Uncertainty-Aware Multimodal MRI Fusion for HIV-Associated Asymptomatic Neurocognitive Impairment Prediction

Zige Chen[1†], Haonan Qin[1†], Wei Wang[2], Zhongkai Zhou[2], Chen Zhao[1], Yuqi Fang[1*] and Caifeng Shan[1*]

[1] School of Intelligence Science and Technology at Nanjing University, Jiangsu, China
yqfang@nju.edu.cn
[2] Department of Radiology, Beijing Youan Hospital, Capital Medical University, Beijing, China
[†]Equal contribution [*]Co-corresponding author

Abstract. Asymptomatic neurocognitive impairment (ANI) is an early stage of HIV-associated neurocognitive disorder. Recent studies have investigated magnetic resonance imaging (MRI) for ANI analysis, but most of them rely on single modality, neglecting to utilize complementary information derived from multiple MRI modalties. For a few multimodal MRI fusion studies, they usually suffer from "modality laziness", where dominant modalities suppress weaker ones due to misalignment and scale disparities, limiting fusion efficacy. To address these issues, we propose Uncertainty-aware Multimodal MRI Fusion (UMMF), a novel framework integrating structural MRI, functional MRI, and diffusion tensor imaging for ANI identification. The UMMF employs modality-specific encoders with an uncertainty-aware alternating unimodal training strategy to reduce modality dominance and enhance feature extraction. Moreover, a random network prediction method is designed to estimate uncertainty weights for each modality, enabling robust uncertainty-aware fusion that prioritizes reliable modalities. Extensive experiments demonstrate UMMF's superior performance over SOTA methods, achieving significant improvements in prediction accuracy. Additionally, our approach can help identify critical brain regions associated with ANI, offering potential clinical biomarkers for its early intervention. Our code is available at https://github.com/IsaacKingCzg/IK_MICCAI25_UMMF.

Keywords: HIV-associated ANI · Multimodal MRI fusion · Alternating unimodal training · Uncertainty weighting.

1 Introduction

Asymptomatic neurocognitive impairment (ANI) is an early stage of HIV-associated neurocognitive disorder [1]. ANI typically presents with no obvious clinical symptoms, but if not identified and treated promptly, it can eventually progress to HIV-associated dementia, an irreversible stage. Therefore, early diagnosis of ANI is crucial for timely treatment and disease prevention.

Currently, clinical methods for diagnosing ANI primarily rely on using a series of neurocognitive scales to assess patients [2]. But these assessment methods have a certain degree of subjectivity, leading to varying interpretations among clinicians [3]. To address this issue, Magnetic Resonance Imaging (MRI) is increasingly used for ANI analysis, which provides clinicians with an objective companion diagnostic tool. The commonly used MRI includes structural MRI (sMRI), functional MRI (fMRI) and diffusion tensor imaging (DTI). And many studies have investigated brain analysis with these modalities [4,5,6,7,8,9], but they typically utilize a single modality and ignore cross-modal information derived from multiple MRI modalities.

Recently, the advancement of multimodal fusion [10,11,12,13] techniques has been developed to further enhance prediction outcomes. For instance, Zhu et al. [11] used an attention-based method to combine DTI and fMRI for brain disease diagnosis, while Fang et al. [12] employed attention-based deep learning techniques to predict ANI by combining sMRI and fMRI. Zhao et al. [14] conducted neonatal brain development using sMRI, and DTI. However, existing technologies have some key limitations: (i) For disease prediction tasks, most methods only consider two MRI modalities, without fully accounting for the inherent interdependencies among sMRI, fMRI, and DTI, while comprehensive MRI information is crucial for ANI diagnosis [1]. (ii) Existing multimodal fusion methods commonly suffer from the issue of modality laziness [15], which occurs due to imperfect modality alignment and differing data scales, causing the dominant modality to suppress weaker modalities during joint optimization [16,17]. This ultimately prevents the full potential of modality fusion from being realized [18]. (iii) Most multimodal frameworks treat different modalities equally, overlooking their varying contributions to ANI prediction and the impact of sample uncertainty on prediction results.

