text{roi}}(\mathbf{z}_{\text{roi}}^{(i)}). \quad (3)$$

For a mini-batch of size B , we compute cross-view similarities as:

$$S_{ij} = \frac{\text{sim}(\mathbf{p}_{\text{img}}^{(i)}, \mathbf{p}_{\text{roi}}^{(j)})}{\tau}, \quad (4)$$

where τ is a temperature parameter and $\text{sim}(\cdot, \cdot)$ denotes cosine similarity.

Using these similarities, the bidirectional InfoNCE loss [5] is defined as:

$$\mathcal{L}_{\text{con}} = -\frac{1}{2B} \sum_{i=1}^B \left[\log \frac{\exp(S_{ii})}{\sum_{j=1}^B \exp(S_{ij})} + \log \frac{\exp(S_{ii})}{\sum_{j=1}^B \exp(S_{ji})} \right]. \quad (5)$$

where S_{ii} corresponds to positive cross-view pairs from the same subject, and S_{ij} for $i \neq j$ correspond to negative pairs across different subjects. This formulation promotes cross-view alignment of representations from the same subject while preserving inter-subject discrimination.

2.3 Fusion and Classification Objective

The aligned embeddings are concatenated to form a joint representation:

$$\mathbf{z}_{\text{fuse}}^{(i)} = [\mathbf{z}_{\text{img}}^{(i)}; \mathbf{z}_{\text{roi}}^{(i)}].$$

A classifier $h(\cdot)$ maps the fused representation to class logits $\hat{y}_i$. The overall training objective combines the cross-entropy classification loss and the contrastive loss:

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{con}} \quad (6)$$

where $\mathcal{L}_{\text{cls}}$ is the standard cross-entropy loss for classification, and λ controls the trade-off between discriminative supervision and cross-view alignment.

3 Experiments and Results

We evaluate the proposed joint imaging–ROI framework along four dimensions: (i) comparison between single-branch (imaging-only, ROI-only) and joint learning; (ii) ablations on encoder choices and fusion mechanisms; (iii) robustness under missing-view settings; and (iv) interpretability analyses to quantify branch-specific contributions and complementarity.

Dataset Description We conducted experiments on two public structural MRI datasets: ADHD-200 [22] and the Autism Brain Imaging Data Exchange (ABIDE) [7]. ADHD-200 includes 776 subjects (ADHD vs. typically developing controls, TDC), and ABIDE contains 1112 subjects (autism spectrum disorder vs. TDC). For both cohorts, we used the officially released fMRIPrep-preprocessed imaging data [10], in which the anatomical volumes were skull-stripped and normalized to MNI space, then cropped and resized to $96 \times 96 \times 96$. Performance is evaluated using accuracy (Acc), area under the ROC curve (AUC), and F1-score (F1). All results are reported as mean $\pm$ standard deviation over 5-fold cross-validation, stratified by diagnostic label and split at the subject level to prevent data leakage.

Single-Branch vs. Joint Learning To establish strong single-branch baselines and evaluate the unified framework, we employ three 3D imaging encoders (ViT3D* [19], RAE-ViT [15], and 3DSC-TF [25]), together with two ROI-based models (DNN [4] and NeuroGraph [26]). Comparisons are conducted across ROI-only, imaging-only, and joint configurations. For a fair comparison, all models were tuned with Optuna.

As shown in Table 1, joint learning consistently outperforms both imaging-only and ROI-only baselines on ADHD-200 and ABIDE across backbone choices. These improvements indicate that global volumetric features and ROI-level graph representations capture complementary information. Importantly, the performance gains persist across multiple imaging encoders, suggesting that the proposed framework is backbone-agnostic rather than tailored to a specific architecture. We further validated the gains using paired t -test on Acc and DeLong’s test on AUC, confirming that the joint model significantly outperforms both

Table 1: Single-branch vs. joint learning performance on ADHD-200 and ABIDE.

