CMD-AD: Cross-Modality Distillation from fMRI-EEG for Single-Modality Alzheimer's Disease Diagnosis

Xin Lin[1†], Yuxiao Liu[1†], Wenfei Yu[1†], Yulong Dou[1], Weiqi Li[1], Minhui Tan[1], Yuanzhe He[1], Zhiming Cui[1(✉)], and Dinggang Shen[1,2,3(✉)]

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
[2] Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200230, China
[3] Shanghai Clinical Research and Trial Center, Shanghai 201210, China
cuizm.neu.edu@gmail.com, Dinggang.Shen@gmail.com

Abstract. Alzheimer's disease (AD) disrupts both the spatial organization and temporal dynamics of brain functional connectivity, which can be jointly characterized by fMRI and EEG. fMRI-derived functional connectivity networks (FCNs) encode multiscale spatial topology, and EEG-derived FCNs capture multiband temporal coupling patterns, providing complementary information for AD diagnosis. However, paired fMRI-EEG acquisitions are rarely available in clinical practice, and existing fusion methods require both modalities at inference, limiting real-world deployment. To address this challenge, we propose **CMD-AD** (**C**ross-**M**odality **D**istillation for **AD** diagnosis), a three-stage framework that distills fused spatiotemporal knowledge into single-modality encoders to improve their performance. In Stage 1, hierarchical graph attention aggregation (HGAA) encoders are independently pretrained on each modality to extract multiscale fMRI spatial features and multiband EEG temporal features. In Stage 2, scale-wise bidirectional cross-modality fusion (BCMF) integrates fMRI and EEG representations at each parcellation scale, producing fused spatiotemporal targets. In Stage 3, teacher-student distillation transfers the fused spatiotemporal targets back to each modality-specific encoder, enabling accurate single-modality diagnostic performance without paired inputs. Evaluated on 8 multicenter datasets, CMD-AD improves fMRI-only accuracy from 73.51% to 75.50% and EEG-only accuracy from 63.52% to 65.02%, with AUC increasing from 80.88% to 83.70% and from 71.63% to 73.77%, respectively. These results demonstrate that cross-modality training can effectively enhance single-modality deployment for AD diagnosis. Our code is available at `https://github.com/yizhixiaolinya/Code-MICCAI2026_fMRI-EEG.git`.

Keywords: fMRI-EEG fusion, Cross-modality distillation, Functional connectivity network, Alzheimer's disease, Graph neural networks.

1 Introduction

Alzheimer's disease (AD) induces spatiotemporal disruptions in brain function that can be assessed using functional magnetic resonance imaging (fMRI) and electroen-

† Equal contribution.

cephalography (EEG) [2]. These modalities provide complementary views: fMRI measures blood-oxygen-level-dependent (BOLD) signals with high spatial resolution [10], whereas EEG records neural electrical activity with millisecond temporal resolution across frequency bands [3]. Functional connectivity networks (FCNs) are constructed by extracting region-of-interest (ROI) time series and computing pairwise connectivity, yielding ROI-by-ROI matrices [5]. Accordingly, fMRI-derived FCNs are widely used to characterize spatial network organization across multiscale parcellations, whereas EEG FCNs are typically computed across multiple frequency bands to capture multiband temporal coupling patterns [4].

Motivated by this complementarity, fMRI-EEG fusion frameworks span linear models (*e.g.*, CCA, jICA) [32,17] and deep methods that learn non-linear cross-modality interactions (*e.g.*, MCSP, HC-GNN) [27,29]. However, three challenges still limit their clinical applications. **First**, most frameworks require paired fMRI-EEG at inference [27,29], whereas clinical acquisitions often provide only single-modality [7]. **Second**, their fusion strategies are frequently coarse and lack explicit scale-wise correspondence between multiscale fMRI parcellations and EEG band-informed representations, weakening their cross-modality alignment. **Third**, spatiotemporal representation learning within each modality remains suboptimal. fMRI encoders often under-exploit multiscale FCN organization and long-range interregional interactions [14], whereas EEG encoders often treat frequency bands as generic channels rather than modeling band-specific temporal coupling and cross-band relationships [27].