To address these limitations, we propose **U**ncertainty-aware **M**ultimodal **MR**I **F**usion (UMMF), a novel framework integrating sMRI, fMRI, and DTI for ANI prediction. Specifically, we first extract features from each modality using separate encoders. Then an uncertainty-aware alternating unimodal training strategy to ensure independent optimization and reduce modality dominance, where random network prediction [19] method is introduced to estimate uncertainty weights for each modality, enhancing fusion robustness. Our contributions are summarized as follows:

- We propose a novel framework, termed UMMF, which integrates sMRI, fMRI, and DTI for ANI prediction, considering the combined impact of these modalities.
- We introduce an uncertainty-aware unimodal alternating training method to address modality laziness, with uncertainty-weighted fusion to improve robustness.
- We validate our method through extensive experiments and identify brain regions as potential clinical biomarkers for ANI analysis.

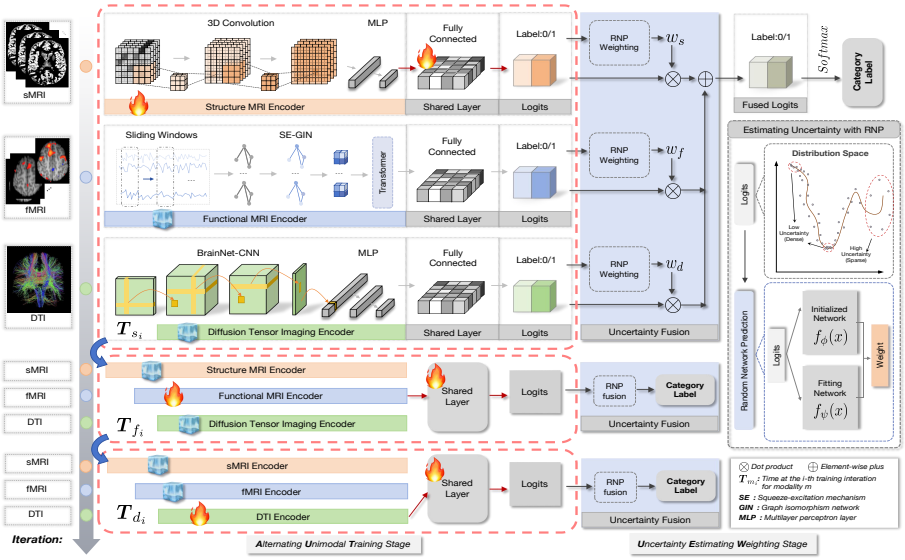

Fig. 1. Overall framework of our proposed **U**ncertainty-aware **M**ultimodal **M**RI **F**usion (UMMF). We design an uncertainty-aware alternating unimodal training strategy based on sMRI, fMRI and DTI, which consists **A**lternating **U**nimodal **T**raining stage and **U**ncertainty **E**stimating **W**eighting stage.

2 Methodology

2.1 Overview

An overview of the proposed Uncertainty-aware Multimodal MRI Fusion for HIV-associated ANI prediction is illustrated in Fig. 1. We design three different encoders for sMRI, fMRI, and DTI, respectively, to ensure that features from all three modalities are fully extracted. We then design a shared layer for the three modalities to capture the common features across them. In terms of training strategy, we design an uncertainty-aware alternating unimodal training approach, which consists of two stages. In the first stage, only the shared layer and the encoder of a single modality are updated during each training phase, addressing the issue of modality laziness. In the second stage, we train the random network prediction [19] module to fit the distribution space of the logits obtained from each modality's network, estimating the uncertainties of the three modalities and performing adaptive weighting, which ensures better fusion of the logits features from all three modalities. Finally, the fused features are input into a Softmax function for ANI prediction.

2.2 Multimodal Feature Encoding

Structural MRI Feature Encoder. To capture high-resolution anatomical features, we use a 3D CNN [20], processing T1-weighted sMRI volumes through

four sequential blocks with $5 \times 5 \times 5$ convolutions, Leaky ReLU activation, max pooling, and batch normalization. The resulting 3D feature maps are vectorized and compressed into a 256-dimensional structural embedding $F_s \in \mathbb{R}^{256}$ using three MLP layers, capturing multi-scale structural patterns from local cortical thickness to global volumetric changes, providing anatomical correlates for ANI-related neurodegeneration.

