Alzheimer's Disease Recognition Based on Adaptive Graph Normalization Flow for Incomplete Multimodal Data Fusion

Yaqin Li, Yihong Dong(✉), Yanan Wu, Haihao Yan, and Linlin Gao

Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China
dongyihong@nbu.edu.cn

Abstract. Multimodal data holds significant value in the diagnosis of Alzheimer's disease (AD). However, in real-world applications, factors such as privacy protection, acquisition costs, and sensor failures often lead to data missingness, posing challenges for incomplete multimodal learning. Currently the artificial intelligence-based diagnostic methods for AD on incomplete multimodal data have gained increasing attention. However, existing approaches typically overlook modality distribution discrepancies and suffer from severe performance degradation under recovery paradigms lacking reconstruction experience. To address this challenge, we propose an Adaptive Graph Distribution Consistency Modal Recovery Network Based on Normalizing Flows (AGDiC) to tackle incomplete multimodal learning in neuroimaging. We develop a novel framework integrating adaptive graph learning with normalizing flows and a modality regularization strategy. This framework focuses adaptive graph attention features on modality distributions while ensuring distribution consistency of recovered data, and employs masked cross-attention to facilitate multimodal fusion. Unlike conventional methods, our model can handle arbitrary modality missingness during both training and inference phases without relying on reconstruction experience. Extensive experiments are conducted using three neuroimaging modalities from the ADNI dataset: sMRI, fMRI and PET. Results demonstrate that our method achieves state-of-the-art performance and exhibits remarkable stability across various random missing rates.

Keywords: Incomplete multimodal learning · Alzheimer's disease · Normalizing Flows · Graph Neural Network · Multimodal fusion.

1 Introduction

Alzheimer's Disease (AD) is an irreversible neurodegenerative disorder, and early diagnosis is critical for improving prognosis and delaying disease progression [4][18]. Currently, structural MRI (sMRI), functional MRI (fMRI), and Positron Emission Tomography (PET) data can each reflect the brain's biomarker features from different perspectives. However, compared to sMRI, the acquisition of fMRI and PET is more challenging. In particular, PET faces challenges in real-world applications due to the high radiation risk and cost of imaging, which often results in incomplete multimodal learning [3][23].

For incomplete multimodal learning, existing approaches can be divided into non-recovery methods and recovery methods. Non-recovery methods can be broadly classified into grouping strategy-based, correlation maximization-based, knowledge distillation-based, and latent space-based [25][12][8][24]. For example, Zhou et al. [29] and Chen et al. [5] address the issue of missing modalities by establishing modality relationships through deep representations, projecting this latent space into the label space for AD diagnosis. However, non-recovery methods still struggle with handling complementary information among modalities and lack flexibility. Recovery methods estimate and reconstruct missing modalities, which can generally be categorized into zero-based recovery, average-based recovery, and deep learning-based recovery [15][26][16][11]. Deep learning-based methods tend to offer more advantages. For instance, Pan et al. [14] proposed a feature-consistency generative adversarial network to estimate missing PET images in MRI for AD diagnosis, whereas Wang et al. [20] synthesized more discriminative PET images via joint learning for multimodal fusion. However, they overlooked the distribution discrepancies among modalities. Subsequently, Wang Y et al. [21] introduced a flow-based distribution consistency recovery paradigm for modality reconstruction. Nevertheless, current recovery paradigms assume that modality data in the training stage are complete or available. Specifically, they often rely on complete real data during training to empirically correct the generated modalities and only support missing modality inputs at the testing stage. They fail to address the more realistic recovery paradigm in which there is a lack of "reconstruction experience." Under such a paradigm, the performance of existing models degrades significantly, which is the core issue we aim to tackle.

To overcome the challenges mentioned above, we propose a novel Adaptive Graph Distribution Consistency Modal Recovery Network Based on Normalizing Flows (AGDiC), which can adapt to the missing modalities during both the training and inference processes of multimodal tasks while ensuring accurate AD classification. The main contributions of our work are as follows: 1) Unlike existing methods, our AGDiC model can address the recovery paradigm that lacks reconstruction experience, supporting the missing of any modality during both the training and inference stages. To the best of our knowledge, this is the first work on incomplete multimodal learning in neuroimaging under this paradigm. 2) We develop a novel modality regularizer based on normalizing flows, aimed at capturing modality distribution differences and ensuring the stability of the recovered modalities. At the same time, we integrate adaptive layer memory attention graph learning with flow-based generation strategies to precisely capture the key structural features of each modality. 3) Through multimodal masked cross-attention, we facilitate effective modality fusion. Experimental results on three modalities (sMRI, fMRI, and PET) from the ADNI dataset demonstrate that our method achieves significant performance improvements compared to existing techniques in multiple random missing scenarios.