To address these challenges, we propose **CMD-AD** (Fig. 1), a three-stage cross-modality distillation framework for enhanced single-modality diagnosis. **Stage 1** first pretrains HGAA-based modality-specific encoders to learn discriminative single-modality representations. **Stage 2** then replaces within-modality attention with scale-wise bidirectional cross-attention (BCMF), enabling fMRI and EEG representations to interact at each parcellation scale and producing fused spatiotemporal representations. **Stage 3** finally distills this fused multimodal knowledge back into each single-modality encoder through teacher-student learning, enabling inference when only fMRI or EEG is available. CMD-AD yields consistent gains: fMRI-only accuracy improves from 73.51% to 75.50% and EEG-only accuracy improves from 63.52% to 65.02%, with AUC increasing from 80.88% to 83.70% and from 71.63% to 73.77%, respectively, validating that cross-modality training effectively enhances single-modality AD diagnosis.

2 Method

2.1 fMRI and EEG FCN Graph Construction

We construct fMRI and EEG FCN graphs using shared multiscale Schaefer parcellations [21]. For fMRI, voxel-wise BOLD signals are averaged into ROI time series at each parcellation scale. For EEG, electrode recordings are projected onto cortical source space via sLORETA [18] and aggregated to the same ROIs, establishing scale-wise ROI correspondences that enable subsequent fusion and distillation.

Let $m \in \{\text{fMRI}, \text{EEG}\}$ and $s \in \{1, \ldots, S\}$ (with $s = 1$ as the coarsest scale). At scale s, we define a multichannel FCN graph $\mathcal{G}_s^m = (\mathbf{A}_s^m, \mathbf{H}_s^m)$ with B_{fMRI}=1 and B_{EEG}=5

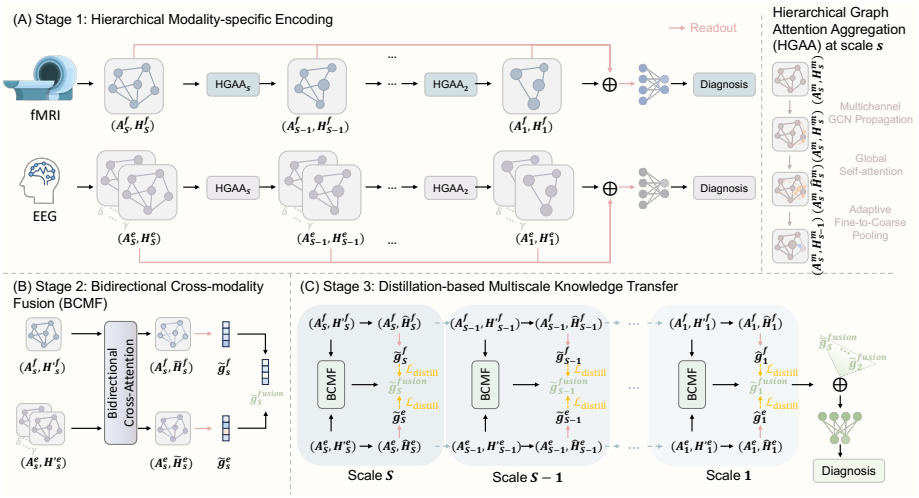

Fig. 1. CMD-AD overview. (A) Stage 1 learns hierarchical modality-specific representations by stacking HGAA blocks over multiscale FCN graphs. (B) Stage 2 replaces the intra-modality global self-attention with BCMF, enabling scale-wise fusion between fMRI and EEG. (C) Stage 3 performs multiscale distillation to supervise each modality-specific encoder.

channels. For each channel c, the adjacency matrix $\mathbf{A}_s^{m,c}$ is computed via Pearson correlation (fMRI) or orthogonalized power envelope correlation (oPEC, EEG) [31], and sparsified by retaining the top 20% off-diagonal edges by absolute value while preserving signed weights [28]. Each node feature concatenates its connectivity fingerprint (FCN row) with the mean and standard deviation of the corresponding ROI time series [13]. Across scales, voxel-overlap weights $\mathbf{W}_{s\to s-1}$ define valid fine-to-coarse assignments $\Omega_{s,r} = \{v \mid \mathbf{W}_{s\to s-1}(v,r) > 0\}$ for hierarchical pooling.

2.2 Stage 1: Hierarchical Modality-Specific Encoding

Stage 1 pretrains two modality-specific encoders for fMRI and EEG that share the same HGAA architecture but use independent parameters. Each HGAA block comprises multichannel GCN propagation, global self-attention, and atlas-constrained adaptive pooling. Stacking these blocks from scale S to 1 yields scale-wise representations that serve as inputs for fusion in Stage 2 and distillation in Stage 3.