Functional MRI Feature Encoder. To extract hierarchical spatiotemporal representations from fMRI, we use a hybrid graph-transformer architecture. After AAL atlas parcellation (116 ROIs), sliding window segmentation constructs dynamic functional connectivity graphs based on Pearson correlation. A four-layer graph isomorphism network (GIN) [21] with squeeze-excitation blocks models spatial interactions, while a two-head transformer captures temporal dynamics. The final feature vector $F_f \in \mathbb{R}^{256}$ is obtained through summation pooling and three MLP layers, preserving spatial connectivity and temporal transitions linked to ANI-related disruptions.

Diffusion Tensor Imaging Feature Encoder. We use BrainNet-CNN [22] to model white matter microstructural changes via structural connectivity matrices. After segmenting the brain into 116 regions-of-interest (ROIs) using the AAL atlas, a 116×116 fiber density matrix is constructed from DTI tractography. The encoder processes connectivity patterns with edge-to-edge convolutional blocks, using orthogonal 1D filters to capture multi-scale features, followed by edge-to-node spatial aggregation and node-to-graph global pooling. Three fully connected layers with dropout ($p = 0.5$) project features into a 256-dimensional vector $F_d \in \mathbb{R}^{256}$, preserving hierarchical patterns from local to global connections. This architecture targets HIV-related white matter degeneration with neuroimaging-optimized learnable filters.

2.3 Uncertainty-Aware Alternating Unimodal Training

In multimodal MRI fusion tasks, joint optimization strategies often allow dominant modalities with richer disease-related information to overshadow weaker ones [16,17], leading to modality laziness [15]. And most multimodal frameworks treat all modalities equally, neglecting their different contributions to ANI prediction. To resolve these, we design an uncertainty-aware alternating unimodal training method, which consists of two stages: alternating unimodal training (AUT) stage and uncertainty estimating weighting (UEW) stage. The first stage aims to independently optimize each modality's encoder through sequential single-modal learning phases. Then the second stage employs the random network prediction (RNP) [19] method to perform adaptive uncertainty weighting.

Alternating Unimodal Training Stage. Let the full dataset be partitioned by modality as $P = \{P_s, P_f, P_d\}$ corresponding to sMRI, fMRI, and DTI, respectively. For modality m, its data is defined as:

$$P_m = (X_m, Y_m) = \{(x_{m_n}, y_{m_n})\}_{n=1}^{N}, \tag{1}$$

where N denotes the total number of subjects. We design modality-specific prediction functions as composite mappings:

$$h_m = g \circ e_m, \tag{2}$$

where, e_m represents the encoder for modality m, while g is a shared single linear layer across all modalities (Fig. 1). During AUT stage, we perform I iterations aligned with the subject count N. At each iteration i, we sequentially optimize the three modalities through dedicated training phases T_m, where $m \in \{sMRI, fMRI, DTI\}$. For modality m, the objective minimizes:

$$\mathcal{L}_{T_m} = \mathbb{E}_{(x,y)\sim P_m} \left[\ell(g(e_m(x;\theta_m));\phi), y\right], \tag{3}$$

where θ_m and ϕ denote trainable parameters of encoder e_m and shared layer g, respectively. This alternating unimodal optimization ensures each encoder learns discriminative features without interference from dominant modalities, effectively addressing the "modality laziness" issue. We also apply Recursive Least Squares correction [15] to orthogonalize the shared layer gradient at each update, reducing gradient vanishing and preserving cross-modal information.

Uncertainty Estimating Weighting Stage. As shown in Fig. 1, after completing the AUT stage, the UEW stage estimates modality-specific uncertainties through RNP modules. For each modality m, the shared layer outputs logits l_m, which are processed by a dedicated RNP_m unit. Each RNP_m contains a *Fitting Network* f_ψ trained to approximate outputs of a fixed-weight *Initialized Network* f_ϕ with random initialization, aiming to fit the distribution space of l_m. The fitting network better approximates low-uncertainty samples in densely populated regions while struggling with high-uncertainty samples in sparse regions.

