Textile & Leather Review

ISSN 2623-6281 | www.tlr-journal.com | 10.31881/TLR

Hybrid Transformer Graph Attention Framework for Early Mild Cognitive Impairment Identification Using Multimodal Brain Networks

Pengfei Su, Wei Kong, Shuaiqun Wang

How to cite: Su P, Kong W, Wang S. Hybrid Transformer Graph Attention Framework for Early Mild Cognitive Impairment Identification Using Multimodal Brain Networks. Textile & Leather Review. 2026; 9:4034-4069. https://doi.org/10.31881/TLR.2026.4034

How to link: https://doi.org/10.31881/TLR.2026.4034

Published: 25 April 2026

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

Hybrid Transformer Graph Attention Framework for Early Mild Cognitive Impairment Identification Using Multimodal Brain Networks

Pengfei Su, Wei Kong*, Shuaiqun Wang

College of Information Engineering, Shanghai Maritime University, 1550 Haigang Ave., Shanghai 201306, P. R. China
Pengfei Su and Wei Kong contributed equally to this work, they are both first authors.
*weikong@shmtu.edu.cn

Article
https://doi.org/10.31881/TLR.2026.4034
Published 25 April 2026

ABSTRACT

Objectives: Alzheimer's disease (AD) begins with subtle symptoms, making early detection of mild cognitive impairment (MCI) crucial for timely intervention. Current deep learning models for brain imaging often process different modalities separately, failing to account for demographic variations and the need for integrated analysis of structural and functional data from the same brain regions. This approach can result in reduced sensitivity to early cognitive decline due to oversmoothing effects in Graph Convolutional Networks (GCNs).

Methods: This study proposes a hybrid multi-channel transformer graph attention network (HMT-GAT) for MCI identification. First, a structurally constrained fused brain network is constructed by incorporating DTI-derived anatomical information into rs-fMRI-based functional connectivity estimation. A locally weighted clustering coefficient (LWCC) is then used to extract multi-scale local topological features from the fused network. Demographic and acquisition-related variables, including acquisition site, age, and sex, are further integrated into a sparsely connected population graph to model inter-subject relationships.

Results: Under the same evaluation protocol, HMT-GAT achieved competitive and generally superior performance compared with implemented baseline models. For NC vs. EMCI classification, HMT-GAT obtained an ACC of 87.97 %, SEN of 80.46%, and SPE of 91.45%. For NC vs. LMCI classification, it achieved an ACC of 87.63%, SEN of 95.13%, and SPE of 94.46%, indicating balanced classification performance for MCI-related identification tasks.

Discussion:Interpretability analysis identified disease-related regions, including the inferior temporal gyrus and amygdala, suggesting that HMT-GAT may provide biologically meaningful evidence for MCI-related brain network alterations within the AD continuum.

KEYWORDS

alzheimer's disease, brain imaging patterns, graph convolutional network, LWCCl, HMT-GAT

INTRODUCTION

Alzheimer's disease (AD) and mild cognitive impairment (MCI) are neurodegenerative disorders that often remain undiagnosed until advanced stages, underscoring the need for early detection. Early identification of MCI can potentially delay or prevent progression to AD, improving patient outcomes [1,2]. Integrating neuroimaging modalities, such as resting-state functional MRI (rs-fMRI) and Diffusion Tensor Imaging (DTI), allows for capturing both functional brain activity and structural connectivity disruptions associated with cognitive decline [3,4]. The fusion of these modalities offers a comprehensive understanding of neurodegenerative diseases, crucial for accurate early diagnosis [5,6]. Deep learning models have demonstrated promise in neuroimaging-based disease classification, enabling the integration of multi-modal data and accounting for functional and structural brain changes, thus providing a reliable approach for early AD detection [7,8].

Representative studies have explored multi-scale graph convolution and local weighted clustering coefficients for MCI-related brain network analysis [9,10] MSE-GCN uses parallel, multi-scale graph convolutions to capture hierarchical connectivity patterns, and its LWCC-based fusion integrates complementary features from both modalities [11,12]. This design improved classification accuracy in MCI detection. However, MSE-GCN has notable limitations: by merging modalities into a single graph, it does not explicitly disentangle structural versus functional information, so one modality can overwhelm the other during message passing. Its deep GCN layers also risk over-smoothing, causing node representations to become too similar across the network. Moreover, MSE-GCN ignores non-imaging covariates (e.g. age, gender) when building the graph, potentially reducing robustness to demographic variability.

Other graph-based multimodal studies further introduced population graph construction, non-imaging information, and multi-channel filtering to improve disease classification performance [13,14]. Like MSE-GCN, MMP-GCN fuses DTI and fMRI into a single subject-level graph, but it introduces three key mechanisms to improve performance. It applies a DTI-strength penalty when constructing functional connectivity, thereby biasing edges to reflect underlying white-matter strength. It also constructs a multi-center attention graph over subject nodes, where edge weights are learned based on factors such as data source (site), gender, scanner model and disease status. This attention mechanism lets the model weight connections according to known cohort and demographic differences. Finally, MMP-GCN employs a multi-channel convolution scheme and a label-guided pooling strategy: different filters are applied to subgroups of features, and a pooling operator uses the known diagnostic labels to evaluate and retain high-quality node features [15,16]. These

innovations yielded state-of-the-art accuracy on AD/MCI benchmarks. Still, MMP-GCN has limitations: it continues to process a single fused graph without separate modality channels, so structural and functional signals remain entangled. Its use of label-informed pooling risks overfitting (since training labels guide the pooling) and may not generalize well to new cohorts. The overall design is also quite complex (multi-center attention, multi-channel filters), which can demand large datasets and careful tuning to avoid instability

Multi-scale and phenotype-aware graph learning has also been investigated in neurological and psychiatric imaging studies, showing the value of incorporating auxiliary information into graph construction [17,18] MAMF-GCN creates parallel graph “channels” for different atlas scales: an encoder network integrates non-imaging covariates (phenotypic attributes) to compute subject similarities, while other channels extract node features from fMRI under each atlas definition. An adaptive attention module then fuses these specific (atlas-dependent) and common embeddings, learning weights that highlight the most relevant features for diagnosis. This multi-scale, multi-modal fusion significantly improved performance on psychiatric and neurological datasets. However, MAMF-GCN’s graph construction relies on predefined atlas templates, so it cannot dynamically adapt to individual differences in connectivity. Its multi-channel architecture also adds heavy computation. Like the other models, MAMF-GCN may suffer from over-smoothing in deep GCN layers, and although attention is used, there is no explicit mechanism to keep structural and functional representations fully separate beyond the parallel channel design. prior multi-modal GCN methods have made important strides in fusing DTI and fMRI for early MCI detection, but each has gaps [19,20]. None enforces a fully disentangled treatment of modalities, deep GCNs can still over-smooth, atlas-based graphs lack flexibility, and most ignore richer demographic factors.