Multichannel GCN Propagation. Each HGAA block begins with multichannel GCN propagation over B_m channels to encode local FCN topology. For EEG, the B_{EEG}=5 channels correspond to canonical frequency bands, preserving band-specific features for multiband temporal coupling modeling. For modality m at scale s and channel c, ROI features are updated as:

$$\tilde{\mathbf{H}}_s^{m,c} = \sigma\left((\mathbf{D}_s^{m,c})^{-\frac{1}{2}}\mathbf{A}_s^{m,c}(\mathbf{D}_s^{m,c})^{-\frac{1}{2}}\mathbf{H}_s^{m,c}\mathbf{W}_{\text{GCN}}^{m}\right), \tag{1}$$

where $\mathbf{W}^{m}_{\mathrm{GCN}} \in \mathbb{R}^{d\times d}$ is shared across channels within modality m, and $\mathbf{D}^{m,c}_{s}$ is the degree matrix of $\mathbf{A}^{m,c}_{s}$.

Band-Preserving Global Self-Attention. To model long-range ROI interactions at each scale, we apply a global self-attention module. We first average $\tilde{\mathbf{H}}^{m,c}_{s}$ over channels c to obtain a channel-aggregated representation $\bar{\mathbf{H}}^{m}_{s}$, then apply multi-head self-attention $\mathrm{SA}^{m}(\cdot)$ to $\bar{\mathbf{H}}^{m}_{s}$ and inject the resulting global context back into each channel via a residual connection:

$$\widehat{\mathbf{H}}^{m,c}_{s} = \mathrm{LayerNorm}\big(\tilde{\mathbf{H}}^{m,c}_{s} + \mathrm{SA}^{m}\big(\bar{\mathbf{H}}^{m}_{s}\big)\big)\,. \tag{2}$$

Here, $\bar{\mathbf{H}}^{m}_{s}$ is used only to compute a shared global attention context, while the residual update is applied to each per-channel representation $\tilde{\mathbf{H}}^{m,c}_{s}$. For EEG, this design introduces shared global interactions while preserving frequency-band-specific representations in separate channels.

Adaptive Fine-to-Coarse Pooling. We apply atlas-constrained adaptive pooling to aggregate ROI features from scale s to $s-1$ along the hierarchical parcellation, thereby capturing hierarchical spatial organization. For each coarse ROI r, we attend over its assigned fine ROIs $v \in \Omega_{s,r}$ and compute

$$\mathbf{H}^{m,c}_{s-1}(r,:) = \mathrm{LayerNorm}\left(\sum_{v\in\Omega_{s,r}} a_{r,v}\,\widehat{\mathbf{H}}^{m,c}_{s}(v,:)\right), \tag{3}$$

where $a_{r,v}$ are normalized masked-attention weights over $\Omega_{s,r}$. The pooled features are paired with the scale-$(s-1)$ FCN adjacency for the next HGAA block.

Graph Readout and Multiscale Classification. Finally, we obtain scale-wise graph embeddings and train each modality-specific encoder with single-modality supervision. At scale s, we aggregate channels to produce a single ROI feature matrix and apply global average pooling over R_s ROIs to form $\mathbf{g}^{m}_{s}$. The multiscale embedding is $\mathbf{g}^{m}_{\mathrm{multi}} = \mathrm{Concat}(\mathbf{g}^{m}_{1},\ldots,\mathbf{g}^{m}_{S})$, followed by a modality-specific classifier with loss $\mathcal{L}^{m}_{1} = \mathrm{CE}(\hat{y}^{m}, y)$. The fMRI and EEG branches are trained independently using all available single-modality data.

2.3 Stage 2: Bidirectional Cross-Modality Fusion

Stage 2 replaces the intra-modality global self-attention in Stage 1 with **BCMF** to enable scale-wise bidirectional interaction between modalities. At each scale s, we form channel-aggregated ROI features $\bar{\mathbf{H}}^{m}_{s}$ by averaging $\tilde{\mathbf{H}}^{m,c}_{s}$ over channels c and compute cross-attention in both directions. The cross-modality context is integrated into the target modality through residual connections, and for EEG, the same context is broadcast to each band-specific channel:

$$\widehat{\mathbf{H}}^{m,c}_{s} = \mathrm{LayerNorm}\Big(\tilde{\mathbf{H}}^{m,c}_{s} + \mathrm{CrossAttn}^{m\leftarrow m'}\Big(\bar{\mathbf{H}}^{m}_{s}, \bar{\mathbf{H}}^{m'}_{s}\Big)\Big), \tag{4}$$

where $(m,m') \in \{(\mathrm{fMRI},\mathrm{EEG}),(\mathrm{EEG},\mathrm{fMRI})\}$. The cross-attention context is computed from the channel-aggregated representations $\bar{\mathbf{H}}^{m}_{s}$ and $\bar{\mathbf{H}}^{m'}_{s}$, and is then injected into each target-channel feature through the residual connection. We obtain $\tilde{\mathbf{g}}^{\mathrm{fMRI}}_{s}$ and