We optimize RNP_m by minimizing the Mean Squared Error between both networks' outputs with L_2 regularization:

$$\psi^* = \arg\min_{\psi} \sum_{n=1}^{N} \|f_\psi(l_{m_n}) - f_\phi(l_{m_n})\|_2^2 + \lambda\|\psi\|_2^2, \tag{4}$$

where ψ denote trainable parameters of f_ψ, and λ controls regularization strength. For each modality $m \in \{sMRI, fMRI, DTI\}$, we compute its fusion map u_m using the trained RNP_m module:

$$u_m = \|f_\psi(l_m) - f_\phi(l_m)\|_2^2, \tag{5}$$

where lower values indicate higher reliability (denser regions in feature space). Next, we can obtain the weight w for each modality by cross-assigning u_m:

$$w_s = u_f + u_d, \tag{6}$$

where w_s is the weights of sMRI. The final fused logits L are obtained through uncertainty-weighted summation:

$$L = w_s \otimes l_s + w_f \otimes l_f + w_d \otimes l_d, \tag{7}$$

where w_f and w_d are the weights of fMRI and DTI, $\otimes$ denotes element-wise multiplication, dynamically scaling each modality's prediction confidence.

3 Experiment and Discussion

Materials and Image Preprocessing We use the ANID dataset from Beijing Youan Hospital, consisting of 68 HIV-associated ANI patients and 69 HCs, each with matching sMRI, fMRI, and DTI data. All sMRI data are preprocessed with FreeSurfer [23], including bias field correction, motion correction, intensity normalization, MNI registration, and skull stripping. fMRI data are processed with DPARSF [24], involving steps such as discarding the first 10 volumes, slice timing correction, motion correction, bandpass filtering (0.01-0.10 Hz), nuisance signal removal, MNI normalization, and partitioning into 116 ROIs based on the AAL atlas. The regional mean fMRI time series are extracted for each subject. DTI assesses white matter structure by measuring water diffusion, constructing a 116×116 symmetric fiber length matrix based on tractography and fiber path similarities.

Experimental Setup We conducted all experiments using a 3-fold cross-validation, repeated 5 times. The following hyperparameters were used across all experiments: training was performed for 70 epochs with the Adam optimizer, a batch size of 6, and a learning rate of 6×10^{-4}. The regularization parameter λ in Eq. (4) was set to 0.0001. A detailed description of the encoder configurations is provided in Section 2.2. To ensure a fair comparison, all baseline methods were tuned using the same procedure, with hyperparameters optimized for optimal performance under identical conditions.

Competing Methods We use the ANID dataset for ANI vs. HC classification and compare the proposed UMMF with ten competing methods, including: 1) DA-MIDL [5] and 2) DL4AD [6], which use DA-MIDL and DL4AD models for sMRI prediction; 3) GCN [25] and 4) GAT [26], which capture fMRI features using graph convolutional networks and graph attention networks, respectively; 5) ConCeptCNN [7] and 6) GCN-A [9], which analyze DTI using convolution-based and graph neural network-based strategies, respectively; 7) ASFF [12], which uses an attention-enhanced strategy to fuse fMRI and sMRI; 8) MTAN [11],

Table 1. Results (%) of UMMF and ten competing methods on ANID dataset.

Method	AUC (%)	ACC (%)	F1 (%)	SEN (%)	SPE (%)	PRE (%)
DA-MIDL	57.29±1.05	56.38±6.12	56.10±11.11	58.73±17.96	59.90±10.29	60.28±1.71
DL4AD	56.72±7.26	56.72±7.26	58.02±5.89	59.86±6.11	52.96±6.40	56.39±6.27
GCN	63.53±8.88	57.70±8.24	58.07±6.54	60.85±15.91	57.59±17.07	59.31±13.04
GAT	65.80±13.10	57.70±13.33	59.81±9.22	63.77±15.52	56.46±28.89	61.65±19.88
ConCeptCNN	61.45±9.53	57.94±11.88	55.69±15.47	55.56±19.57	60.32±4.49	56.53±11.11
GCN-A	60.17±4.23	56.35±2.24	55.83±7.24	57.14±13.47	55.56±11.88	56.44±1.90
ASFF	68.66±5.73	65.24±4.90	63.98±4.85	61.91±6.73	68.57±9.33	66.86±6.25
MTAN	71.50±8.72	65.87±4.89	66.51±5.69	68.25±8.09	63.49±5.94	65.14±4.74
TMF	57.94±2.97	57.75±4.39	55.19±4.98	52.38±7.78	63.49±5.94	58.97±2.91
MaskGNN	72.94±4.12	63.45±6.61	63.17±8.37	64.56±12.79	62.32±5.42	62.47±5.72
UMMF (Ours)	**74.35±1.96**	**69.05±5.83**	**68.96±6.11**	**69.05±5.83**	**73.94±2.24**	**69.88±5.48**