Model	Acc	AUC	F1
ADHD			
<i>ROI-only</i>			
DNN	64.77 ± 4.88	67.73 ± 5.55	63.40 ± 5.60
NeuroGraph	63.48 ± 2.74	65.51 ± 3.23	50.25 ± 6.62
<i>Imaging-only</i>			
ViT3D*	66.84 ± 2.63	66.57 ± 6.19	64.62 ± 4.21
RAE-ViT	64.26 ± 4.05	63.99 ± 7.37	61.75 ± 5.25
3DSC-TF	68.65 ± 4.99	70.92 ± 4.93	68.42 ± 4.85
<i>Joint</i>			
NeuroGraph + ViT3D*	68.13 ± 1.98	69.70 ± 2.49	67.25 ± 1.72
NeuroGraph + RAE-ViT	67.35 ± 4.86	69.76 ± 7.72	66.85 ± 5.24
NeuroGraph + 3DSC-TF	69.29 ± 4.44	72.73 ± 4.17	69.01 ± 4.51
ABIDE			
<i>ROI-only</i>			
DNN	57.24 ± 4.44	62.84 ± 3.90	56.02 ± 4.93
NeuroGraph	61.09 ± 1.23	60.90 ± 3.41	59.52 ± 1.26
<i>Imaging-only</i>			
ViT3D*	56.94 ± 0.95	54.51 ± 2.38	55.57 ± 1.95
RAE-ViT	57.95 ± 2.51	56.78 ± 2.06	55.75 ± 3.43
3DSC-TF	59.17 ± 2.39	57.91 ± 2.16	58.33 ± 3.12
<i>Joint</i>			
NeuroGraph + ViT3D*	58.25 ± 2.68	58.15 ± 3.49	57.04 ± 3.62
NeuroGraph + RAE-ViT	60.38 ± 3.12	60.44 ± 4.56	59.55 ± 2.75
NeuroGraph + 3DSC-TF	62.54 ± 1.79	64.08 ± 2.14	61.71 ± 1.43

single-branch baselines across datasets ($p < 0.05$). These observations motivate the encoder selection for subsequent fusion ablations.

Encoder and Fusion Ablation To further analyze architectural choices, we fix the imaging encoder to the strongest-performing backbone (3DSC-TF) and evaluate two ROI encoders (DNN and NeuroGraph) under three fusion strategies: simple concatenation (Concat), bidirectional cross-attention (Cross-attn), and the proposed contrastive alignment (Contra).

Table 2 reveals two consistent findings. First, NeuroGraph consistently outperforms DNN across datasets and fusion mechanisms, indicating that graph-based message passing more effectively captures inter-regional connectivity than independent node-level features. Second, contrastive alignment achieves the best overall performance in most settings, demonstrating that explicitly aligning heterogeneous representations in a shared latent space produces more compatible embeddings for downstream classification.

Table 2: Encoder and joint learning ablation on ADHD-200 and ABIDE.

Imaging Encoder	ROI Encoder	Fusion	Acc	AUC	F1
ADHD					
3DSC-TF	DNN	Concat	59.87 ± 4.07	66.73 ± 3.56	58.62 ± 3.94
		Cross-attn	60.65 ± 7.63	66.49 ± 4.62	59.11 ± 8.77
		Contra	62.84 ± 4.20	68.69 ± 4.62	61.06 ± 5.79
	NeuroGraph	Concat	66.32 ± 4.53	65.94 ± 7.88	64.17 ± 6.23
		Cross-attn	68.00 ± 4.23	69.53 ± 5.94	67.71 ± 4.32
		Contra	69.29 ± 4.44	72.73 ± 4.17	69.01 ± 4.51
ABIDE					
3DSC-TF	DNN	Concat	56.24 ± 2.78	60.24 ± 3.15	51.61 ± 9.64
		Cross-attn	57.84 ± 6.60	59.35 ± 5.98	54.37 ± 4.09
		Contra	57.85 ± 4.16	62.71 ± 5.26	57.35 ± 4.04
	NeuroGraph	Concat	60.89 ± 1.91	61.73 ± 4.47	57.85 ± 4.54
		Cross-attn	61.70 ± 5.12	64.63 ± 4.17	61.01 ± 4.83
		Contra	62.54 ± 1.79	64.08 ± 2.14	61.71 ± 1.43