To address these limitations, this study proposes HMT-GAT for multimodal graph learning in early MCI identification. Unlike approaches that process imaging modalities independently or rely mainly on late fusion, HMT-GAT first constructs a structurally constrained fused brain network by incorporating DTI-derived anatomical connectivity into rs-fMRI-based functional connectivity estimation. On this fused network, LWCC is used to characterize multi-scale local weighted clustering patterns rather than serving as a standalone latent-factor fusion method. This choice is motivated by the fact that MCI-related brain alterations may appear as subtle changes in local network organization, which can be overlooked when multimodal information is projected only into a global latent space. Compared with fusion strategies such as canonical correlation analysis or joint non-negative matrix factorization, LWCC preserves region-wise topological information and

produces interpretable node-level descriptors that are suitable for subsequent population graph construction and graph attention learning [21,22]. Related sparse CCA and joint NMF studies have also shown that latent-space fusion can be useful for multimodal biomedical data integration [23]. In addition, HMT-GAT incorporates demographic and acquisition-related information into the population graph and applies a multi-channel graph attention architecture to improve feature filtering. Overall, the proposed framework aims to capture complementary structural and functional information while maintaining graph topology and regional interpretability.

MATERIALS AND METHODS

Workflow of this study

A detailed framework description of the proposed HMT-GAT method is depicted in Figure 1.

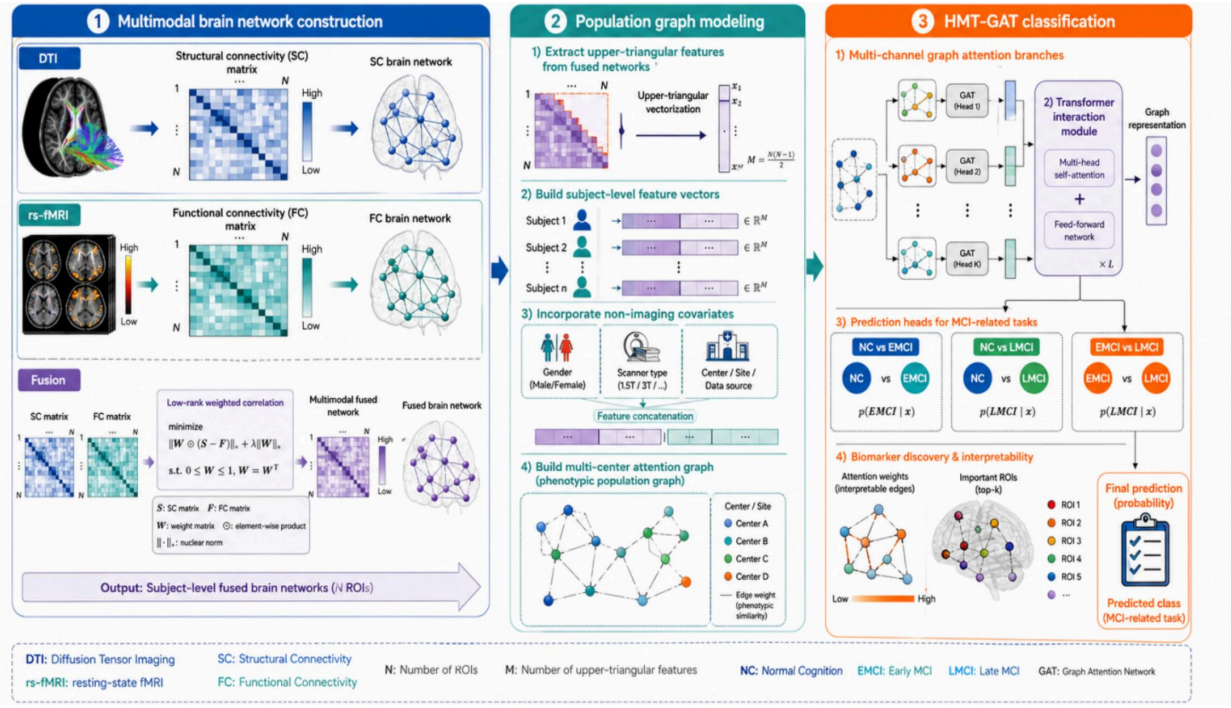

Figure 1. Overall workflow of the proposed HMT-GAT framework for MCI classification using multimodal brain networks

This study presents a structured workflow for early mild cognitive impairment (MCI) identification within the Alzheimer's disease (AD) continuum. As shown in Fig. 1, the proposed HMT-GAT framework consists of three main stages: multimodal brain network construction, population graph modeling with non-imaging information, and graph-based classification using the multi-channel attention architecture.

In the second stage, the upper triangular elements of the fused network are extracted and reorganized into subject-specific descriptors. These descriptors are then integrated into a cross-center population graph, where each node represents an individual subject and each edge is defined according to similarities in both brain connectivity patterns and phenotypic information, such as demographic characteristics and scanner-related factors. This design enables the model to capture inter-subject relationships while reducing the impact of site-related heterogeneity in multi-center data.

The third stage processes the graph using a multi-channel Graph Attention Network (GAT) equipped with dynamic multi-head self-attention. This mechanism learns refined connectivity features within each modality subgraph and incorporates non-imaging phenotypic data adaptively, mitigating biases related to cohort diversity and acquisition settings. The resulting unified embeddings—combining imaging and phenotypic information—are classified to discriminate between patients with mild cognitive impairment (MCI) and healthy controls (HC).

The logical transition between stages is essential to the model's success. The initial fusion of multimodal data enables a seamless shift to graph-based representation in the second stage, effectively capturing complex feature relationships. This graph encoding facilitates the subsequent integration of phenotypic data in the final stage, where a GAT with self-attention refines embeddings by correcting for cohort heterogeneity, significantly enhancing prediction accuracy. The combined use of imaging and phenotypic features ensures a robust and interpretable model, even under small-sample conditions. This integrated approach produces a unified embedding that incorporates both brain connectivity and phenotypic context, which is then classified for early MCI identification detection. The entire architecture is trained end-to-end with cross-entropy loss, emphasizing generalization ability, and progresses coherently from fusion to graph modeling to phenotype adjustment, ensuring scientific rigor and translational relevance.

Brain Network Construction

Fused Brain Network Construction

The first step of the proposed framework is to construct a fused brain connectivity network that jointly reflects functional interactions and structural constraints. A fusion strategy is implemented whereby DTI-derived structural connectivity constrains the estimation of functional connectivity obtained from resting-state fMRI signals. Rather than treating functional and structural modalities as independent or sequential inputs, this approach embeds anatomical priors directly into the functional network construction process. The resulting

connectivity matrix—termed the fused brain network—captures not only temporally correlated activity patterns but also respects the physical wiring constraints of the brain, thereby yielding a more coherent and neurobiologically plausible representation for downstream analysis.