which designs a triplet attention network to fuse fMRI and DTI; 9) TMF [14], which employs an attention mechanism to fuse DTI and sMRI; 10)MaskGNN [27], which uses a Masked Graph Neural Network, where sMRI, fMRI, and DTI features are used as node features. For all methods, we perform three-fold cross-validation on the dataset and record the average results. Six metrics are used for evaluation, including area under the ROC curve (AUC), accuracy (ACC), F1-score (F1), sensitivity (SEN), specificity (SPE), and precision (PRE).

Classification Results Table 1 reports the mean and standard deviation results of various methods for ANI diagnosis. It can be observed that methods utilizing three modalities generally outperform those employing only one or two modalities, demonstrating that fusing information from multiple MRI modalities enhances diagnostic accuracy. Furthermore, our UMMF outperforms another three-modal fusion method (MaskGNN); for instance, UMMF improves AUC and ACC scores by 1.41% and 5.6%, respectively, while also yielding a lower standard deviation. This superior performance may be attributed to the fact that MaskGNN merely concatenates sMRI, fMRI, and DTI features and feeds them into a GNN—treating each modality equally and training them uniformly—which gives rise to the modality laziness problem and consequently results in suboptimal outcomes.

Ablation Study As shown in the upper part of Table 2 (from UMMF-S to UMMF-SD), we investigate the influence of each modality on ANI diagnosis, where S refers to sMRI, F refers to fMRI and D refers to DTI. The results show that two-modal fusion outperforms single-modality methods, and three-modal fusion yields the best performance, highlighting the advantages of UMMF in addressing modality laziness and the necessity of three-modal fusion.

We also investigate the effect of the Uncertainty-aware alternating unimodal training method: 1) UMMF-oA, which replaces the alternating unimodal training by joint optimization strategy, and sMRI features, DTI features and fMRI features are directly concatenated for prediction; 2) UMMF-oR, which removes RNP module and directly sums the three logits for prediction. As shown in the

Table 2. Ablation study for the eight variants of the proposed UMMF.

Method	AUC (%)	ACC (%)	F1 (%)	SEN (%)	SPE (%)	PRE (%)
UMMF-S	59.69±5.80	59.85±5.67	59.02±6.75	59.85±5.67	63.44±6.09	59.27±6.76
UMMF-F	67.01±1.22	63.49±4.49	63.40±4.29	63.49±4.49	63.39±11.77	66.28±5.60
UMMF-D	52.59±7.85	56.67±4.25	55.94±4.93	56.67±4.25	66.50±4.02	56.34±5.36
UMMF-SF	67.71±12.47	61.90±10.29	61.99±10.19	61.90±10.29	65.39±11.09	62.73±9.63
UMMF-DF	68.70±1.47	65.08±1.12	65.03±1.06	65.08±1.12	62.28±6.27	66.32±1.71
UMMF-SD	60.53±4.03	60.32±9.59	59.98±9.49	60.32±9.59	57.11±18.83	61.15±8.92
UMMF-oA	52.93±4.35	54.81±5.54	53.97±5.87	54.81±5.54	64.33±11.73	55.28±6.12
UMMF-oR	53.20±3.77	53.17±4.05	53.22±4.00	53.17±4.05	58.33±6.24	53.96±4.11
UMMF(Ours)	**74.35±1.96**	**69.05±5.83**	**68.96±6.11**	**69.05±5.83**	**73.94±2.24**	**69.88±5.48**

lower part of Table 2, removing either module results in a significant performance drop, highlighting the critical role of each component in our framework. The simple concatenation approach based on joint optimization is affected by modality laziness, causing the subordinate modalities to underperform and limiting the potential of the dominant modality. The strategy of simply summing the logits fails to account for the varying contributions of each modality, resulting in heightened sensitivity of the model under uncertainty.

