Multi-Spatial Granger Causality Features Fusion Network for Alzheimer's Disease Classification

Zhiwei Song[1], Jingming Li[1], Hu Yu[1], and Xiaojuan Guo[1,2] ✉

1 School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China
2 Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing 100875, China
gxj@bnu.edu.cn

Abstract. By leveraging complementary Euclidean and graph-based spatial information from structural Magnetic Resonance Imaging (sMRI), the effective fusion of multi-spatial brain features holds the potential to enhance the classification accuracy for Alzheimer's Disease (AD). Existing deep learning models often rely on simplistic methods such as concatenation, weighted summation, and self-attention to integrate Euclidean and graph spatial features. However, these models neglect the causal relationships between feature domains and labels, resulting in redundancies and limiting the classification accuracy. In this study, we propose a Multi-Spatial Granger Causality Features Fusion Network (MSGCFNet). Specifically, the MSGCFNet consists of a Multi-Spatial Features Encoder (MSFEN) module that extracts Euclidean and graph spatial features, a Multi-Spatial Granger Causality Features Disentanglement (MSGCFD) module that uses Granger causality-based learning to disentangle the causal dependencies within Euclidean and graph spatial features, and a Multi-Spatial Features Fusion Classification (MSFFC) module that employs a bidirectional cross-attention mechanism to robustly fuse the disentangled features from the two spatial features. Additionally, we design a multi-spatial Granger causal contrast disentanglement loss function that effectively minimizes the bias and redundancy of the disentangled features. Experimental results demonstrate that MSGCFNet achieves classification accuracies of 93.6% for Alzheimer's Disease (AD) vs. Normal Controls (NC) and 83.4% for Early Mild Cognitive Impairment (EMCI) vs. Late Mild Cognitive Impairment (LMCI) tasks, highlighting its superior classification performance. The code is available at https://github.com/FindBrain/MSGCFNet.

Keywords: Granger causality, multi-spatial features fusion, sMRI, Alzheimer's disease.

1 Introduction

Structural magnetic resonance imaging (sMRI) [1] is a cornerstone in Alzheimer's disease (AD) diagnosis due to its clinical accessibility and ability to detect brain atrophy [2]. In deep learning-based methods for AD diagnosis, Convolutional neural networks (CNNs) [3,4] typically characterize the gray matter volume (GMV) [5] of Euclidean spatial in sMRI, while graph neural networks (GNNs) [6,7] characterize morphological

brain networks [8,9] in graph space. However, these studies primarily focus on features within a single spatial domain, neglecting the potential benefits of integrating complementary information from both spatial representations, and preventing the full utilization of sMRI's diagnostic potential for AD.

Recent studies have increasingly explored the use of deep learning technology to leverage multi-spatial brain features from sMRI for enhancing Alzheimer's disease (AD) diagnosis [10-12]. For instance, in [10], the author proposed a two-channel European-graph spatial feature representation framework based on CNN-Transformer and GCN models for early AD diagnosis. In [11], the author developed a multi-spatial information fusion framework, incorporating key brain regions via multi-scale CNN and GCN models, to improve AD classification performance. However, existing fusion models often ignore the causal relationships between different spatial features and target labels, leading to redundancies and biases. Furthermore, these fusion approaches often rely on simplistic methods such as concatenation, addition or attention, which fail to capture the complex nonlinear interactions between features, limiting the models' ability to make full use of complementary information.

To address these challenges, we propose a Multi-Spatial Granger Causality Feature Fusion Network (MSGCFNet), which leverages Granger causality-based learning to enhance the integration of causal features from both Euclidean and graph spaces. Our approach effectively disentangles these features while preserving causal relationships, minimizing bias, and reducing redundancy. By focusing on causal dependencies, only the most relevant and non-redundant information is retained during the diagnostic process. Furthermore, a bidirectional cross-attention mechanism is employed to robustly fuse the decoupled features, capturing complex cross-domain interactions between Euclidean and graph spatial attributes in sMRI data.

In summary, the contributions of our work are as follows:

1) We propose MSGCFNet, a deep learning framework for multi-spatial brain feature integration that leverages Granger causality-based learning to preserve causal dependencies from multi-spatial features and a bidirectional cross-attention mechanism to capture complex cross-domain interactions, thereby enhancing AD diagnosis.
2) We design a Multi-spatial Granger causal contrast disentanglement loss function to ensure the preservation of relevant and non-redundant information from both Euclidean and graph spatial features.
3) MSGCFNet incorporates a bidirectional cross-attention mechanism to capture complex interactions between the decoupled causality features, enhancing the ability of multi-spatial feature fusion to improve AD classification accuracy.