Missing-View Robustness. In practical clinical scenarios, imaging modalities or ROI-derived features may be unavailable due to acquisition or preprocessing failures [12,28]. To assess robustness under missing views, we simulate view absence by zero-masking one branch for a random subset of subjects at rates of 10%, 30%, and 50% on the ADHD-200 cohort, applied consistently during both training and testing. As illustrated in Fig. 2, masking either branch leads to an initial performance drop compared to the full-view setting. However, increasing the missing rate from 10% to 50% results in only moderate additional degradation. The bounded decline suggests that per-branch supervision preserves individual discriminative capacity, while cross-view contrastive alignment facilitates implicit knowledge transfer between branches. Consequently, the remaining view can partially compensate for the missing modality.

Contribution Analysis To interpret learned decision patterns, we apply Grad-CAM [11] to derive region-level contribution maps from the test sets across 5-fold cross-validation on the ADHD-200 cohort, as shown in Fig. 3. Such attribution visualizations are widely adopted in medical imaging to assess clinical face validity [30]. Compared with unimodal models, the joint framework produces more spatially coherent contribution maps. The imaging-only branch yields relatively diffuse patterns, whereas the ROI-only branch produces sharper but more heterogeneous activations. In contrast, the joint model highlights regions consistently supported across both representations, suggesting synergistic integration.

Regions with higher contributions in the joint model predominantly involve frontal, sensorimotor, orbitofrontal, and limbic systems, which have been repeatedly implicated in ADHD by large-scale and meta-analytic neuroimaging studies. For example, prior meta-analyses report ADHD-related alterations in

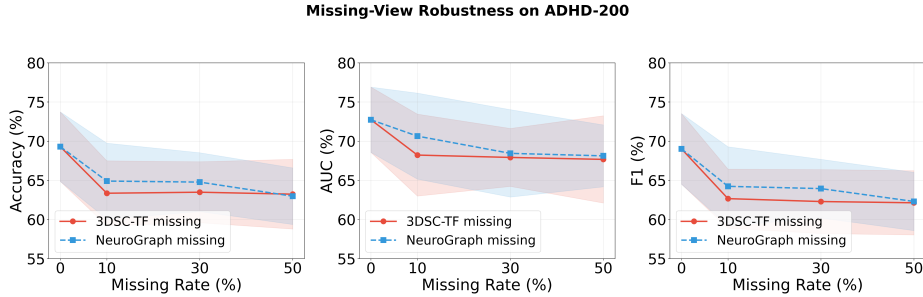


Fig. 2: Missing-view robustness on ADHD-200. Shaded regions denote ± 1 std over five-fold cross-validation.

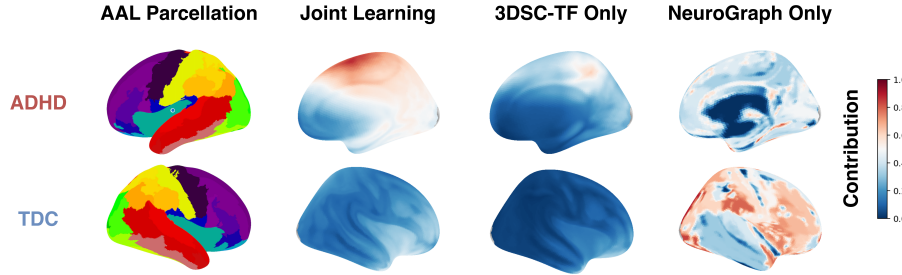


Fig. 3: Contribution maps for ADHD and TDC. The AAL parcellation (first column) provides anatomical reference; remaining columns show normalized contribution scores (0–1) for Joint Learning, 3DSC-TF Only, and NeuroGraph Only.

the superior frontal, precentral, and orbitofrontal cortices [3], as well as volumetric differences in limbic structures such as the hippocampus [14]. Functional evidence further implicates orbitofrontal and supplementary motor regions in reward processing and inhibitory control [13,17,23]. While attribution maps do not establish causal neurobiological mechanisms, the alignment between model-sensitive regions and established ADHD-related circuits supports the clinical plausibility of the learned representations and clarifies how the two branches contribute jointly.