Discriminative Brain Regions. Based on the UMMF multimodal deep learning model, which integrates Grad-CAM [28] for sMRI and DTI with attention analysis for fMRI, we successfully validated the thalamus, parahippocampal gyrus, and cerebellum as cross-modal biomarkers for HIV-associated neurocognitive disorder [29,30,31]. Specifically, (1) the model accurately captured multilayer damage features in the thalamus, including structural atrophy, white matter microstructural abnormalities (reduced MD), and metabolic decline, aligning with the classical theory of thalamic dysfunction as an "information integration hub" [29,30]; (2) it revealed a spatiotemporal coupling pattern between parahippocampal atrophy and default network disconnection [31], as well as dynamic compensatory reorganization in the cerebello-cortical circuitry [31]. This study demonstrates that UMMF can identify key brain regions associated with ANI, with interpretable features that align closely with existing unimodal evidence, further validating the effectiveness of UMMF.

4 Conclusion

In this paper, we introduced an Uncertainty-aware Multimodal MRI Fusion (UMMF) framework for predicting HIV-associated asymptomatic neurocognitive impairment (ANI). By alternatively training modality-specific encoders for sMRI, fMRI and DTI, UMMF effectively reduces modality dominance and enhances feature extraction using an uncertainty-aware alternating unimodal training strategy. Our experiments demonstrate that UMMF outperforms other methods, significantly improving prediction accuracy and identifying ANI-related brain regions. Future work will focus on exploring disease-specific encoders, investigating alternative uncertainty estimation methods, and extending the

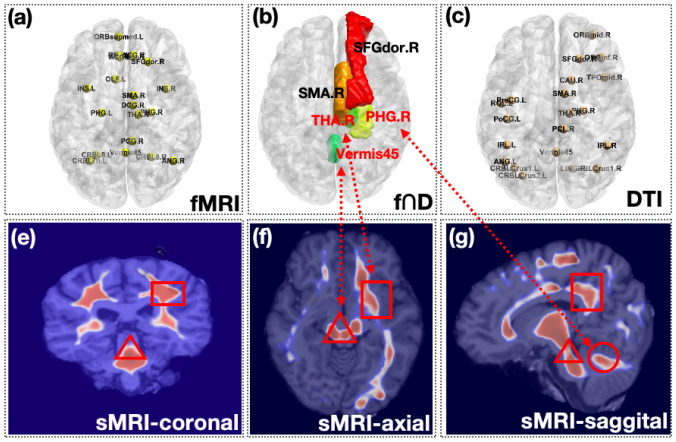

Fig. 2. Illustration of brain regions associated with ANI, identified through fMRI (a), DTI (c), and sMRI (e, f, g) analyses. Panel (b) highlights the overlapping regions of interest (ROIs) between DTI and fMRI.

framework to other brain diseases to enhance its generalizability and clinical applicability.

Acknowledgments. This work was supported by AI & AI for Science Project of Nanjing University (4810-14380006), Fundamental Research Funds for the Central Universities (4810-16003302), Open Project of the Henan Clinical Research Center for Infectious Diseases (No. KFKT202401), and Beijing Hospital Authority Clinical Medicine Development Special Funding Support (No. ZLRK202333).

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

References

1. Han, S., et al.: Altered regional homogeneity and functional connectivity of brain activity in young HIV-infected patients with asymptomatic neurocognitive impairment. Frontiers in Neurology **13** (2022) 982520
2. Wei, J., et al.: The prevalence of frascati-criteria-based HIV-associated neurocognitive disorder (HAND) in HIV-infected adults: A systematic review and meta-analysis. Frontiers in Neurology **11** (2020) 581346
3. Gandhi, N.S., et al.: A comparison of performance-based measures of function in HIV-associated neurocognitive disorders. Journal of Neurovirology **17** (2011) 159–165
4. Zhang, H., et al.: Classification of brain disorders in rs-fMRI via local-to-global graph neural networks. IEEE Transactions on Medical Imaging **42**(2) (2022) 444–455