2 Methodology

As illustrated in Fig.1, the proposed MSGCFNet consists of three parts: (1) Multi-Spatial Features Encode (MSFEN) module that extracts both Euclidean spatial features from gray matter volume (GMV) and the graph spatial features from the Regional Radiomics Similarity Network (R2SN) [8] (Fig. 1a); (2) Multi-Spatial Granger Causality

Features Disentanglement (MSGCFD) module that disentangles the causality features related to AD classification from GMV and R2SN (Fig. 1b); (3) Multi-Spatial Features Fusion Classification (MSFFC) module that fuses the features of the Euclidean and graph spaces (Fig. 1c).

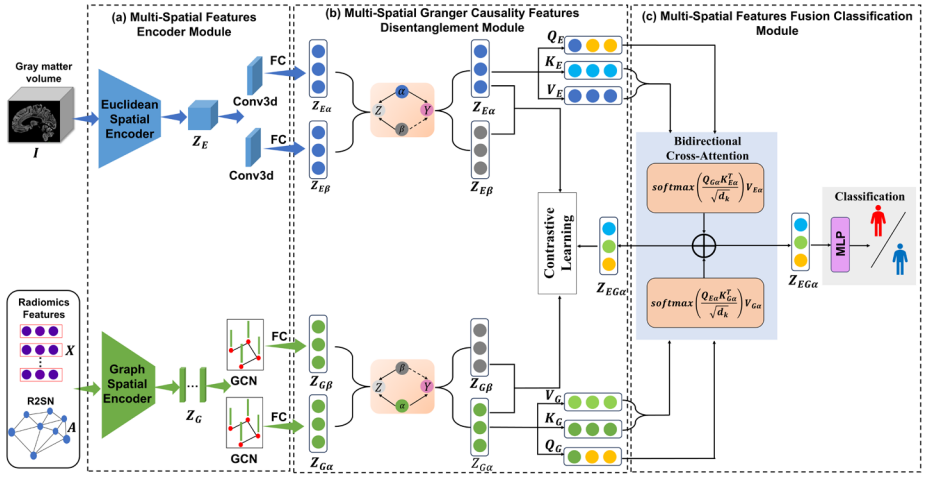

Fig. 1. Architecture of the proposed MSGCFNet model. (a) Multi-Spatial Features Encoder Module; (b) Multi-Spatial Granger Causality Features Disentanglement Module; (c) Multi-Spatial Features Fusion Classification Module.

2.1 Multi-Spatial Features Encoder Module

To represent the Euclidean spatial features, we adopt the ResNet model [13] as the backbone network to extract the initial latent features Z_E from the GMV. As shown in Fig.2 and Fig.1(a), the Euclidean spatial encoder includes two modules. Module 1, before the first ResBlock, consists of instance normalization, 3D convolution, batch normalization, ReLU activation, and average pooling. Module 2 includes eight ResBlocks, each with two 3×3×3 convolutions, one 1×1×1 convolution, three batch normalizations, two ReLU activations, and residual connections. Three-dimension GMV images is used as inputs, denoted as $I \in R^{C\times L\times W\times H}$ ($C \times L \times W \times H$ represents the channels, length, width, and height of the images, respectively.). Finally, we formulate the above feature learning process as shown in Equation (1)

$$Z_E = Resbolck\ Module\big(Conv\ Module(I)\big), \tag{1}$$

where $Conv\ Module(\cdot)$ represents the Module 1 part in Fig.2, and Resblock Module $(\cdot)$ represents the Module2 part in Fig.2. Finally, the Euclidean latent feature Z_E was obtained from Euclidean spatial encoder.