$\tilde{\mathbf{g}}_s^{\text{EEG}}$ via the same readout procedure, and form the fused embedding using a learnable scale-wise mixing weight $\lambda_s = \sigma(\tilde{\lambda}_s)$:

$$\mathbf{g}_s^{\text{fuse}} = \lambda_s \, \tilde{\mathbf{g}}_s^{\text{fMRI}} + (1 - \lambda_s) \, \tilde{\mathbf{g}}_s^{\text{EEG}}. \tag{5}$$

Fused embeddings are concatenated across scales for multimodal prediction $\hat{y}^{\text{fuse}}$ and retained as distillation targets for Stage 3.

2.4 Stage 3: Distillation-Based Multiscale Knowledge Transfer

Stage 3 employs teacher-student knowledge distillation to inject the fused spatiotemporal knowledge learned in Stage 2 into each single-modality encoder, enabling robust inference from a single-modality alone. We freeze the fusion branch of Stage 2 as the teacher. For each modality $m \in \{\text{fMRI}, \text{EEG}\}$, a student encoder is initialized from Stage 1 and trained to jointly match the teacher's fused multiscale representations and its softened predictive distribution. Specifically, given fused teacher targets $\{\mathbf{g}_s^{\text{fuse}}\}_{s=1}^{S}$ and student embeddings $\{\mathbf{g}_s^{m}\}_{s=1}^{S}$, the multiscale feature distillation loss is defined as:

$$\mathcal{L}_{\text{feat}}^{m} = \sum_{s=1}^{S} \left\| \mathbf{g}_s^{m} - \mathbf{g}_s^{\text{fuse}} \right\|_2^2 . \tag{6}$$

Additionally, letting $\mathbf{z}^{\text{fuse}}$ and $\mathbf{z}^{m}$ denote the teacher and student logits, respectively, we distill soft targets with temperature τ via:

$$\mathcal{L}_{\text{KD}}^{m} = \tau^2 \, \text{KL}\left(\text{Softmax}\left(\frac{\mathbf{z}^{\text{fuse}}}{\tau} \right) \,\middle\|\, \text{Softmax}\left(\frac{\mathbf{z}^{m}}{\tau} \right) \right). \tag{7}$$

To preserve modality-specific discriminability, a supervised cross-entropy loss $\mathcal{L}_{\text{cls}}^{m} = \text{CE}(\hat{y}^{m}, y)$ is also included, yielding the overall objective function of Stage 3:

$$\mathcal{L}_3 = \sum_{m \in \{\text{fMRI},\text{EEG}\}} \left(\mathcal{L}_{\text{cls}}^{m} + \mathcal{L}_{\text{distill}}^{m} \right), \qquad \mathcal{L}_{\text{distill}}^{m} = \eta \, \mathcal{L}_{\text{feat}}^{m} + \kappa \, \mathcal{L}_{\text{KD}}^{m}, \tag{8}$$

where $\mathcal{L}_{\text{distill}}^{m}$ combines feature-level and logit-level distillation, and η and κ control their relative contributions. At inference, the teacher fusion branch is discarded, and only the distilled student encoder is used to predict $\hat{y}^{m}$ from the available modality.

3 Datasets and Data Preprocessing

We collect **8** datasets spanning **fMRI-only**, **EEG-only**, and **paired fMRI-EEG** cohorts (Table 1). The fMRI-only cohorts comprise ADNI and OASIS [8,15]; the EEG-only cohorts comprise ADSZ, ADFTD, APAVA, and ADFSU [1,16,23,25]; and the paired cohorts comprise BrainLat and PEARL-Neuro [20,6], which are used for fusion in Stage 2 training and distillation in Stage 3 (BrainLat: 15 AD and 17 NC pairs; PEARL-Neuro: 78 NC pairs). Accordingly, Stage 1 pretrains modality-specific encoders using all available single-modality data. Stage 2 optimizes the fusion model using paired

Table 1. Demographic and clinical characteristics of the datasets used in this study.