In conventional sparse representation (SR)-based methods, the functional connectivity matrix W is estimated under l1-norm or l2-norm regularization to enforce sparsity. In order to incorporate structural priors derived from DTI, we further embed a DTI-related penalty term into the optimization framework and simultaneously introduce a structural-consistency regularizer. Accordingly, the objective function can be written as:

$$\min_{W} \frac{1}{2} \| Y - XW \|_F^2 + \lambda \| C \odot W \|_1 + \mu \| W - S \|_F^2 \quad (1)$$

Y is the target matrix, X is the input data matrix (e.g., the functional connectivity matrix obtained via Pearson correlation), and W is the weight matrix to be optimized. This term represents the prediction error, aiming to make X transformed by W as close as possible to Y. Allowing different input and output matrices provides greater flexibility. λ is the regularization parameter, C is a weighting matrix, and $\odot$ denotes element-wise multiplication. This sparse regularization enhances the model's generalization ability and reduces overfitting. S (the DTI-derived reference matrix) is obtained by using fractional anisotropy (FA) as feature vectors and partitioning the brain space into 90 ROIs via the AAL template on DTI images-regional structural connectivity was represented by the mean fractional anisotropy of the fiber bundles linking paired regions. The resulting subject-specific structural network was then organized as a 90 × 90 connectivity matrix. This term ensures W remains close to S , improving model stability and potential generalization. This formulation preserves structural prior information and reduces redundant functional connections, thereby improving the stability of the fused network.

Rationale for using LWCC-based topological feature extraction

After constructing the fused brain connectivity network, it is necessary to transform the high-dimensional connectivity matrix into informative and compact graph descriptors. In this study, LWCC was selected because it characterizes local weighted clustering patterns around each brain region while preserving the topological organization of the fused network. This property is important for MCI identification, since early pathological changes may appear as subtle alterations in local connectivity organization rather than as isolated connection-level changes.

Alternative multimodal fusion methods, such as canonical correlation analysis (CCA) and joint non-negative matrix factorization (joint NMF), are useful for learning shared latent representations across modalities. However, these methods mainly project multimodal data into a latent feature space and may reduce the explicit region-to-region topological structure of brain networks. In contrast, LWCC operates directly on the weighted brain connectivity graph and produces node-level topological descriptors that retain local clustering information, edge-weight information, and regional interpretability. Therefore, LWCC is more consistent with the graph-based design of HMT-GAT, where each subject is represented by topology-aware features derived from the fused brain network.

In addition, LWCC has relatively low computational complexity and does not introduce a large number of trainable parameters. This is beneficial under the small-sample setting of the present study, especially considering the limited number of LMCI subjects. For these reasons, LWCC was adopted as the main topological feature extraction strategy after fMRI–DTI network fusion.

In Equation (1), the structural connectivity penalty matrix C plays a pivotal role in enhancing model performance. To ensure robustness, both the structural connectivity patterns and the inter-group variability in connectivity strength across the population are taken into account. Specifically, given a training set comprising T subjects with known diagnostic labels, the cohort is stratified into two groups according to their clinical classification. Additionally, heterogeneity arising from multi-center data acquisition is incorporated to better reflect real-world variability and improve generalizability across sites. The training subjects were divided into two diagnostic groups, and the corresponding DTI-based connectivity matrices were organized separately for subsequent analysis,SC+=[SC1+, SC2+, …, SCT1+] and SC−=[SC1−, SC2−, …, SCT2−], where T1 and T2 are the numbers of subjects in each group (T1+T2= T). A group-difference matrix SC# was then constructed to quantify disparities in structural connectivity strength between the two diagnostic groups:

$$\mathbf{SC}^{\#} = \left| \frac{\frac{1}{T_1}\sum_{i=1}^{T_1} \mathbf{Sc}^{i+} - \frac{1}{T_2}\sum_{j=1}^{T_2} \mathbf{sc}^{j-}}{\frac{1}{T_1}\sum_{i=1}^{T_1} \mathbf{sc}^{i+} + \frac{1}{T_2}\sum_{j=1}^{T_2} \mathbf{sc}^{j-}} \right| \tag{2}$$

Here, SC#∈R90×90matrix characterizes region-wise differences in DTI-derived connectivity between the two groups, and SC∈R90×90denotes the DTI connectivity network. $SCi+$ and $SCj-$ are the DTI connectivity strength matrices for subjects i and j from different groups.

For each subject, the structural connectivity penalty matrix C , with elements Cij , is defined based on its structural connectivity matrix SC and strength diversity matrix $SC#$:

$$\mathbf{C}_{ij} = \exp\left(\frac{-\mathbf{sc}_{ij}^2}{\sigma_1}\right) \times \left(1 + \exp\left(\frac{-\mathbf{SC}_{ij}^{\#}}{\sigma_2}\right)\right) \tag{3}$$

where $\sigma 1$ and $\sigma 2$ are the average standard deviations of SC and $SC\#$ across all subjects.

LWCC Feature Extraction

Based on the above rationale, a multi-scale LWCC feature extraction strategy was applied to the fused fMRI–DTI brain connectivity network. Instead of treating the fused matrix as a simple vector of independent edges, LWCC summarizes the local weighted clustering structure around each brain region. LWCC values were computed at multiple neighborhood scales and then concatenated to form topology-aware subject-level features for classification. In brain network models, local connectivity patterns at different scales can reveal distinct functional and structural properties; hence, LWCC is evaluated for various neighbor orders (e.g., 1-hop, 2-hop, 3-hop) and the resulting multi-scale coefficients are concatenated. Formally, for each node i, the multi-scale LWCC is defined as:

$$f_i^{(k)} = \frac{2\sum_{j:j\in\varepsilon_i^{(k)}}(w_{ij})^{\alpha}}{\mid \varepsilon_i^{(k)} \mid (\mid \varepsilon_i^{(k)} - 1) \mid} \tag{4}$$

$f_i^{(k)}$: LWCC value of node i at k -hop neighbors. $\varepsilon_i^{(k)}$: Set of k -hop neighbors for node i . w_{ij} : Connection weight between nodes i and j : Exponent controlling weighted clustering (e.g., α=1/3), adjusting the influence of weights.

Integration of Non-Imaging Data into Model Construction

To further enhance the biological relevance and population-level robustness of the constructed brain graph, demographic and phenotypic information is explicitly integrated into the modeling process. Rather than treating non-imaging covariates (e.g., gender, site, scanner type) as isolated metadata, a dynamic multi-head self-attention mechanism is introduced to transform these discrete attributes into meaningful phenotypic similarities between nodes. These learned similarities are subsequently fused with image-derived affinities to

construct a more expressive and context-aware adjacency matrix. The structure of the proposed non-imaging information encoding module is illustrated in Fig. 2.