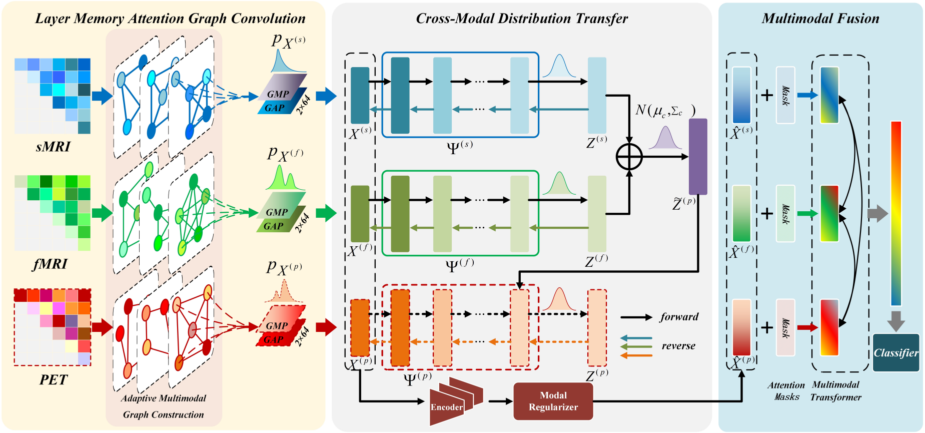

Fig. 1. The AGDiC model consists of three modules: layer memory attention graph convolution, cross-modal distribution transfer, and multimodal fusion. For simplicity, it is assumed that PET modality is missing in this figure, but in practice, any modality may be missing, as long as at least one modality is available for each subject.

2 Method

Our AGDiC framework is shown in Fig. 1. For clarity and without loss of generality, it is assumed that the PET modality is missing. The model consists of three main components: layer memory attention graph convolution, cross-modal distribution transfer, and multimodal fusion. First, specific brain network matrices are constructed for sMRI, fMRI, and PET as the initial features. Next, a set of graphs is adaptively learned for each modality to accommodate layer memory attention at different stages. Then, cross-modal distribution transfer is used to estimate the distribution of the missing modality, and empirical correction is performed through a modality regularizer after the three encoders. Finally, multimodal fusion is carried out through masked cross-attention for classification.

2.1 Adaptive Multimodal Graph Learning

Let the tuple $\left(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \ldots, \boldsymbol{X}^{(M)}\right)$ denote the M different modalities of a sample. Define an indicator $\lambda \in \{0, 1\}$, where $\lambda_m = 0$ indicates that the m-th modality is missing, and otherwise, $\lambda_m = 1$. Thus, the goal is to recover the missing modalities $M_{\text{miss}} = \{\, m \mid \lambda_m = 0\}$ based on the observable modalities $M_{\text{obs}} = \{\, m \mid \lambda_m = 1\}$. Considering the differences among fMRI, sMRI, and PET data, we construct a specific brain network matrix for each as the initial feature. First, all modality data are parcellated into ROIs using the AAL brain atlas template [19]. For fMRI, the ROI average time series is obtained, and then the Pearson correlation coefficient (PCC) is computed between ROI time series: $\boldsymbol{X}_{i,j}^{\{f\}} = \text{PCC}\big(s_i, s_j\big) \in \mathbb{R}^{N \times N}$, which serves as the initial fMRI feature. For PET, a Fisher transform-based approach is employed to compute the correlation coefficient between the i-th and j-th ROIs, yielding the individual feature matrix for PET [2]: $\boldsymbol{X}_{i,j}^{\{p\}} = \left[\exp\left(2 \times E(i,j)\right) - 1\right] / \left[\exp\left(2 \times E(i,j)\right) + 1\right]$, where $E(i,j)$ denotes the connectivity matrix of the PET SUVR feature between the i-th and j-th ROIs, i.e., a weight matrix calculated based on the difference in the effect size of average uptake between individual subjects and the average NC subjects [27]. For sMRI, gray matter features are first extracted, and the same computation method as for the PET individual features is applied to obtain the individual feature $\boldsymbol{X}_{i,j}^{\{s\}}$. At this point, the initial individual feature matrices for fMRI, PET, and sMRI, $\boldsymbol{X}^{\{f,p,s\}} \in \mathbb{R}^{116\times116}$, are used as the input to our model.