Dataset	Modality	Group	N	M/F	Age (y)	Clinical score (MMSE/MoCA)
BrainLat [20]	fMRI	AD	46	19/27	76.17 ± 8.33	- / 15.95 ± 4.15
		NC	83	44/39	70.10 ± 7.61	- / 27.02 ± 3.18
	EEG	AD	35	15/20	76.66 ± 7.87	- / 16.92 ± 3.45
		NC	45	20/25	69.63 ± 8.09	- / 25.45 ± 3.41
PEARL-Neuro [6]	fMRI EEG	NC	78	39/39	55.29 ± 3.14	- / -
ADNI [8]	fMRI	AD	133	70/63	72.80 ± 7.30	22.90 ± 2.50 / -
		NC	565	235/330	74.64 ± 7.25	29.10 ± 1.10 / -
OASIS [15]	fMRI	AD	43	28/15	73.15 ± 6.62	25.58 ± 3.35 / -
		NC	207	107/100	72.30 ± 4.46	28.87 ± 1.37 / -
ADFTD [16]	EEG	AD	36	12/24	66.39 ± 7.89	17.75 ± 4.50 / -
		NC	29	18/11	67.90 ± 5.40	30.00 ± 0.00 / -
ADFSU [25]	EEG	AD	160	-	-	- / -
		NC	12	-	-	- / -
APAVA [23]	EEG	AD	12	5/7	72.8 ± 8.0	- / -
		NC	11	-	-	- / -
ADSZ [1]	EEG	AD	24	-	72.00 ± 11.00	- / -
		NC	24	-	69.00 ± 16.00	- / -

fMRI-EEG data, and Stage 3 further utilizes all available data to transfer the learned multimodal knowledge back to single-modality models.

Resting-state fMRI scans are processed using a standardized pipeline [26]: We discard the first 5-10 volumes, perform slice-timing and motion correction, exclude subjects with excessive motion (translation > 2 mm or rotation > 2°), regress nuisance signals (white matter, cerebrospinal fluid, and motion parameters), normalize to MNI space with 3 mm FWHM smoothing, and apply band-pass filtering (0.01-0.1 Hz; native TR). Multiscale ROI time series are then extracted to compute Pearson correlation FCNs. EEG signals are preprocessed via ICA-based artifact removal, band-pass filtering (0.5-49 Hz), and resampling to 128 Hz. Electrode signals are projected onto cortical source space using sLORETA [18], decomposed into five canonical frequency bands (δ: 0.5-4 Hz; θ: 4-8 Hz; α: 8-13 Hz; β: 13-30 Hz; γ: 30-49 Hz), and band-limited FCNs are estimated using oPEC to mitigate volume-conduction artifacts.

4 Experiments and Results

4.1 Implementation Details

All experiments are implemented in PyTorch [19] and run on one NVIDIA A100 (40 GB) GPU using Adam [11] (batch size 32, learning rate 10^{-3}, weight decay 0.001). We use subject-level five-fold cross-validation, holding out 10% of the subjects in each training split for validation. ComBat harmonization parameters are estimated on the training split and applied to the held-out validation and test splits. For distillation in Stage 3, η and κ control feature- and logit-level distillation, and τ controls teacher-logit softness. We select them on the validation split using AUC, the selected setting is $\eta = 0.5$,

Table 2. Single-modality comparison and ablation of CMD-AD for AD vs. NC classification. Best results in each block are in bold.

A. Comparison with Single-modality Methods (identical inputs)								
Method	**fMRI-only**				**EEG-only**			
	ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE
GCN [12]	65.28±2.57	71.12±1.66	66.69±2.60	64.95±2.61	58.59±2.17	64.19±2.74	59.94±1.78	56.78±2.73
BrainNetCNN [9]	68.74±2.76	74.19±2.83	71.18±1.63	68.16±3.08	60.74±3.11	65.34±2.43	61.45±3.02	59.79±3.57
GAT [24]	68.40±2.90	73.75±2.13	69.85±3.05	68.05±2.97	60.31±2.25	64.89±2.52	62.19±2.09	57.78±2.83
BrainGNN [13]	71.95±3.25	78.19±3.14	72.56±3.71	71.81±3.20	62.67±2.42	68.75±2.91	63.69±2.48	61.31±2.80
CMD-AD (w/o distillation)	73.51±1.03	80.88±1.13	**81.99±1.46**	71.49±1.22	63.52±1.15	71.63±1.23	**68.18±1.58**	57.29±1.43
CMD-AD (w distillation)	**75.50±1.17**	**83.70±1.07**	81.09±1.00	**74.17±1.28**	**65.02±1.11**	**73.77±1.24**	65.93±1.27	**63.82±1.26**
B. Ablation Study (single-modality distill)								
Variant	**fMRI-only (Distill)**				**EEG-only (Distill)**			
	ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE
w/o multiscale	67.10±1.94	70.60±2.29	69.84±2.24	66.45±1.88	59.23±1.56	63.80±2.72	59.57±3.54	58.79±1.25
w/o multiband	70.65±2.88	75.80±2.77	75.71±3.41	69.45±2.81	59.67±3.46	64.80±2.89	59.95±3.50	59.29±3.70
w/o BCMF (self-attn)	72.99±2.95	78.80±1.95	77.49±2.00	71.91±3.21	63.96±2.19	71.20±2.41	65.19±2.17	62.32±2.90
Full CMD-AD (BCMF)	**75.50±1.17**	**83.70±1.07**	**81.09±1.00**	**74.17±1.28**	**65.02±1.11**	**73.77±1.24**	**65.93±1.27**	**63.82±1.26**