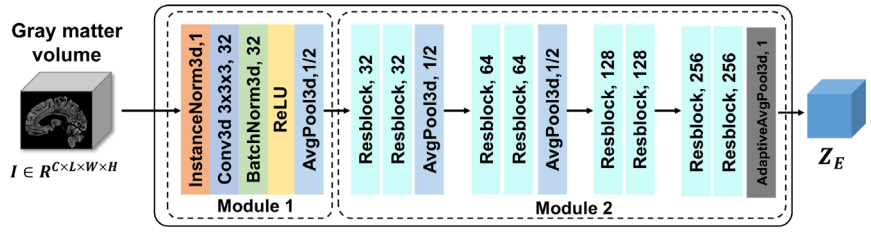

Fig. 2. Architecture of the Euclidean Spatial Encoder

Additionally, as shown in Fig.3 and Fig.1(a), a graph spatial encoder is used to learn the relationship between node features $X \in R^{Roi\times Rad}$ (where Roi represents the number of brain regions, and Rad represents the number of radiomics features per region) and edges of the R2SN [8] $A \in R^{Roi\times Roi}$ based on GCN [14]. Specifically, the graph spatial encoder consists of two GCN modules. Each GCN module includes a GCN layer, a layerNorm, and a LeakyReLU activation layer. The GCN operation generalizes the traditional convolution operator to graph data by defining a filter in the graph spectral domain, and the calculation process of the convolution operator encoding node features is shown in Equation (2):

$$Z_G = LayerNorm(\sigma\left(\hat{A}\sigma\left(LayerNorm\left(\hat{A}XW^{(0)}\right)\right)W^{(1)}\right)), \tag{2}$$

where $\hat{A}$ is the Laplacian-normalized adjacency matrix derived from A, $W^{(\cdot)}$ is a trainable matrix. $\sigma(\cdot)$ represents the LeakReLU activation function. Finally, we obtained the latent feature Z_G in the graph structural data.

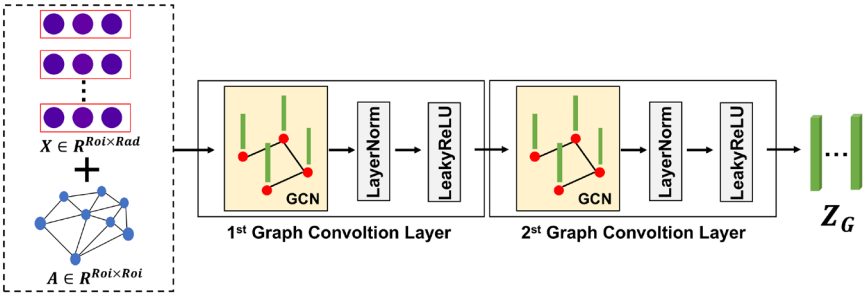

Fig. 3. Architecture of the Graph Spatial Feature Encoder

To resolve the size mismatch between $Z_E \in R^{N\times 256\times 1\times 1\times 1}$ and $Z_G \in R^{N\times 90\times 25}$, feature transformation is applied. Specifically, after Euclidean encoding, two independent 3D convolutions and feature flattening transform Z_E into $Z_{E\alpha} \in R^{N\times 256}$ and $Z_{E\beta} \in R^{N\times 256}$ (where N represent batch size). Similarly, after graph spatial encoding, two independent graph convolutions, feature flattening and fully connected (FC) layers are applied to transform Z_G into $Z_{G\alpha} \in R^{N\times 256}$ and $Z_{G\beta} \in R^{N\times 256}$.

2.2 Multi-Spatial Granger Causality Features Disentanglement Module

In the context of joint representation of both Euclidean and graph space, the model aims to disentangle potential features that are related and unrelated to label Y. Therefore, it is crucial to extract the underlying causality characteristics of both spatial types from the backbone network. Therefore, we design a Multi-Spatial Granger Causality Features Disentanglement (MSGCFD) module, which is conducted on a multi-spatial Granger causal contrast disentangling loss function (L_{Mgccl}), which includes multi-spatial Granger causal disentanglement loss function ($L_{Mgcausal}$), multi spatial contrastive loss function ($L_{Mcontrastive}$) and Binary CrossEntropy loss function [15] (L_{BCE}). The potential features related to Y in Euclidean and graph spaces are disentangled through Granger causal calculation and contrast learning, effectively reducing the influence of redundant information on the target task label Y. Additionally, the AD diagnostic task label is denoted as Y.