Treating ROIs as nodes and constructing edges based on the relationships between ROIs is a common approach, which facilitates subsequent feature extraction using Graph Neural Networks (GNNs). We adopt an adaptive graph construction method for modeling each modality to flexibly extract multimodal structural features [6]. Specifically, an Multi-Layer Perceptron (MLP) is used to adaptively learn the adjacency matrix $\boldsymbol{A}_{ij}$ for each modality for feature updating to prevent over-smoothing in the GNN. This is achieved by constraining the most relevant nodes through a sparse graph neighbor attention mechanism:

$$\boldsymbol{A}^{(k)} = \left\| \mathrm{MLP}^{(k)}\left(U(\boldsymbol{x}_i^{(k)}, \boldsymbol{x}_j^{(k)})\right) \right\|, \quad \text{s.t.} \quad \boldsymbol{A}_i^{(k)} \leq \theta V, \quad \theta \in (0,1] \tag{1}$$

where k denotes the number of graph convolution layers, θ represents the edge construction ratio, and $\boldsymbol{A}_i^{(k)} \in \mathbb{R}^{1\times V}$. $U(\cdot,\cdot)$ calculates the cosine similarity between the features of the i-th and j-th nodes, which is then passed through an MLP to learn a node attention matrix. Finally, the sparse attention adjacency matrix for each layer of graph convolution is given by $\hat{\boldsymbol{A}}^{(k),m} = \boldsymbol{A}^{(k),m} + \boldsymbol{I}, \quad m \in M_{\text{obs}}$, where $\boldsymbol{I}$ is the identity matrix. Following the representation in [10], the output of the layer memory attention graph convolution at the $(k+1)$-th layer can be expressed as: $F_k(\boldsymbol{X}^{(k)}, \boldsymbol{W}^{(k)}) = \rho(\hat{\boldsymbol{A}}^{(k)} \boldsymbol{X}^{(k)} \boldsymbol{W}^{(k)}) \in \mathbb{R}^{V\times h}$. To avoid over-smoothing and overfitting of neighboring node information, we choose to use the "early memory" of intermediate layers for later data transmission. The output of the $(k+1)$-th layer GNN can be expressed as:

$$\boldsymbol{X}^{(k+1)} = \left[\boldsymbol{X}^{(k)}, F_k\left(\boldsymbol{X}^{(k)}, \boldsymbol{W}^{(k)}\right)\right] \in \mathbb{R}^{V\times(d+kh)} \tag{2}$$

where $W^{(k)} \in \mathbb{R}^{d\times h}$. After passing through n layers of GNN, the output $X^{(k+n)}$ is subjected to global max pooling (GMP) and global average pooling (GAP), yielding the deep features for each modality, $X^{(m)}, m \in M_{\text{obs}}$.

2.2 Flow-based Distribution Transfer and Modality Regularizer

Normalizing flows, proposed by Dinh et al. [7], are a classical generative probabilistic model, also known as flow-based generative models. As suggested in [21], let $\Psi^{(m)}$ represent the normalizing flow model for modality m, and its inverse transformation is denoted as $(\Psi^{(m)})^{-1}$. For example, when fMRI and sMRI modalities are available and the PET modality is missing, $\boldsymbol{X}^{(f)}$ and $\boldsymbol{X}^{(s)}$ can be input into $\Psi^{(f)}$ and $\Psi^{(s)}$ to obtain the multimodal latent states under Gaussian distribution, $\boldsymbol{Z}^{(f)} \sim \mathcal{N}(\mu_c, \Sigma_c)$ and $\boldsymbol{Z}^{(s)} \sim \mathcal{N}(\mu_c, \Sigma_c)$. We perform averaging to sample the latent representation of the missing PET modality as $\tilde{\boldsymbol{Z}}^{(p)} \leftarrow (\boldsymbol{Z}^{(f)}, \boldsymbol{Z}^{(s)})/2 \sim \mathcal{N}(\mu_c, \Sigma_c)$. Then, we inject this into $(\Psi^{(p)})^{-1}$ to generate a sample $\tilde{X}^{(p)}$ with the PET modality distribution. The process is as follows:

$$\tilde{\boldsymbol{X}}^{(p)} = (\Psi^{(p)})^{-1} \left([\Psi^{(f)}(\boldsymbol{X}^{(f)}) + \Psi^{(s)}(\boldsymbol{X}^{(s)})]/2 \right) \tag{3}$$