$\kappa = 0.5$, and $\tau = 2$. In Stages 2 and 3, class-balanced oversampling is applied only within the paired training split after subject-level splits are fixed; validation and test distributions remain unchanged. Test subjects remain unseen throughout all stages. We report mean±std across folds for accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC).

4.2 Results

Comparison Experiments. We compare CMD-AD with representative single-modality FCN classifiers under identical inputs (Table 2, Block A). Without distillation, CMD-AD achieves 73.51% fMRI-only ACC and 63.52% EEG-only ACC, outperforming the strongest baseline BrainGNN, which reaches 71.95% and 62.67% ACC for fMRI-only and EEG-only diagnosis, respectively. Compared with BrainGNN, CMD-AD improves ACC/AUC by 1.56/2.69% for fMRI-only diagnosis and by 0.85/2.88% for EEG-only diagnosis. These gains before distillation indicate that the hierarchical modality-specific encoder already provides stronger single-modality representations by modeling multiscale spatial network organization for fMRI and frequency-specific temporal coupling for EEG. After cross-modality knowledge distillation, the full CMD-AD further increases fMRI-only ACC/AUC to 75.50%/83.70% and EEG-only ACC/AUC to 65.02%/73.77%. Relative to CMD-AD without distillation, distillation improves ACC/AUC by 1.99/2.82% for fMRI-only diagnosis and by 1.50/2.14% for EEG-only diagnosis. It also improves specificity with modest sensitivity changes: fMRI-only SPE increases from 71.49% to 74.17%, with SEN changing from 81.99% to 81.09%, and EEG-only SPE increases from 57.29% to 63.82%, with SEN changing from 68.18% to 65.93%. These results indicate that the fused multiscale spatiotemporal representations learned from paired fMRI-EEG data transfer complementary cross-modality information to single-modality models, improving robustness when only one modality is available.

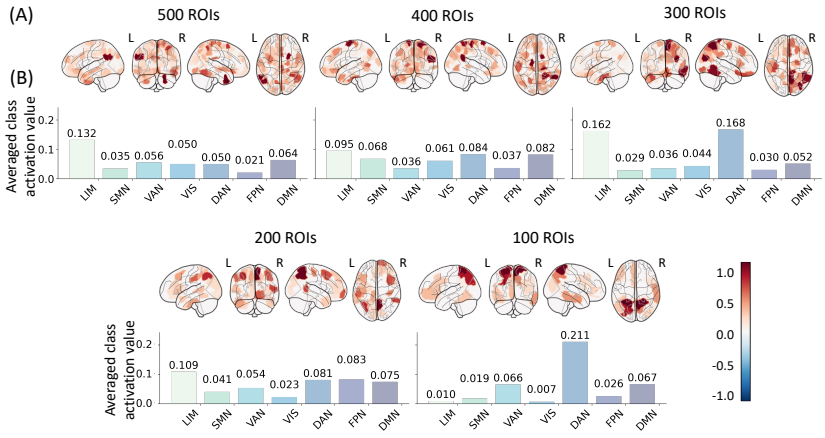

Fig. 2. Interpretability of CMD-AD. (A) Group-averaged ROI-level Grad-CAM maps. (B) Network-level saliency at each scale.