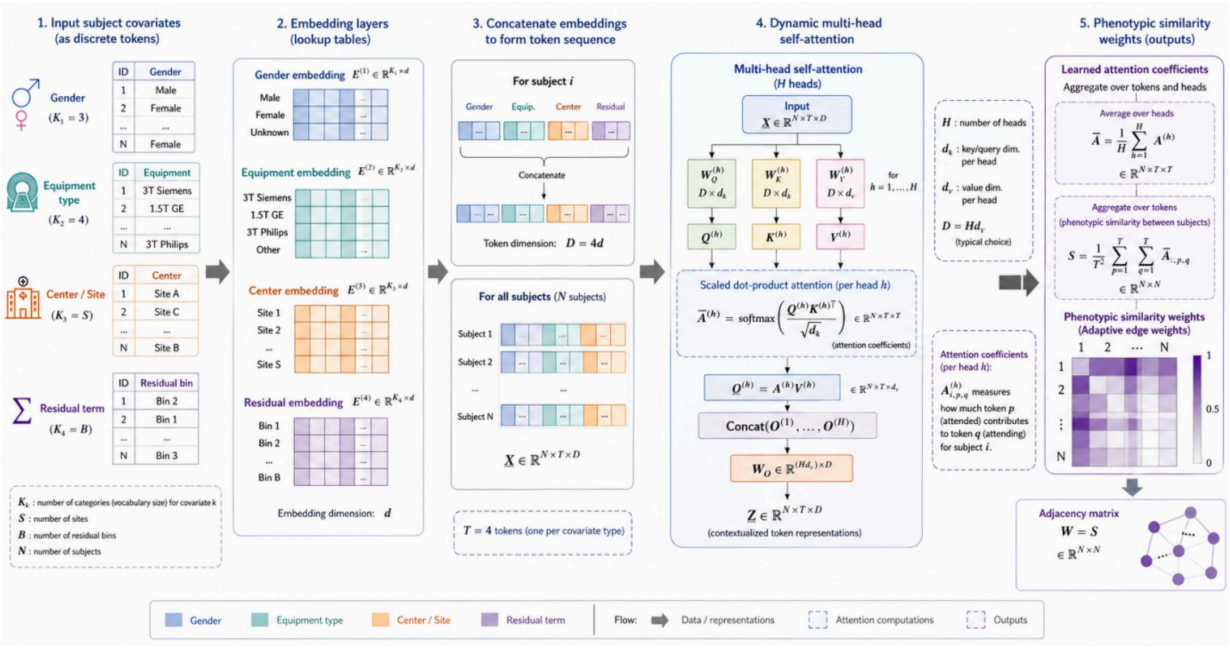

Figure 2. Dynamic multi-head self-attention module for encoding non-imaging covariates and generating phenotypic similarity weights

Learning of non-imaging attention coefficients

The non-imaging attributes used as input include gender, scanner type, acquisition equipment, and imaging center/site. These categorical variables were encoded as discrete indices and then mapped into low-dimensional embedding vectors through trainable embedding layers. Diagnostic labels were used only to define supervised classification tasks and to compute the classification loss during training; they were not used as input attributes for test samples during model evaluation.

For a subject $pair(i, j)$, attribute-level similarity scores were computed from the corresponding embeddings. Specifically, s_ij^s, s_ij^g, s_ij^e , and s_ij^c denote the pairwise similarity scores derived from scanner type, gender, acquisition equipment, and imaging center/site, respectively. These scores describe whether two subjects share similar non-imaging characteristics and provide phenotypic information for population graph construction. The five attention coefficients a_s, a_g, a_e, a_c , and a_0 are not manually assigned. Instead, they are learnable parameters optimized jointly with the classification objective during model training. Before

normalization, a trainable coefficient vector $\beta = [\beta_s, \beta_g, \beta_e, \beta_c, \beta_0]$ is introduced. The normalized attention coefficients are obtained through a softmax function:

$$[a_s, a_g, a_e, a_c, a_0] = s([\beta_s, \beta_g, \beta_e, \beta_c, \beta_0]) \tag{5}$$

Here, a_s, a_g, a_e , and a_c represent the learned contributions of scanner type, gender, acquisition equipment, and imaging center/site to the phenotypic similarity graph. The residual base term a_0 represents a task-adaptive baseline weight. It is introduced to preserve a minimum non-imaging contribution when two subjects do not share strong similarity in the observed categorical attributes or when some attribute information is weakly informative. Therefore, a_0 functions as an intercept-like residual component rather than as an additional clinical variable. The phenotypic similarity between subjects i and j is computed as:

$$S_{pheno}(i,j) = a_s s_{ij}^s + a_g s_{ij}^g + a_e s_{ij}^e + a_c s_{ij}^c + a_0 \tag{6}$$

The learned phenotypic similarity matrix is incorporated into the final adjacency matrix through element-wise modulation:

$$A_{final} = A_{img} \odot S_{pheno} \tag{7}$$

where A_{img} denotes the image-derived adjacency matrix obtained from the fused brain network, S_{pheno} denotes the learned phenotypic similarity matrix, and $\odot$ denotes the Hadamard product. In this way, non-imaging covariates do not replace the imaging-based graph structure; instead, they adaptively adjust the strength of inter-subject edges according to demographic and acquisition-related similarities. It should be noted that the attention coefficients are learned separately for each classification task under the corresponding training folds. Therefore, the values reported in Table 1 are the average learned coefficients obtained after model training, rather than fixed prior weights. The variation of a_s, a_g, a_e, a_c , and a_0 across NC vs. EMCI, NC vs. LMCI, and EMCI vs. LMCI tasks indicates that the proposed mechanism does not treat all non-imaging attributes as equally important. Instead, it learns task-specific attribute importance from the training data.

Table 1. Learned attention weights across classification tasks

Tasks/Atten coef	as	ag	ae	ac	a0
NC vs EMCI	0.203	0.191	0.213	0.208	0.179
NC vs LMCI	0.209	0.192	0.206	0.205	0.183
EMCI vs LMCI	0.196	0.189	0.212	0.214	0.184

The embedding layer maps discrete categorical features into continuous low-dimensional vectors to capture semantic similarity. It takes discrete integer indices as input (e.g., gender=0, equipment type=2) and outputs continuous vectors (e.g., gender mapped to an 8-dimensional vector, equipment type mapped to an 8-dimensional vector). The formula is as follows:

$$e_{\text{gender}} = Embedding(g_i), \quad e_{\text{center}} = Embedding(c_i), \quad etc \tag{8}$$

All embedded vectors of non-imaging information are concatenated into a joint vector, which serves as input for the subsequent self-attention mechanism:

$$e_{combined} = [e_{gender}; e_{equipment}; e_{center}; e_{status}] \tag{9}$$

The input data is transformed into continuous vectors through the embedding layer, then linearly projected into Q (Query), K (Key), and V (Value) matrices:

$$Q = EW_Q, K = EW_K, V = EW_V \tag{10}$$

where W_Q , W_K , and W_V are learnable weight matrices, and d_k denotes the dimensionality of each attention head. For each head h , the attention score is calculated as:

$$\text{Attention}_h = \text{softmax}\left(\frac{Q_h K_h^T}{\sqrt{d_k}}\right) V_h \tag{11}$$

$Q_h K_h^T$ The similarity between nodes is computed (dot product), d_k with a scaling factor to prevent gradient vanishing/explosion, then normalized into probability distributions representing the attention weights between nodes. The outputs from all heads are concatenated and linearly transformed to obtain the final result:

$$\text{MultiHead}\ (Q,K,V)=\text{Concat}\ (\text{Attention}_1,\text{Attention}_2,...,\text{Attention}_H)W_O \tag{12}$$

here W_O ∈ RH is the output weight matrix. The phenotypic similarity matrix Spheno is generated from the multi-head self-attention output to produce a dynamic weight matrix:

$$S_{\text{pheno}}=\text{softmax}\ (\text{MultiHead}\ (Q,K,V)) \tag{13}$$

The fused multi-modal adjacency matrix is combined with the non-imaging adjacency matrix through element-wise multiplication:

$$A=S_{\text{feat}}\odot S_{\text{pheno}} \tag{14}$$

where $\odot$ denotes element-wise multiplication (Hadamard product).