To ensure that all latent space representations are discriminative under a multivariate Gaussian distribution, we introduce labels to adaptively learn the Gaussian distribution for specific classes. The loss function for cross-modal distribution transfer, L_{MDT}, can be defined as:

$$\mathcal{L}_{\mathrm{MDT}} = -\sum_{m \in M_{\mathrm{obs}}} \left[\log p_{\boldsymbol{Z}^{(m)}} \left(\boldsymbol{Z}^{(m)} \mid y = c \right) + \log \left| \det \left(\partial \boldsymbol{Z}^{(m)} / \partial \boldsymbol{X}^{(m)} \right) \right| \right] \tag{4}$$

where the first term represents the log-density of $\boldsymbol{Z}^{(m)}$ under its own class condition, and the second term represents the log-determinant of the normalizing flow model for modality m.

Previous incomplete multimodal learning models tend to rely on real experience to correct the recovered modalities. However, when the model lacks reconstruction experience during training, such models may suffer severe degradation. Therefore, even without such external assistance, it is necessary to ensure the stability of the flow-based recovery strategy and to encourage the modality encoders to perform self-correction. To this end, the output $\hat{\boldsymbol{X}}^{(m)}$ from modality-specific encoders is used both for cross-modal fusion and as input to the modality regularizer to assist in decision-making. The modality regularization loss is:

$$\mathcal{L}_{\mathrm{AR}} = -\sum_{m}^{M} \sum_{i}^{N} \sum_{c}^{C} y_{ic}^{m} \log(p_{ic}^{m}) \tag{5}$$

where M is the number of modalities, N is the number of samples, and C is the number of classes. y_{ic} is an indicator function, which is 1 if sample i belongs to class c, and otherwise is 0. p_{ic} is the probability predicted by the model that sample i belongs to class c.

2.3 Masked Cross-Attention Modal Fusion

Modal interactions provide new information that may negatively affect predictions [17] [28]. Inspired by the Transformer [22], we use self-attention to capture these interactions, introducing a masking mechanism to limit unnecessary ones. The masked self-attention is: $\mathrm{Attn}(Q, K, V) = \mathrm{softmax}\left((QK^T)/\sqrt{d_z} + \Pi\right) V$, where Q, K, and V are the Queries, Keys, and Values obtained from the Tokens, and d_z is the projection dimension. Π is the mask matrix, where each element of $\mathit{\Pi}$ is 0 for interactions between observable modalities, and otherwise is $-\infty$, to reduce the noise generated by unnecessary interactions. Finally, the fusion features are passed to the classifier for classification. The total loss is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \alpha\mathcal{L}_{\text{MDT}} + \beta\mathcal{L}_{\text{AR}} \tag{6}$$

where $\mathcal{L}_{\text{CE}}$ is the cross-entropy loss, and α and β are adjustable hyperparameters.

3 Experiments

3.1 Materials and Experimental Setup

We use data from three modalities from ADNI2 [1], as shown in Table 1, which include resting-state fMRI, T1-weighted MPRAGE sMRI, and FDG-PET. The preprocessing workflow for all raw imaging data follows the methods of Zhang et al. [27] and Gu et al. [9]. Not every subject has complete modality data. In the classification tasks for Healthy Controls (HC), Early Mild Cognitive Impairment (EMCI), and Late Mild Cognitive Impairment (LMCI), there is inherent missingness in the multimodal data. For convenience, the modality missing rate is defined as $\eta = 1 - (\sum_{i=1}^{N}\sum_{m=1}^{M} J_{im})/(N \times M)$, where N is the number of subjects, M is the number of modalities, and J_{im} is an indicator function, which is 1 if modality is present for a subject, and 0 otherwise. According to this definition, the default missing rate for the three-class task is approximately 0.328.

Table 1. The data information used in the experiment.

Category	Number of samples			Default
	sMRI	fMRI	PET	Miss Rate
HC/EMCI/LMCI	211/185/170	162/150/131	52/38/42	0.328

All experiments used 5-fold cross-validation with 100 epochs. The Adam optimizer, with a learning rate of 0.0001 and hyperparameters $\alpha = 0.1$, $\beta = 0.5$, and edge ratio is 0.3. The model ran on a GeForce RTX 4080 GPU with PyTorch. The modality missing rate was consistent across training and testing, ensuring at least one modality per subject. Additionally, we conducted experiments with random modality deletion and increased missing rates of 0.4 and 0.5. Evaluation metrics included accuracy (Acc), recall (Rec), and F1 score (F1).