Ablation Study. We ablate CMD-AD under single-modality inference after distillation (Table 2, Block B) and summarize performance using ACC and AUC. Removing the multiscale hierarchy causes the largest degradation: fMRI-only ACC/AUC decreases from 75.50%/83.70% to 67.10%/70.60%, and EEG-only ACC/AUC decreases from 65.02%/73.77% to 59.23%/63.80%. This suggests that multiscale spatial network organization is essential *not only* for fMRI representation learning *but also* for establishing reliable cross-modality alignment across atlas scales. Collapsing EEG bands into a single channel also degrades performance, reducing fMRI-only ACC/AUC to 70.65%/75.80% and EEG-only ACC/AUC to 59.67%/64.80%. The decrease in EEG performance indicates the importance of frequency-specific temporal coupling, while the concurrent decrease in fMRI performance suggests that removing EEG band structure weakens the temporal guidance transferred to the fMRI model during distillation. Replacing BCMF with intra-modality self-attention yields a smaller but consistent decrease, with fMRI-only ACC/AUC dropping to 72.99%/78.80% and EEG-only ACC/AUC dropping to 63.96%/71.20%. This confirms that bidirectional cross-modality fusion provides useful complementary information beyond within-modality encoding alone. Overall, these results show that multiscale modeling, multiband EEG representation, and BCMF jointly support the final single-modality performance.

Interpretability. We interpret CMD-AD using Grad-CAM [22] applied to ROI representations of the fused branch before graph readout, with saliency computed with respect to the AD-NC logit. Fig. 2(A) displays group-averaged ROI saliency across Schaefer scales (100-500), and Fig. 2(B) summarizes saliency aggregated within Yeo-7 subnetworks [30]. Association systems are consistently more salient than primary sensory systems across all scales. This cross-scale consistency indicates atlas-robust evidence usage, rather than reliance on a specific parcellation.

5 Conclusion

We present CMD-AD, a three-stage framework that transfers multimodal knowledge between fMRI and EEG to enhance single-modality AD diagnosis. CMD-AD first pretrains hierarchical modality-specific encoders to learn discriminative single-modality representations, then learns fused multiscale spatiotemporal representations from paired subjects via scale-wise bidirectional cross-attention, and finally distills these fused representations and softened logits back into each modality-specific encoder. Evaluated on multicenter cohorts, CMD-AD improves fMRI-only ACC from 73.51% to 75.50% and EEG-only ACC from 63.52% to 65.02%, demonstrating that cross-modality training can effectively support practical single-modality deployment when paired fMRI-EEG data are limited.

Acknowledgments. This work was supported in part by Brain Science and Brain-like Intelligence Technology-National Science and Technology Major Project (No. 2025ZD0215400), National Natural Science Foundation of China (grant numbers 62131015, U23A20295, 82441023, 82394432), the China Ministry of Science and Technology (STI2030-Major Projects-2022ZD0209000, STI2030-Major Projects-2022ZD0213100), Key-Area Research and Development Program of Guangdong Province (2023B0303040001), and HPC Platform of ShanghaiTech University.

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

References

1. Alves, C.L., Pineda, A.M., Roster, K., Thielemann, C., Rodrigues, F.A.: Eeg functional connectivity and deep learning for automatic diagnosis of brain disorders: Alzheimer's disease and schizophrenia. Journal of Physics: complexity **3**(2), 025001 (2022)
2. Arbabyazd, L., Petkoski, S., Breakspear, M., Solodkin, A., Battaglia, D., Jirsa, V.: State-switching and high-order spatiotemporal organization of dynamic functional connectivity are disrupted by alzheimer's disease. Network Neuroscience **7**(4), 1420–1451 (2023)
3. Başar, E., Güntekin, B.: Review of delta, theta, alpha, beta, and gamma response oscillations in neuropsychiatric disorders. Supplements to Clinical neurophysiology **62**, 303–341 (2013)
4. Betzel, R.F., Bassett, D.S.: Multi-scale brain networks. Neuroimage **160**, 73–83 (2017)
5. Biswal, B., Zerrin Yetkin, F., Haughton, V.M., Hyde, J.S.: Functional connectivity in the motor cortex of resting human brain using echo-planar mri. Magnetic resonance in medicine **34**(4), 537–541 (1995)
6. Dzianok, P., Kublik, E.: Pearl-neuro database: Eeg, fmri, health and lifestyle data of middle-aged people at risk of dementia. Scientific Data **11**(1), 276 (2024)
7. Huster, R.J., Debener, S., Eichele, T., Herrmann, C.S.: Methods for simultaneous eeg-fmri: an introductory review. Journal of Neuroscience **32**(18), 6053–6060 (2012)
8. Jack Jr, C.R., Barnes, J., Bernstein, M.A., Borowski, B.J., Brewer, J., Clegg, S., Dale, A.M., Carmichael, O., Ching, C., DeCarli, C., et al.: Magnetic resonance imaging in alzheimer's disease neuroimaging initiative 2. Alzheimer's & Dementia **11**(7), 740–756 (2015)
9. Kawahara, J., Brown, C.J., Miller, S.P., Booth, B.G., Chau, V., Grunau, R.E., Zwicker, J.G., Hamarneh, G.: Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage **146**, 1038–1049 (2017)