For the $L_{Mgcausal}$, the causal graph in Fig.1(b) illustrates the causal relationships. Specifically, we assume that the input Euclidean features Z_E and graph features Z_G are influenced by latent factors α (represented as $Z_{E\alpha}$ and $Z_{G\alpha}$) and β (represented as $Z_{E\beta}$ and $Z_{G\beta}$). Here α encapsulates the latent factors that have a direct causal relationship with the target label Y, while β represents latent factors that exhibit a spurious correlation with Y. Therefore, we assume that the latent factors α and β are disentangled or independent, such that α has a direct causal impact on label Y. Therefore, based on the proofs provided [16] and [17], we design the multi-spatial granger causal disentangle loss function (called $L_{Mgcausal}$) to calculate the potential features in the label-related Euclidean and graph spaces. $L_{Mgcausal}$ can be defined as:

$$L_{Mgcausal} = I\left(Z_{E\alpha}; Z_{E\beta}\right) + I\left(Z_{G\alpha}; Z_{G\beta}\right) - I\left(Z_{E\alpha}; Y|Z_{E\beta}\right) - I\left(Z_{G\alpha}; Y|Z_{G\beta}\right), \quad (3)$$

$$I(\alpha;\beta) = H(\alpha) + H(\beta) - H(\alpha,\beta), \quad (4)$$

$$I(\alpha;Y|\beta) = H(\alpha|\beta) - H(\alpha|Y,\beta) = H(\alpha,\beta) + H(Y,\beta) - H(\beta) - H(\alpha,Y,\beta), \quad (5)$$

where $I(\cdot;\cdot)$ denotes mutual information and $H(\cdot)$ denotes entropy or joint entropy.

2.3 Multi-Spatial Features Fusion Classification Module

MSGCFD disentangled the Euclidean spatial features $Z_{E\alpha}$ and graph spatial features $Z_{G\alpha}$ related to the target task through $L_{Mgcausal}$. To enhance the interaction between features, we propose a Multi-Spatial Features Fusion Classification (MSFFC) module based on a multi-spatial bidirectional cross-attention module (Fig.1(c)) for effectively fusing Euclidean and graph spatial features, resulting in the fused features $Z_{EG\alpha}$.
Furthermore, to reduce the impact of irrelevant features $Z_{E\beta}$ and $Z_{G\beta}$ on the fused features $Z_{EG\alpha}$, we introduce a multi spatial contrastive loss function (called $L_{Mcontrastive}$) to diminish the relationship between $Z_{EG\alpha}$, $Z_{E\beta}$, and $Z_{G\beta}$. Finally, the fused features $Z_{EG\alpha}$ are passed through a multi-layer perceptron (MLP) [18] for the subsequent classification task, enabling more precise task-driven feature representations. Bidirectional cross-attention and $L_{Mcontrastive}$ formulas are defined as:

$$Z_{EG\alpha} = softmax\left(\frac{Q_{E\alpha}K_{G\alpha}^{T}}{\sqrt{d_k}}\right)V_{G\alpha} + softmax\left(\frac{Q_{G\alpha}K_{E\alpha}^{T}}{\sqrt{d_k}}\right)V_{E\alpha}, \tag{6}$$

$$L_{Mcontrastive} = -\frac{1}{N}\sum_{i=1}^{N}\left[log\sigma(S^{(i)}(Z_{E\alpha}, Z_{EG\alpha})) + log\sigma(S^{(i)}(Z_{G\alpha}, Z_{EG\alpha})) - log\sigma(S^{(i)}(Z_{E\alpha}, Z_{E\beta})) + log\sigma(S^{(i)}(Z_{G\alpha}, Z_{G\beta}))\right], \tag{7}$$

where $S^{(i)}$ represents the cosine similarity calculation of sample i, $\sigma(\cdot)$ denotes the Sigmoid function.

The Binary CrossEntropy (BCE) loss function [15] is used to optimize the classification task. To sum up, the proposed multi-spatial granger causal contrastive loss ($Loss_{Mccl}$) is shown as follows:

$$L_{Mgccl} = L_{BCE} + \lambda(L_{Mcausal} + L_{Mcontrastive}), \tag{8}$$

where λ is the hyper-parameter.

3 Experiments

3.1 Dataset and Pre-processing

We selected 1,481 subjects from the Alzheimer's Disease Neuroimaging Initiative database (ADNI) database (http://adni.loni.usc.edu) (Table 1). Gray matter volume (GMV) maps were segmented from 1.5T T1-weighted scans using CAT12, bias-corrected, and registered to MNI space, resliced to 2 mm³ isotropic voxels. Following [8], twenty-five Radiomics features were calculated for 90 AAL-derived regions, normalized via min–max scaling, and the R2SN [8] (90 × 90) was generated by Pearson correlation based on Radiomics features.

Table 1. Demographic information about the database of ADNI.