3.2 Comparison with Other Methods

The comparison methods can be divided into two categories: 1) Non-recovery methods, including the latent space-based OLFG [5] and G-SLC [13]; 2) Recovery-based methods, including the GAN-based DSDL [14], flow-based DiCMoR [21], and joint learning-based JLCM [20]. All methods were evaluated under the same

Table 2. The evaluation results of each model under different modal missing rates in different tasks are represented as the average of 5-fold cross-validation.

Task	Method	Modal Miss Rate η								
		Default			0.4			0.5		
		Acc	Rec	F1	Acc	Rec	F1	Acc	Rec	F1
HC vs EMCI vs LMCI	OLFG (2023)	53.54	53.41	52.45	51.76	51.11	50.12	49.65	49.61	48.26
	G-SLC (2024)	61.67	61.11	61.1	60.61	60.17	60.08	57.07	56.25	56.09
	DSDL (2021)	59.00	58.57	57.51	54.60	54.78	53.27	48.23	47.31	44.40
	DiCMoR (2023)	65.37	65.01	65.13	63.26	63.23	62.98	57.24	57.06	56.90
	JLCM (2024)	62.36	62.29	61.39	53.89	53.47	52.89	49.64	48.37	44.76
	Ours	**75.96**	**75.76**	**75.70**	**70.67**	**70.55**	**70.45**	**62.54**	**62.36**	**62.17**
EMCI vs LMCI	OLFG (2023)	60.28	59.27	56.19	58.31	59.02	55.15	57.46	57.09	52.69
	G-SLC (2024)	68.45	68.23	68.20	67.89	67.81	67.62	65.07	64.79	64.68
	DSDL (2021)	73.80	73.72	73.41	69.01	68.51	67.77	60.28	58.91	51.65
	DiCMoR (2023)	71.27	71.12	71.09	70.42	70.31	70.28	66.48	66.41	66.25
	JLCM (2024)	73.80	73.60	73.32	70.14	70.21	69.19	65.92	66.03	64.66
	Ours	**80.56**	**80.73**	**80.53**	**74.37**	**74.62**	**74.27**	**69.30**	**69.32**	**69.11**

experimental settings and appropriately adapted. Comparative experimental results are shown in Table 2, which includes two tasks: the three-class classification (HC vs EMCI vs LMCI) and the binary classification (EMCI vs LMCI). As shown in the table, our method achieves the best performance metrics across all tasks. First, recovery-based methods generally outperform non-recovery methods, and they exhibit higher stability across different missing rates. Second, the compared recovery-based methods, which rely on reconstruction experience during training, perform worse when effective reconstruction is lacking. In the three-class task, our method improves accuracy by 10.59%, 7.41%, and 5.3% compared to the second-best method in different scenarios. Similarly, in the binary classification task, our method also shows significant performance improvements.

3.3 Ablation Study and Parameter Sensitivity Analysis

We conducted an ablation study to validate the effectiveness of key modules in our model. The implementations are as follows: 1) **w/o ReFlow:** Remove the flow-based generation module with regularization; instead, a zero-filling method is used to recover missing modalities. 2) **w/o MAT:** Remove the masked cross-attention in the multimodal fusion module; instead, feature concatenation is used for modality fusion. 3) **w/o AR:** Remove the auxiliary regularization module, so that the encoder output is used solely for multimodal fusion.

In the HC vs EMCI vs LMCI task, the experimental results are shown in Table 3. First, ReFlow has the greatest impact on performance: when ReFlow is removed, the full model's performance drops by 9.0% (Acc) and 9.21% (F1) under the default setting. Second, removing MAT results in a smaller performance decrease, with drops of 1.76% (Acc) and 1.92% (F1) under the default missing rate. Interestingly, when AR is removed, Acc increases by 0.71% under the de-

Table 3. Comparison of Ablation Results, presented as the average of 5-fold.