Multi-channel Pooling GAT

Once the enriched adjacency matrix—infused with functional, structural, and phenotypic affinities—is obtained, the framework transitions to a refined graph-based learning phase. In this stage, a multi-channel Graph Attention Network (GAT) is employed to extract nuanced connectivity patterns that differentiate Mild Cognitive Impairment (MCI) from normal cognition. This approach enhances the traditional GCN by integrating a learnable attention mechanism, enabling the model to automatically assign varying importance to different neighbors rather than treating all adjacent nodes uniformly.

Formally, let each node i in the graph possess input features xi and neighbor set $\mathcal{N}(i)$. Through a shared linear transformation W∈RF′×F, intermediate features Wx_i are computed. Attention scores between j and i are calculated as:

$$e_{ij}=LeakyReLU\big(\mathbf{a}^{\top}\big[\mathbf{W}\mathbf{x}_i\parallel\mathbf{W}\mathbf{x}_j\big]\big) \tag{15}$$

These coefficients are normalized via a softmax over i 's neighbors:

$$\alpha_{ij}=\frac{\exp(e_{ij})}{\sum_{k\in N(i)}\exp(e_{ik})} \tag{16}$$

This enables each neighborhood's contribution to be adaptively weighted based on learned relevance. The node's updated representation is:

$$\mathbf{h}_i' = \sigma\left(\sum_{j\in N(i)} \alpha_{ij}\mathbf{W}\mathbf{x}_j\right) \tag{17}$$

To enhance expressivity, the model employs multi-head attention:

$$\mathbf{h}_i^{multi} = \Big\|_{k=1}^{K} \sigma\left(\sum_{j\in N(i)} \alpha_{ij}^k \mathbf{W}^k \mathbf{x}_j\right) \tag{18}$$

where outputs from K parallel attention heads are concatenated, or averaged in the final layer. Each modal-specific channel—for functional, structural, or phenotype-enhanced graphs—utilizes an independent GAT branch. Outputs are then pooled via attention-guided selection (akin to Self-Attention Graph Pooling) to emphasize diagnostically significant sub-networks. The pooled representations from all channels are fused, delivering a hierarchically enriched graph embedding for downstream MCI classification. This multi-channel GAT design takes advantage of learnable attention to both distinguish informative connections and integrate cross-modal signals, resulting in embeddings that are both biologically meaningful and diagnostically powerful.

To further enhance the specificity of the graph filtering process and improve the model's ability to discriminate between different disease states, a specialized graph filtering mechanism is introduced. This mechanism employs a multi-channel approach, where distinct graph filters are assigned to feature subsets based on their statistical characteristics, and a pooling mechanism is utilized to prune edges according to the disease status of the training samples.

In this study, the multi-channel mechanism operates on disjoint feature subsets. Specifically, after feature ranking, the full feature matrix $X \in \mathbb{R}^{N\times M}$ is divided into W non-overlapping subsets:

$$\frac{\parallel Mean(\mathbf{X}^+) - Mean(\mathbf{X}^-) \parallel}{Std(\mathbf{X}^+) + Std(\mathbf{X}^-)} \tag{19}$$

where M_w denotes the number of features assigned to the w -th channel. For a given w the first w −1 channels contain M / W features, and the remaining features are assigned to the last channel. Therefore, each feature is used in one and only one channel. The graph filters in different channels are independent, but they operate on non-overlapping feature dimensions. This design avoids duplicating the full feature space across all channels.

Each feature subset is then associated with a channel-specific graph filter and processed by an independent lightweight graph branch. Importantly, each branch receives only its assigned feature subset rather than the full feature vector. Therefore, although the filters are channel-specific, their input dimensions are reduced according to the feature partition, which keeps the total number of channel-specific transformation parameters controlled. The outputs of all channels are concatenated and passed to the subsequent classification module. This design allows statistically different feature groups to be filtered separately while preserving the overall feature dimension after concatenation.

Inspired by previous feature selection and statistical filtering strategies, where feature mean and standard deviation are commonly used to characterize discriminative differences, the ranking criterion for splitting features is defined as:

$$X = [X_1, X_2, ..., X_W], \quad X_p \cap X_q = \varnothing (p \neq q), \quad \sum_{w=1}^{W} M_w = M \tag{20}$$

In Eq. (20), assume that the graph contains N subjects, with T samples used for training and M features for each subject. Let $T1$ and $T2$ denote the numbers of training samples in the two diagnostic groups, where $T1 + T2 = T$. $X^+ \in R^{T1 \times M}$ and $X^- \in R^{T2 \times M}$ represent the feature matrices of the two groups. Mean (·) and Std (·) denote the mean and standard deviation functions, respectively. Features with larger ranking scores are considered more discriminative and are assigned earlier in the ranked feature list.

Selection of channel number W

The channel number W was selected through a validation-based hyperparameter search rather than being learned during backpropagation. In this study, candidate values were set to $W \in 1, 2, 3, 4, 5, 6$. For each candidate value, feature partitioning, graph filtering, and classification were performed under the same stratified 10-fold cross-validation protocol. The validation performance was compared using the mean ACC across the evaluated MCI-related classification tasks. The final value of W was selected according to the best mean

validation performance while avoiding unnecessary feature fragmentation. In the present experiments, W =5 was adopted as the default setting for the final model, while W =4 was also reported because it achieved the best performance in the NC vs. EMCI task. The test fold was not used for selecting W .

Parameter analysis of the multi-channel design

The parameter-efficiency claim of the proposed multi-channel mechanism holds under the disjoint feature partitioning setting described above. In a conventional single-channel graph filtering layer, if each subject has M input features and the hidden dimension is Q_1 , the feature transformation contains $M \times Q_1$ trainable parameters.