10. Kim, S.G., Ogawa, S.: Biophysical and physiological origins of blood oxygenation level-dependent fmri signals. Journal of Cerebral Blood Flow & Metabolism **32**(7), 1188–1206 (2012)
11. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. International Conference on Learning Representations (ICLR) (2015), https://arxiv.org/abs/1412.6980, arXiv:1412.6980
12. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
13. Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Staib, L.H., Ventola, P., Duncan, J.S.: Braingnn: Interpretable brain graph neural network for fmri analysis. Medical Image Analysis **74**, 102233 (2021)
14. Liu, M., Zhang, H., Shi, F., Shen, D.: Hierarchical graph convolutional network built by multiscale atlases for brain disorder diagnosis using functional connectivity. IEEE Transactions on Neural Networks and Learning Systems **35**(11), 15182–15194 (2023)
15. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. Journal of cognitive neuroscience **19**(9), 1498–1507 (2007)
16. Miltiadous, A., Tzimourta, K.D., Afrantou, T., Ioannidis, P., Grigoriadis, N., Tsalikakis, D.G., Angelidis, P., Tsipouras, M.G., Glavas, E., Giannakeas, N., et al.: A dataset of scalp eeg recordings of alzheimer's disease, frontotemporal dementia and healthy subjects from routine eeg. Data **8**(6), 95 (2023)
17. Moosmann, M., Eichele, T., Nordby, H., Hugdahl, K., Calhoun, V.D.: Joint independent component analysis for simultaneous eeg–fmri: principle and simulation. International Journal of Psychophysiology **67**(3), 212–221 (2008)
18. Pascual-Marqui, R.D., et al.: Standardized low-resolution brain electromagnetic tomography (sloreta): technical details. Methods find exp clin pharmacol **24**(Suppl D), 5–12 (2002)
19. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems **32**, 8024–8035 (2019)
20. Prado, P., Medel, V., Gonzalez-Gomez, R., Sainz-Ballesteros, A., Vidal, V., Santamaría-García, H., Moguilner, S., Mejia, J., Slachevsky, A., Behrens, M.I., et al.: The brainlat project, a multimodal neuroimaging dataset of neurodegeneration from underrepresented backgrounds. Scientific Data **10**(1), 889 (2023)
21. Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.: Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. Cerebral cortex **28**(9), 3095–3114 (2018)
22. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 618–626 (2017)
23. Smith, K., Escudero, J.: The complex hierarchical topology of eeg functional connectivity. Journal of Neuroscience Methods **276**, 1–12 (2017)
24. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
25. Vicchietti, M.L., Ramos, F.M., Betting, L.E., Campanharo, A.S.: Computational methods of eeg signals analysis for alzheimer's disease classification. Scientific Reports **13**(1), 8184 (2023)
26. Wang, J., Wang, X., Xia, M., Liao, X., Evans, A., He, Y.: Gretna: a graph theoretical network analysis toolbox for imaging connectomics. Frontiers in human neuroscience **9**, 386 (2015)

27. Wei, X., Zhao, K., Jiao, Y., Carlisle, N.B., Xie, H., Fonzo, G.A., Zhang, Y.: Multi-modal cross-domain self-supervised pre-training for fmri and eeg fusion. Neural Networks **184**, 107066 (2025)
28. Xu, J., Chen, Y., Dong, X., Lan, M., Huang, T., Bian, Q., Cheng, J., Ke, Y.: Brainood: Out-of-distribution generalizable brain network analysis. arXiv preprint arXiv:2502.01688 (2025)
29. Yang, Y., Ye, C., Cai, G., Song, K., Zhang, J., Xiang, Y., Ma, T.: Hypercomplex graph neural network: Towards deep intersection of multi-modal brain networks. IEEE Journal of Biomedical and Health Informatics (2024)
30. Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al.: The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of neurophysiology (2011)
31. Zhang, Y., Wu, W., Toll, R.T., Naparstek, S., Maron-Katz, A., Watts, M., Gordon, J., Jeong, J., Astolfi, L., Shpigel, E., et al.: Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography. Nature Biomedical Engineering **5**(4), 309–323 (2021)
32. Zhuang, X., Yang, Z., Cordes, D.: A technical review of canonical correlation analysis for neuroscience applications. Human brain mapping **41**(13), 3807–3833 (2020)