4 Conclusion

In this work, we proposed a unified cross-view contrastive framework for joint imaging–ROI representation learning in brain disorder classification. By aligning heterogeneous embeddings in a shared latent space, the framework enables principled integration of volumetric and ROI-level graph representations. Experiments on ADHD-200 and ABIDE demonstrate consistent improvements over single-branch baselines across multiple backbones. Interpretability analyses further reveal complementary neuroanatomical patterns captured by the two repre-

sentations. These results suggest that cross-view alignment provides an effective strategy for integrating complementary brain representations in neuroimaging-based diagnosis.

Acknowledgments. This work was supported in part by NIH (R01LM013519, RF1AG077820), NSF (IIS-2319451, MRI-2215789), DOE (DE-SC0025801), and Lehigh University (CORE and RIG).

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

References

1. Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., Calhoun, V.: Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. *Nature Communications* **12**(1), 353 (2021)
2. Alp, S., Akan, T., Bhuiyan, M.S., Disbrow, E.A., Conrad, S.A., Vanchiere, J.A., Kevil, C.G., Bhuiyan, M.A.: Joint transformer architecture in brain 3d mri classification: its application in alzheimer’s disease classification. *Scientific Reports* **14**(1), 8996 (2024)
3. Chen, C., Sun, S., Chen, R., Guo, Z., Tang, X., Chen, G., Chen, P., Tang, G., Huang, L., Wang, Y.: A multimodal neuroimaging meta-analysis of functional and structural brain abnormalities in attention-deficit/hyperactivity disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **136**, 111199 (2025)
4. Chen, M., Li, H., Wang, J., Dillman, J.R., Parikh, N.A., He, L.: A multichannel deep neural network model analyzing multiscale functional brain connectome data for attention deficit hyperactivity disorder detection. *Radiology: Artificial Intelligence* **2**(1), e190012 (2019)
5. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: *International Conference on Machine Learning*. pp. 1597–1607. PMLR (2020)
6. 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)
7. Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry* **19**(6), 659–667 (2014)
8. Dong, C., Sun, D., Yu, Z., Luo, B.: Multi-view brain network classification based on adaptive graph isomorphic information bottleneck mamba. *Expert Systems with Applications* **267**, 126170 (2025)
9. Dong, K., Liang, W., Hou, T., Lu, Z., Hao, Y., Li, C., Qiu, Y., Kong, N., Cheng, Y., Wen, Y., et al.: Exploring the impact of apoe ϵ 4 on functional connectivity in alzheimer’s disease across cognitive impairment levels. *NeuroImage* **305**, 120951 (2025)
10. 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.: fmripiprep: a robust preprocessing pipeline for functional mri. *Nature Methods* **16**(1), 111–116 (2019)