In the proposed multi-channel setting, the input features are divided into W non-overlapping subsets. The www-th channel contains M_w input features and uses an independent transformation matrix $W_w \in \mathbb{R}^{M_w \times Q_1}$. Therefore, the total number of parameters across all channel-specific filters is:

$$\sum_{w=1}^{W} M_w Q_1 = \left(\sum_{w=1}^{W} M_w \right) Q_1 = M Q_1 \tag{21}$$

After feature ranking according to Eq. (20), the full feature matrix $X \in \mathrm{R}^{\mathrm{N} \times \mathrm{M}}$ is divided into W disjoint subsets, denoted as $X = [X_1, X_2, ..., X_W]$. The ranked features are assigned to different channels in descending order of their discriminative scores. For a given W , the feature allocation follows the deterministic rule described above: the first W − 1 channels contain *M*/*W* features, and the last channel contains the remaining features. Based on these feature subsets and the non-imaging covariates, channel-specific adjacency matrices $A_1, A_2, ..., A_W$ are constructed in parallel. To preserve the global population graph structure, each channel-specific adjacency matrix is refined by the initial adjacency matrix A_0 through element-wise multiplication:

$$\hat{A}_w = A_0 \odot A_w, \quad w = 1, 2, ..., W, \tag{22}$$

Where $\odot$ denotes the Hadamard product. The resulting channel-dependent subgraphs are then processed by their corresponding graph branches for feature filtering.

For graph pooling, let the population graph contain N subjects, including T_1 labeled training samples from Class 1 and T_2 labeled training samples from Class 2. For each node, two class-specific similarity matrices are computed: $S_1 \in \mathbb{R}^{N \times T_1}$, which measures its affinity to labeled samples from Class 1, and $S_2 \in \mathbb{R}^{N \times T_2}$,

which measures its affinity to labeled samples from Class 2. A node-level discriminative score vector is then obtained from the discrepancy between S_1 and S_2 :

$$\mathbf{d} = \| \ Mean(\mathbf{S}_1) - Mean(\mathbf{S}_2) \ \| \tag{23}$$

Nodes are ranked according to their discriminative scores, and the top-k nodes are retained. The selected node subset induces a pooled graph ($\bar{A}, \bar{X}$), where $\bar{A}$ and $\bar{X}$ denote the pooled adjacency matrix and pooled feature matrix, respectively. The overall pooling procedure is shown in Fig. 3.

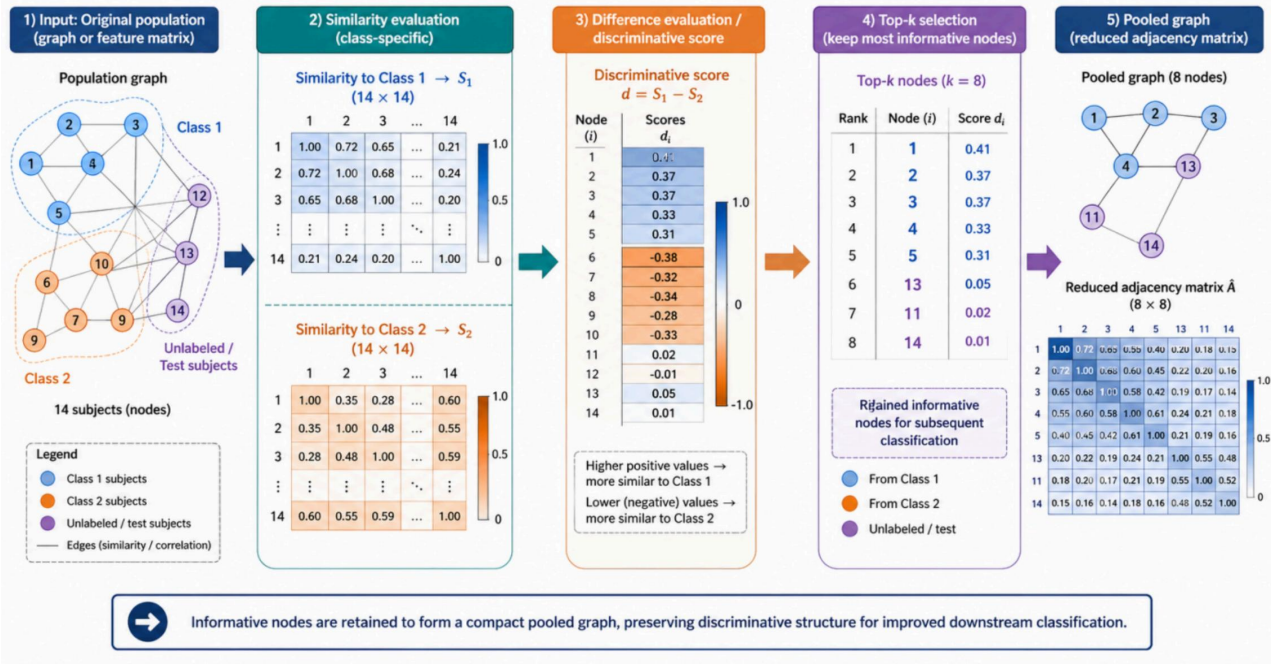

Figure 3. Overview of the proposed pooling mechanism

EXPERIMENTS

Data Sources and preprocessing

Data from a total of 133 participants were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. All participants gave their written informed consent before the data for this paper were collected, and the research was approved by the ethics committee of Shanghai Maritime University. Each subject was assessed using both Diffusion Tensor Imaging (DTI) and resting-state functional MRI (r-fMRI), providing comprehensive multi-modal neuroimaging data. In addition to imaging data, demographic and acquisition-related information, including age, gender, scanner type, and imaging center, was collected to

support phenotype-informed graph modeling. Diagnostic labels, including EMCI, LMCI, and NC, were used only for supervised training and performance evaluation, and were not used as input features for test samples during model evaluation. See Table 2.

Table 2. The demographics of the participants

Groups	EMCI	LMCI	NC
Number	52	21	60
Gender(M/F)	20/32	10/11	30/30
Age(mean ± std)	74.8±6.1	74.5±6.9	75.4±6.5

Functional connectivity (FC) in rfMRI is commonly derived by quantifying the temporal synchronization of BOLD (blood oxygen level-dependent) signal fluctuations across anatomically or functionally defined brain regions [24,25]. The standard analytic pipeline begins with rigorous preprocessing steps—comprising slice timing correction, realignment for head motion, spatial normalization to a common template, spatial smoothing, and temporal band-pass filtering—to mitigate physiological noise and scanner-related artifacts. Subsequently, BOLD time series are extracted from regions of interest (ROIs), identified either through atlas-based parcellation (e.g., AAL) or data-driven approaches such as independent component analysis (ICA). Functional connectivity is then estimated, most frequently via Pearson correlation, yielding a symmetric connectivity matrix that reflects the strength of inter-regional coupling. Alternative statistical metrics, including mutual information, partial correlation, or time-resolved dynamic connectivity, may also be employed to capture more nuanced patterns of neural co-activation. These connectivity profiles are typically visualized as matrices or network graphs, enabling the characterization of intrinsic brain network organization at rest. To enable integrative analysis across imaging modalities, such FC networks provide a biologically grounded scaffold, serving as a fundamental representation of the brain's intrinsic functional architecture for subsequent multi-modal data fusion.