Key Modules	Modal Miss Rate η								
	Default			0.4			0.5		
	Acc	Rec	F1	Acc	Rec	F1	Acc	Rec	F1
w/o ReFlow	66.96	66.69	66.49	63.96	63.77	63.68	57.59	57.27	57.20
w/o MAT	74.20	73.92	73.78	68.72	68.52	68.47	62.37	61.80	61.59
w/o AR	**76.67**	**76.56**	**76.50**	67.66	67.52	67.44	61.31	61.19	61.06
Ours	75.96	75.76	75.70	**70.67**	**70.55**	**70.45**	**62.54**	**62.36**	**62.17**

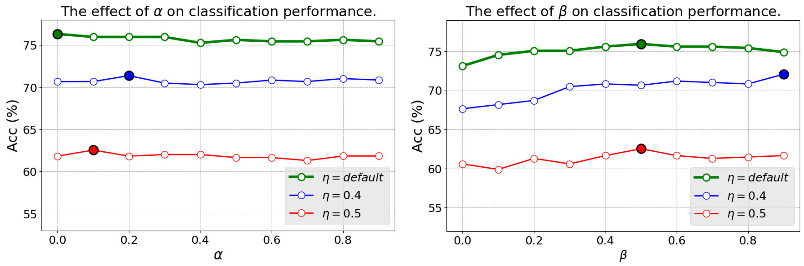

Fig. 2. The Effect of Parameters α and β on Experimental Performance

fault setting compared to the full model. This phenomenon may be because, at low missing rates, the model relies less on recovered data and requires less latent distribution regularization. However, as the missing rate increases, removing AR leads to decreases in Acc of 3.01% (for $\eta = 0.4$) and 1.23% (for $\eta = 0.5$) relative to the full model. This indicates that in the absence of reconstruction experience, a higher missing rate allows AR to better stabilize the modalities.

We analyze the parameters α and β from Eq 6. In the three-class classification task, we conduct experiments by varying one parameter while fixing the other. Fig.2 shows the impact of parameter variations on Accuracy (Acc) under different scenarios. First, the optimal range for α is between 0 and 0.2, indicating that a small α is beneficial for modality recovery. Second, the optimal value for β is around 0.5, suggesting that a larger β helps stabilize classification performance. Overall, the optimization of model performance with respect to both parameters is relatively stable, demonstrating the robustness of the model.

3.4 Conclusions

We propose AGDiC, aiming to address the issues caused by incomplete modalities during model training and inference, especially focus on the paradigm lacking reconstruction experience in recovery strategies. Based on fMRI, sMRI, and PET, AGDiC introduces a novel flow-based adaptive graph learning and modality regularization strategy that not only effectively captures the structural features of each modality but also guides the model to learn precise modality representations. Experimental results on the ADNI dataset demonstrate that our model achieves state-of-the-art performance across multiple tasks and scenarios.

Acknowledgments. This work was partly supported by National Key Research and Development Program of China No. 2023YFC3304202, and Ningbo Municipal Public Welfare Technology Research Project of China No. 2023S023, 2024S055.

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

References

1. Aisen, P.S., Petersen, R.C., Donohue, M., Weiner, M.W., Initiative, A.D.N., et al.: Alzheimer's disease neuroimaging initiative 2 clinical core: progress and plans. Alzheimer's & Dementia **11**(7), 734–739 (2015)
2. Asuero, A.G., Sayago, A., González, A.: The correlation coefficient: An overview. Critical reviews in analytical chemistry **36**(1), 41–59 (2006)
3. Azad, R., Khosravi, N., Dehghanmanshadi, M., Cohen-Adad, J., Merhof, D.: Medical image segmentation on mri images with missing modalities: A review. arXiv preprint arXiv:2203.06217 (2022)
4. Baumgart, M., Snyder, H.M., Carrillo, M.C., Fazio, S., Kim, H., Johns, H.: Summary of the evidence on modifiable risk factors for cognitive decline and dementia: a population-based perspective. Alzheimer's & Dementia **11**(6), 718–726 (2015)
5. Chen, Z., Liu, Y., Zhang, Y., Li, Q., Initiative, A.D.N., et al.: Orthogonal latent space learning with feature weighting and graph learning for multimodal alzheimer's disease diagnosis. Medical Image Analysis **84**, 102698 (2023)
6. Cheng, H., Zhou, J.T., Tay, W.P., Wen, B.: Attentive graph neural networks for few-shot learning. In: 2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR). pp. 152–157. IEEE (2022)
7. Dinh, L., Sohl-Dickstein, J., Bengio, S.: Density estimation using real nvp. arXiv preprint arXiv:1605.08803 (2016)
8. Garcia, N.C., Morerio, P., Murino, V.: Modality distillation with multiple stream networks for action recognition. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 103–118 (2018)
9. Gu, Y., Peng, S., Li, Y., Gao, L., Dong, Y.: Fc-hgnn: A heterogeneous graph neural network based on brain functional connectivity for mental disorder identification. Information Fusion **113**, 102619 (2025)
10. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (2022)
11. Lian, Z., Chen, L., Sun, L., Liu, B., Tao, J.: Gcnet: Graph completion network for incomplete multimodal learning in conversation. IEEE Transactions on pattern analysis and machine intelligence **45**(7), 8419–8432 (2023)
12. Lin, Y., Gou, Y., Liu, Z., Li, B., Lv, J., Peng, X.: Completer: Incomplete multi-view clustering via contrastive prediction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11174–11183 (2021)
13. Ou, Z., Jiang, C., Liu, Y., Zhang, Y., Cui, Z., Shen, D.: A graph-embedded latent space learning and clustering framework for incomplete multimodal multiclass alzheimer's disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 45–55. Springer (2024)
14. Pan, Y., Liu, M., Xia, Y., Shen, D.: Disease-image-specific learning for diagnosis-oriented neuroimage synthesis with incomplete multi-modality data. IEEE transactions on pattern analysis and machine intelligence **44**(10), 6839–6853 (2021)