5. Zhu, W., et al.: Dual attention multi-instance deep learning for alzheimer's disease diagnosis with structural mri. IEEE Transactions on Medical Imaging **40**(9) (2021) 2354–2366
6. Xing, Y., et al.: Classification of smri images for alzheimer's disease by using neural networks. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Springer Nature Switzerland (2022)
7. Chen, M., et al.: Conceptcnn: A novel multi-filter convolutional neural network for the prediction of neurodevelopmental disorders using brain connectome. Medical Physics **49**(5) (2022) 3171–3184
8. Kawahara, J., et al.: Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage **146** (2017) 1038–1049
9. Sang, Y., Li, W.: Classification study of alzheimer's disease based on self-attention mechanism and DTI imaging using GCN. IEEE Access **12** (2024) 24387–24395
10. Sung, Y.L., Li, L., Lin, K., Gan, Z., Bansal, M., Wang, L.: An empirical study of multimodal model merging. arXiv preprint (2023) arXiv:2304.14933.
11. Zhu, Q., et al.: Multimodal triplet attention network for brain disease diagnosis. IEEE Transactions on Medical Imaging **41**(12) (2022) 3884–3894
12. Fang, Y., et al.: Attention-enhanced fusion of structural and functional mri for analyzing HIV-associated asymptomatic neurocognitive impairment. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer Nature Switzerland (2024)
13. Xie, X., Cui, Y., Tan, T., Zheng, X., Yu, Z.: Fusionmamba: dynamic feature enhancement for multimodal image fusion with mamba. Visual Intelligence **2** (2024)
14. Zhao, H., Cai, H., Liu, M.: Transformer based multi-modal mri fusion for prediction of post-menstrual age and neonatal brain development analysis. Medical Image Analysis **94** (2024) 103140
15. Zhang, A., et al.: Multimodal representation learning by alternating unimodal adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2024)
16. Du, C., Teng, J., Li, T., Liu, Y., Yuan, T., Wang, Y., Yuan, Y., Zhao, H.: On uni-modal feature learning in supervised multi-modal learning. In: International Conference on Machine Learning, ICML 2023, Honolulu, Hawaii, USA, PMLR (July 23-29, 2023 2023) 8632–8656
17. Peng, X., Wei, Y., Deng, A., Wang, D., Hu, D.: Balanced multimodal learning via on-the-fly gradient modulation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, IEEE (June 18-24, 2022 2022) 8228–8237
18. Sun, Y., Mai, S., Hu, H.: Learning to balance the learning rates between various modalities via adaptive tracking factor. IEEE Signal Processing Letters **28** (2021) 1650–1654
19. Zhang, A., et al.: Multimodal representation learning by alternating unimodal adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2024)
20. Ji, S., et al.: 3d convolutional neural networks for human action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(1) (2012) 221–231
21. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? In: International Conference on Learning Representations. (2019)
22. Kawahara, J., et al.: Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage **146** (2017) 1038–1049
23. Fischl, B.: Freesurfer. NeuroImage **62**(2) (2012) 774–781

24. Yan, C., Zang, Y.: DPARSF: A matlab toolbox for "pipeline" data analysis of resting-state fMRI. Frontiers in Systems Neuroscience **4** (2010) 1377
25. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
26. Velickovic, P., et al.: Graph attention networks. stat 1050.20 (2017) 10–48550
27. Qu, G., et al.: Integrated brain connectivity analysis with fMRI, DTI, and sMRI powered by interpretable graph neural networks. ArXiv (2024) arXiv-2408.
28. Selvaraju, R., Cogswell, M., Das, A., Vedantam, R., et al.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision. (2017) 618–626
29. Pfefferbaum, A., et al.: Accelerated aging of selective brain structures in human immunodeficiency virus infection: a controlled, longitudinal magnetic resonance imaging study. Neurobiology of Aging **35**(7) (2014) 1755–1768
30. Hammoud, D.A., et al.: Global and regional brain hypometabolism on fdg-pet in treated HIV-infected individuals. Neurology **91**(17) (2018) e1591–e1601
31. O'Connor, E.E., Zeffiro, T.A., Zeffiro, T.A.: Brain structural changes following HIV infection: Meta-analysis. AJNR American Journal of Neuroradiology **39**(1) (2018) 54–62