To integrate DTI information into the functional connectivity network derived from rs-fMRI, structural connectivity was extracted by analyzing white matter fiber tracts. The DTI preprocessing procedure included head-motion correction, eddy-current correction, tensor fitting, and estimation of fractional anisotropy (FA) and principal diffusion directions. Fiber tracking was then performed to reconstruct white matter pathways. Brain regions were defined using the AAL atlas, and inter-regional structural connectivity was quantified using

tract-related measures such as fiber count, tract length, or mean FA. The resulting structural connectivity matrix provided anatomical constraints for multimodal brain network fusion.

Validation strategy and overfitting control under limited LMCI samples

Considering the relatively limited number of LMCI samples in the present cohort, special attention was paid to validation strategy and overfitting control. Although the overall dataset included 133 subjects, the LMCI group contained only 21 participants, which may increase the risk of unstable estimation and reduced generalization ability. To address this issue, a stratified 10-fold cross-validation strategy was adopted. In each fold, samples from NC, EMCI, and LMCI groups were distributed as evenly as possible between the training and testing subsets, so that the minority LMCI class could be represented in both model training and evaluation. The data split was performed at the subject level, and all feature extraction, graph construction, and model training procedures for each fold were conducted using only the corresponding training subset to avoid information leakage.

In addition, several measures were used to reduce overfitting. First, the fused brain network was constructed with sparsity regularization, which constrains redundant connections and reduces the risk of fitting noise in small-sample settings. Second, dropout regularization was applied during model training, and the model was trained with a compact architecture rather than an excessively deep graph network, thereby alleviating over-smoothing and limiting the number of trainable parameters. Third, the multi-channel design partitions features into several subsets without increasing the total parameter scale under the same hidden dimension setting, which helps improve feature filtering while maintaining parameter efficiency. Finally, model performance was evaluated using multiple metrics, including ACC, SEN, SPE, and AUC, instead of relying only on accuracy [26]. This is important for imbalanced MCI classification tasks because sensitivity and specificity provide complementary information about the model's ability to identify clinically relevant minority samples.

Influence of the Regularization Parameter

The construction of the fused brain connectivity network involves a single regularization coefficient, λ, which controls the sparsity level of the estimated network. The influence of λ on classification performance was evaluated by varying its value from 2^{-10} to 2^{-0}. The results show that classification accuracy changes with different values of λ, and relatively stable performance is generally observed when λ falls within the range from 2^{-6} to 2^{-4}.

To clarify the influence of non-imaging covariates on population graph construction, Table 1 reports the learned attention coefficients for each classification task. These coefficients—denoted as as, ag, ae, ac, and a0—correspond respectively to the learned weights for scanner type, gender, acquisition equipment, data center (site), and a residual base term, obtained after training via the multi-head self-attention mechanism. Each coefficient quantifies the relative contribution of its associated attribute to the edge-weighting process in the phenotypic similarity graph.

As shown in Table 1, the learned coefficients are not identical across classification tasks, indicating that the model assigns different importance to non-imaging attributes according to the diagnostic comparison being performed. The coefficients associated with acquisition equipment and imaging center/site are relatively high in all tasks, suggesting that scanner-related and site-related heterogeneity plays an important role in population graph construction. In contrast, the coefficient for gender is lower and more stable, indicating a relatively weaker contribution in this cohort. The residual base term a_0 remains non-zero across tasks, which helps maintain a baseline phenotypic contribution when the observed categorical attributes provide limited discriminative information.

Figure 4 visualizes the task-specific learned attention coefficients reported in Table 1. The figure shows that the attention mechanism assigns different weights to non-imaging attributes across classification tasks, rather than treating all attributes as equally important.

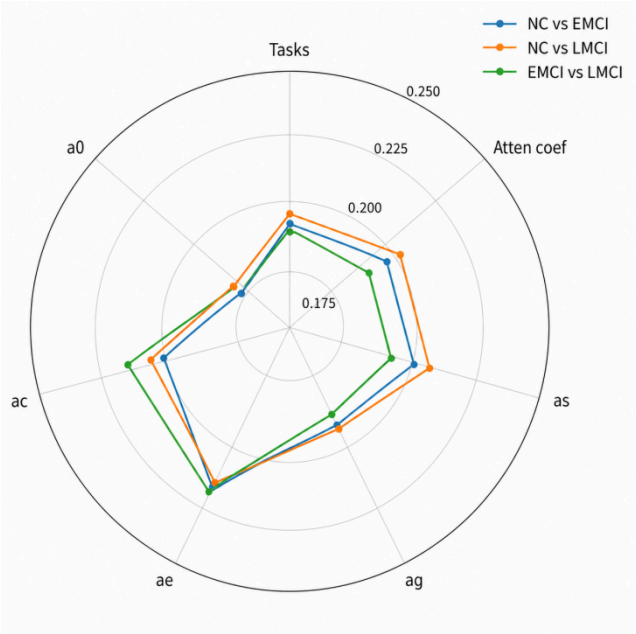

Figure 4. Visualization of learned attention weights for non-imaging attributes across different classification tasks

To evaluate the influence of the regularization parameter λ on classification performance, Figure 5 presents a comparative analysis of accuracy across different λ values.

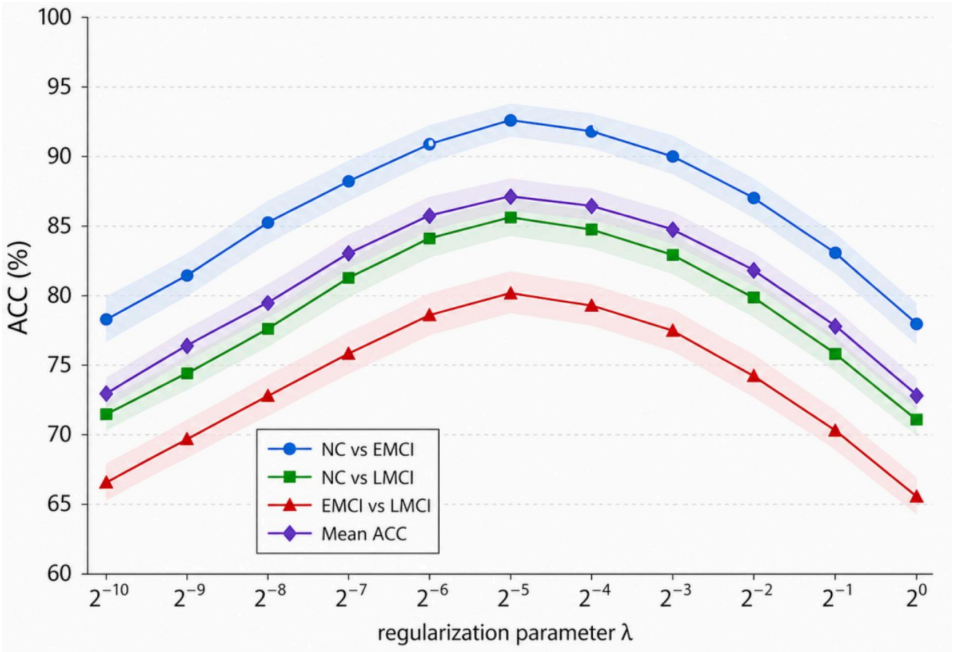

Figure 5. Influence of λ on classification accuracy

Overall, the mean ACC attains its peak when λ is set to 2^{-5}. Varying λ leads to an average ACC fluctuation of 6.2%, suggesting that the regularization setting has a noticeable impact on fused brain network construction and the downstream classification results. Meanwhile, the proposed construction strategy remains relatively concise in terms of formulation and implementation.