11. Fayyaz, A.M., Abdulkadir, S.J., Talpur, N., Al-Selwi, S.M., Hassan, S.U., Sumiea, E.H.: Grad-cam (gradient-weighted class activation mapping): A systematic literature review. *Computers in Biology and Medicine* **198**, 111200 (2025)
12. Gao, X., Shi, F., Shen, D., Liu, M.: Multimodal transformer network for incomplete image generation and diagnosis of alzheimer's disease. *Computerized Medical Imaging and Graphics* **110**, 102303 (2023)
13. Hart, H., Radua, J., Nakao, T., Mataix-Cols, D., Rubia, K.: Meta-analysis of functional magnetic resonance imaging studies of inhibition and attention in attention-deficit/hyperactivity disorder: exploring task-specific, stimulant medication, and age effects. *JAMA Psychiatry* **70**(2), 185–198 (2013)
14. Hoogman, M., Bralten, J., Hibar, D.P., Mennes, M., Zwiers, M.P., Schweren, L.S., van Hulzen, K.J., Medland, S.E., Shumskaya, E., Jahanshad, N., et al.: Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *The Lancet Psychiatry* **4**(4), 310–319 (2017)
15. Jomeiri, A., Habibizad Navin, A., Shamsi, M.: Regional attention-enhanced vision transformer for accurate alzheimer's disease classification using smri data. *Computers in Biology and Medicine* **197**, 111065 (2025)
16. Khvostikov, A., Aderghal, K., Benois-Pineau, J., Krylov, A., Catheline, G.: 3d cnn-based classification using smri and md-dti images for alzheimer disease studies. *arXiv preprint arXiv:1801.05968* (2018)
17. Kringelbach, M.L.: The human orbitofrontal cortex: linking reward to hedonic experience. *Nature Reviews Neuroscience* **6**(9), 691–702 (2005)
18. Kumar, V.S., Lanitha, B.: Ocnn classification model framework: A system for classifying alzheimers disease from smri images. In: 2025 8th International Conference on Computing Methodologies and Communication (ICCMC). pp. 620–627. IEEE (2025)
19. Kunanbayev, K., Shen, V., Kim, D.S.: Training vit with limited data for alzheimer's disease classification: an empirical study. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 334–343. Springer (2024)
20. Luo, X., Wu, J., Yang, J., Xue, S., Beheshti, A., Sheng, Q.Z., McAlpine, D., Sowman, P., Giral, A., Yu, P.S.: Graph neural networks for brain graph learning: A survey. *arXiv preprint arXiv:2406.02594* (2024)
21. Mao, W., Chen, Y., He, Z., Wang, Z., Xiao, Z., Sun, Y., He, L., Zhou, J., Guo, W., Ma, C., et al.: Brain structural connectivity guided vision transformers for identification of functional connectivity characteristics in preterm neonates. *IEEE Journal of Biomedical and Health Informatics* **28**(4), 2223–2234 (2024)
22. Milham, M.P., Fair, D., Mennes, M., Mostofsky, S.H.: The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in Systems Neuroscience* **6** (2012)
23. Nachev, P., Kennard, C., Husain, M.: Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience* **9**(11), 856–869 (2008)
24. Pini, L., Lista, S., Griffo, A., Allali, G., Imbimbo, B.P.: Can brain network connectivity facilitate the clinical development of disease-modifying anti-alzheimer drugs? *Brain Communications* **7**(1), fcae460 (2025)
25. Qiang, Y.R., Zhou, Q.Y., Li, J.N., Xie, M.Y., Cui, X., Zhang, S.W.: Classification of alzheimer's disease by jointing 3d depthwise separable convolutional neural network and transformer. *Expert Systems with Applications* **286**, 127720 (2025)

26. Said, A., Bayrak, R., Derr, T., Shabbir, M., Moyer, D., Chang, C., Koutsoukos, X.: Neurograph: Benchmarks for graph machine learning in brain connectomics. *Advances in Neural Information Processing Systems* **36**, 6509–6531 (2023)
27. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: 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)
28. Uchida, Y., Nishita, Y., Kato, T., Iwata, K., Sugiura, S., Suzuki, H., Sone, M., Tange, C., Otsuka, R., Ando, F., et al.: Smaller hippocampal volume and degraded peripheral hearing among japanese community dwellers. *Frontiers in Aging Neuroscience* **10**, 319 (2018)
29. Zhang, L., Wang, M., Liu, M., Zhang, D.: A survey on deep learning for neuroimaging-based brain disorder analysis. *Frontiers in Neuroscience* **14**, 779 (2020)
30. Zhou, H., He, L., Zhang, Y., Shen, L., Chen, B.: Interpretable graph convolutional network of multi-modality brain imaging for alzheimer’s disease diagnosis. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2022)