15. Parthasarathy, S., Sundaram, S.: Training strategies to handle missing modalities for audio-visual expression recognition. In: Companion Publication of the 2020 International Conference on Multimodal Interaction. pp. 400–404 (2020)
16. Pham, H., Liang, P.P., Manzini, T., Morency, L.P., Póczos, B.: Found in translation: Learning robust joint representations by cyclic translations between modalities. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 6892–6899 (2019)
17. Qian, S., Wang, C.: Com: Contrastive masked-attention model for incomplete multimodal learning. Neural Networks **162**, 443–455 (2023)
18. Tarawneh, R., Holtzman, D.M.: The clinical problem of symptomatic alzheimer disease and mild cognitive impairment. Cold Spring Harbor perspectives in medicine **2**(5), a006148 (2012)
19. 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). https://doi.org/https://doi.org/10.1006/nimg.2001.0978, https://www.sciencedirect.com/science/article/pii/S1053811901909784
20. Wang, C., Piao, S., Huang, Z., Gao, Q., Zhang, J., Li, Y., Shan, H., Initiative, A.D.N., et al.: Joint learning framework of cross-modal synthesis and diagnosis for alzheimer's disease by mining underlying shared modality information. Medical Image Analysis **91**, 103032 (2024)
21. Wang, Y., Cui, Z., Li, Y.: Distribution-consistent modal recovering for incomplete multimodal learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 22025–22034 (2023)
22. Waswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., Polosukhin, I.: Attention is all you need. In: NIPS (2017)
23. Xie, G., Huang, Y., Wang, J., Lyu, J., Zheng, F., Zheng, Y., Jin, Y.: Cross-modality neuroimage synthesis: A survey. ACM computing surveys **56**(3), 1–28 (2023)
24. Yin, Q., Wu, S., Wang, L.: Unified subspace learning for incomplete and unlabeled multi-view data. Pattern Recognition **67**, 313–327 (2017)
25. Yuan, L., Wang, Y., Thompson, P.M., Narayan, V.A., Ye, J., Initiative, A.D.N., et al.: Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. NeuroImage **61**(3), 622–632 (2012)
26. Zhang, C., Cui, Y., Han, Z., Zhou, J.T., Fu, H., Hu, Q.: Deep partial multi-view learning. IEEE transactions on pattern analysis and machine intelligence **44**(5), 2402–2415 (2020)
27. Zhang, Y., He, X., Chan, Y.H., Teng, Q., Rajapakse, J.C.: Multi-modal graph neural network for early diagnosis of alzheimer's disease from smri and pet scans. Computers in Biology and Medicine **164**, 107328 (2023)
28. Zhou, Q., Zou, H., Jiang, H., Wang, Y.: Incomplete multimodal learning for visual acuity prediction after cataract surgery using masked self-attention. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 735–744. Springer (2023)
29. Zhou, T., Liu, M., Thung, K.H., Shen, D.: Latent representation learning for alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data. IEEE transactions on medical imaging **38**(10), 2411–2422 (2019)