Group	Participants	Sex (M/F)	Age (years)
NC	603	277/326	73.46±6.16
AD	282	151/131	74.91±7.69
EMCI	240	129/111	70.89±7.68
LMCI	356	225/131	73.77±7.87

3.2 Implementation Details

The experiments were implemented on Pytorch 1.9.1. The hyperparameters were optimized with Adam optimizer with a learning rate equal to 3×10^{-4} and a weight decay rate equal to 1×10^{-4}. The batch size was set to 8 to fit the GPU memory. The training epochs was set to 100. λ in Eq.8 were experimentally set to 0.05, respectively.

3.3 Comparative Analysis of Classification Performance

To compare classification performance, we conducted experiments using Euclidean spatial features of GMV feature maps, the graph spatial features of R2SNs, and their combination. For GMV maps, we compared four deep learning models: 3D VGG16 [19], 3D ResNet [13], 3D Attention Network (3DAN) [20], and Resnet-Transformer (ResTransformer) [10]. For the R2SN network, we compared five GNN models: Graph Convolutional Network (GCN) [14], Graph Attention Network (GAT) [21], GraphSage [22], GraphTransformer [23], and BC-GCN [24]. For the combined multi-spatial features, we compared two recently models: s^{2}MRI-ADNet [10] and MSRNet [11]. Table 2 summarizes the experimental results for the AD vs NC and LMCI vs EMCI tasks within the ADNI database under the 10-fold cross-validation. The proposed model achieved accuracy (ACC) of 93.6%, F1 score of 90.0%, and an area under the receiver operating characteristic curve (AUC) of 96.4% for the AD vs NC task. For the LMCI vs EMCI task, the model reached an ACC of 83.4%, F1 score of 85.9%, and AUC of 90.4%. The diagnostic classification of the proposed model outperforms all comparison models.

Table 2. Comparison of the proposed model with other deep learning-based methods in classifying AD vs NC and LMCI vs EMCI. The best result is in bold.

Feature	Model	AD vs NC			LMCI vs EMCI		
		ACC (%)	F1 (%)	AUC (%)	ACC (%)	F1 (%)	AUC (%)
R2SN	GCN	81.0±3.7	66.6±6.9	86.3±3.5	76.8±4.5	77.9±3.9	82.0±3.9
	Graphsage	81.5±3.2	62.6±5.0	88.4±3.3	74.5±4.1	78.2±3.5	82.1±2.6
	GAT	82.3±3.3	68.5±6.5	88.8±4.0	73.7±4.0	79.1±2.0	79.8±4.1
	GraphTransformer	81.4±5.0	62.9±12.0	87.4±4.0	66.3±4.2	74.4±3.5	77.3±4.6
	BC-GCN	87.2±3.0	78.2±4.9	92.3±3.4	73.8±5.0	80.1±2.5	77.9±11.5
GMV	3D VGG16	89.1±3.8	82.1±7.1	94.7±2.6	79.2±5.3	82.4±5.2	86.5±3.6
	ResNet	88.5±4.2	79.9±8.6	94.8±2.4	77.9±4.6	81.0±4.5	86.6±3.9
	3DAN	88.1±3.9	80.2±8.4	93.5±3.2	78.7±5.1	81.9±5.1	87.4±4.1
	ResVIT	89.0±2.6	81.8±5.1	94.8±2.6	75.5±6.1	79.2±5.1	81.0±6.0
Combination	s^{2}MRI-ADNet	92.1±2.3	87.7±3.5	96.2±1.5	79.1±4.9	81.7±4.8	86.6±3.8
	MSRNet	92.8±1.8	88.5±2.9	95.6±1.1	79.8±3.3	82.5±2.5	87.1±3.1
	MSGCFNet	**93.6±2.7**	**90.0±3.7**	**96.4±2.5**	**83.4±2.1**	**85.9±1.8**	**90.4±4.3**

3.4 Ablation Analysis

To evaluate the classification performance of the different modules in the proposed MSGCFNet, we consider several configurations where the MSFEN module, the MSGCFD module, and the MSFFC module are either included or excluded (Table 4). The results show that the best performance is achieved when the three modules are fully integrated.

To further validate our proposed fusion approach with causality considerations, Fig.4 presents a t-test violin plot illustrating the differences (including the absolute t-values and p-values) between causal and non-causal features in the AD vs. NC groups and the

LMCI vs. EMCI groups. As shown in Fig.4, compared with the $L_{Mcausal}$ only and the L_{BCE} only, the $Loss_{Mgccl}$ can obtain more significant differences (p<0.0001). In summary, the experimental results in Fig.4 demonstrated the superiority of the proposed loss function $Loss_{Mgccl}$ in improving multi-spatial features fusion and for diagnostic tasks.

Table 3. Ablation results about different modules of the proposed MSGCFNet.

Model	AD vs NC			LMCI vs EMCI		
	ACC (%)	F1 (%)	AUC (%)	ACC (%)	F1 (%)	AUC (%)
MSFEN	90.1±3.5	84.2±5.6	94.8±2.4	77.1±2.6	79.5±2.7	87.2±2.3
MSFEN+·MSGCFD	90.2±3.2	84.5±5.0	95.5±3.0	79.8±4.5	82.6±3.9	86.0±4.1
MSFEN+MSFFC	91.3±3.2	85.9±5.6	95.9±2.9	81.6±4.4	84.4±4.0	88.2±4.2
MSGCFNet	**93.6±2.7**	**90.0±3.7**	**96.4±2.5**	**83.4±2.1**	**85.9±1.8**	**90.4±4.3**

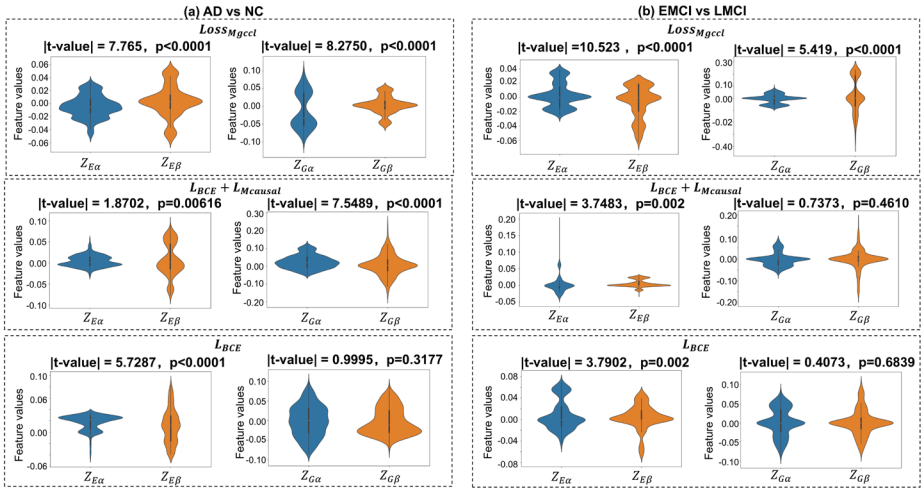

Fig. 4. Violin Plot of Feature Disentanglement Analysis with Different Loss Functions.

4 Conclusion

In summary, the proposed Multi-Spatial Granger Causality Feature Fusion Network (MSGCFNet) effectively integrates multi-spatial brain features between Euclidean and graph spatial domains from sMRI. Furthermore, the MSGCFNet disentangle causal dependencies between different feature domains and robustly fuse them using a bidirectional cross-attention mechanism. The experimental results show that the MSGCFNet exhibits a superior performance to other models in AD classification.

Acknowledgments. This study was supported by the National Natural Science Foundation of China under Grant 62071051, and Interdisciplinary Research Foundation for Doctoral Candidates of Beijing Normal University under Grant BNUXKJC2411.

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

References

1. Yin, T., Cao, M., Yu, J., Shi, T., Mao, X., Wei, X., Jia, Z.: T1-weighted imaging-based hippocampal radiomics in the diagnosis of Alzheimer's disease. Academic Radiology **35**(12), 5183-5192 (2024)
2. Zhou, Y., Li, Y., Zhou, F., Liu, Y., Tu, L.: Learning with domain-knowledge for generalizable prediction of Alzheimer's disease from multi-site structural MRI. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 452-461. Springer (2023)
3. Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O.: Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. Medical Image Analysis **63**, 101694 (2020)
4. Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., Yang, Y., Guo, G., Xiao, M., Du, M., Qu, X.: Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. Frontiers in Neuroscience **12**, 777 (2018)
5. Zhao, K., Ding, Y., Han, Y., Fan, Y., Alexander-Bloch, AF., Han, T., Jin, D., Liu, B., Lu, J.: Independent and reproducible hippocampal radiomic biomarkers for multisite Alzheimer's disease: diagnosis, longitudinal progress and biological basis. Science Bulletin **65**,1103-1113 (2020)
6. Zhou, Z., Wang, Q., An, X., Chen, S., Sun, Y., Wang, G., Yan, G.: A novel graph neural network method for Alzheimer's disease classification. Computers in Biology and Medicine **180**,108869 (2024)
7. Zhang, J., He, X., Qing, L., Chen, X., Liu, Y., Chen, H.: Multi-relation graph convolutional network for Alzheimer's disease diagnosis using structural MRI. Knowledge-Based Systems **270**,110546 (2023)
8. Zhao, K., Zheng, Q., Che, T., Dyrba, M., Li, Q., Ding, Y., Zheng, Y., Liu, Y., Li, S.: Regional radiomics similarity networks (R2SNs) in the human brain: reproducibility, small-world properties and a biological basis. Network Neuroscience **5**(3), 783-797 (2021)
9. Cai, M., Ma, J., Wang, Z., Zhao, Y., Zhang, Y., Wang, H., Xue, H., Chen, Y., Zhang, Y., Wang, C., Zhao, Q, He, K., Liu, F.: Individual‐level brain morphological similarity networks: Current methodologies and applications. CNS Neuroscience & Therapeutics **29**(12), 3713-3724 (2023)
10. Song, Z., Li, H., Zhang, Y., Zhu, C., Jiang, M., Song, L., Wang, Y., Ouyang, M., Hu, F., Zheng, Q.: s2MRI-ADNet: an interpretable deep learning framework integrating Euclidean-graph representations of Alzheimer's disease solely from structural MRI. Magnetic Resonance Materials in Physics, Biology and Medicine **37**, 845-857 (2024)
11. Nan, P., Li, L., Song, Z., Wang, Y., Zhu, C., Hu, F., Zheng, Q.: A multispatial information representation model emphasizing key brain regions for Alzheimer's disease diagnosis with structural magnetic resonance imaging. Quantitative Imaging in Medicine and Surgery **14**(12), 8568-8585 (2024)
12. Fan, C., Yang, H., Peng, L., Zhou, X., Ni, Z., Zhou, Y., Chen, S., Hou, Z.: BGL-Net: A brain-inspired global-local information fusion network for Alzheimer's disease based on sMRI. IEEE Transactions on Cognitive and Developmental Systems **15**,1161-1169 (2022)

13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 770-778 (2016)
14. Kipf, T., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907, (2016)
15. Ruby, U., Yendapalli, V.: Binary cross entropy with deep learning technique for image classification. International Journal of Advanced Trends in Computer Science and Engineering **9**, (2020)
16. O'Shaughnessy, M., Canal, G., Connor, M., Rozell, C., Davenport, M.: Generative causal explanations of black-box classifiers. Advances in Neural Information Processing Systems **33**,5453-5467 (2020)
17. Zheng, K., Yu, S., Chen, B.: Ci-GNN: A granger causality-inspired graph neural network for interpretable brain network-based psychiatric diagnosis. Neural Networks **172**,106147 (2024)
18. Riedmiller, M., Lernen, A.: Multi layer perceptron. University of Freiburg **24**, (2014)
19. Luong, H., Vo, P., Phan, H., Tran, N., Le, H., Nguyen, H.: Fine-tuning VGG16 for Alzheimer's disease diagnosis. In: Conference on Complex, Intelligent, and Software Intensive Systems, Springer, pp. 68-79 (2023)
20. Jin, D., Zhou, B., Han, Y., Ren, J., Han, T., Liu, B., Lu, J., Song, C., Wang, P., Wang D.: Generalizable, reproducible, and neuroscientifically interpretable imaging biomarkers for Alzheimer's disease. Advanced Science **7**,2000675 (2020)
21. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
22. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. Advances in Neural Information Processing Systems **30**,1024-1034, (2017)
23. Zhang, Y., Cao, G., Xu, X., Kang, G.: Graph-based Alzheimer's disease diagnosis with contrastive learning and graph transformer. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, pp. 2368-2373 (2023)
24. Li, Y., Zhang, X., Nie, J., Zhang, G., Fang, R., Xu, X., Wu, Z., Hu, D., Wang, L., Zhang, H., Lin, W., Li, G.: Brain connectivity based graph convolutional networks and its application to infant age prediction. IEEE Transactions on Medical Imaging **41**,2764-2776 (2022)