Comparison of different fusion and feature extraction strategies

To further justify the use of LWCC-based topological feature extraction, we compared several representative fusion and feature extraction strategies under the same downstream HMT-GAT classifier and the same stratified 10-fold cross-validation protocol. The compared strategies included direct fMRI–DTI feature concatenation, CCA-based latent fusion, joint NMF-based latent fusion, the SC-guided fused network without LWCC, and the proposed SC-guided fused network with LWCC. The purpose of this comparison was to examine whether preserving local weighted graph topology provides additional benefit beyond latent-space fusion or direct feature concatenation.

As shown in Table 3, direct fMRI–DTI feature concatenation produced the lowest mean ACC, indicating that simple feature-level integration may not sufficiently capture the relationship between structural and

functional brain connectivity. CCA-based and joint NMF-based fusion achieved higher performance than direct concatenation, suggesting that latent-space fusion can improve multimodal representation to some extent. However, these strategies still performed lower than the SC-guided fused network, probably because they do not explicitly preserve region-wise graph topology. When LWCC was further applied to the SC-guided fused network, the mean ACC increased to 87.80%, with ACC values of 87.97% and 87.63% for NC vs. EMCI and NC vs. LMCI, respectively. These results suggest that LWCC-based topological feature extraction provides complementary local clustering information and is consistent with the graph-based design of HMT-GAT.

Table 3. Comparison of different fusion and feature extraction strategies under the same HMT-GAT classifier

Strategy	Graph topology	NC vs EMCI ACC(%)	NC vs LMCI ACC(%)	Mean ACC (%)
Feature concatenation	Not explicitly preserved	83.42	84.76	84.09
CCA-based fusion	Partly preserved	84.58	85.21	84.90
Joint NMF-based fusion	Partly preserved	85.16	85.83	85.50
SC-guided fusion without LWCC	Preserved	86.31	86.52	86.42
SC-guided fusion + LWCC	Preserved	87.97	87.63	87.80

The classification performance, measured by Accuracy (ACC), Sensitivity (SEN), and Specificity (SPE), is compared across various multi-modal configurations (fMRI, DTI, and fMRI+DTI) and analytical methods (GCN, SVM, MLP, GAT). The results show that models using both fMRI and DTI generally achieve better classification performance than single-modality settings. This finding suggests that integrating structural and functional brain information can improve the discriminative capacity of the model. Among the evaluated methods, HMT-GAT achieves the highest ACC in both NC vs. EMCI and NC vs. LMCI tasks, while maintaining a favorable balance between sensitivity and specificity. Notably, the proposed HMT-GAT model demonstrates significant enhancement in sensitivity, a critical indicator for early-stage detection of Mild Cognitive Impairment (MCI), thus emphasizing its potential clinical utility.

To comprehensively assess the effectiveness of different models and modality combinations, Figure 6 presents a heatmap comparison of classification performance (ACC, SEN, SPE) across multiple methods.

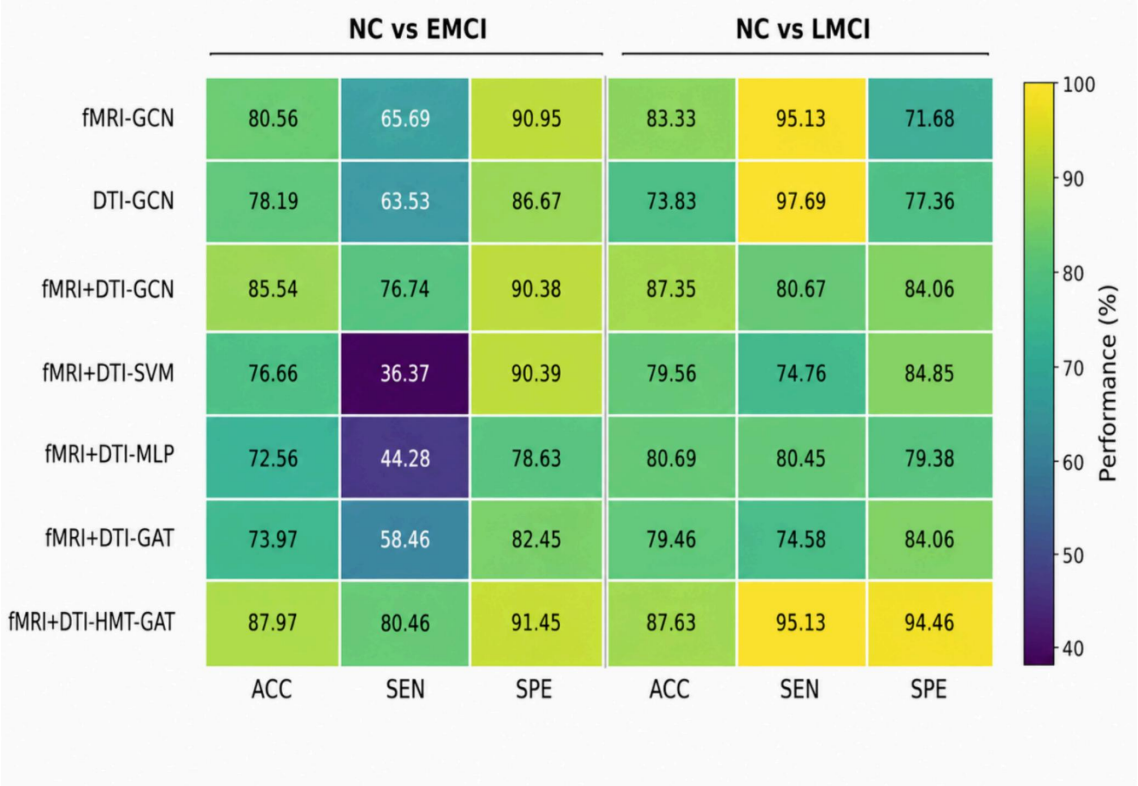

Figure 6. Classification performance comparison of different modality-method combinations on NC vs. EMCI and NC vs. LMCI tasks

The heatmap illustrates the performance of various multi-modal classification methods for distinguishing cognitive states, using key metrics: Accuracy (ACC), Sensitivity (SEN), and Specificity (SPE). It shows that combining modalities like fMRI+DTI generally enhances Sensitivity and Specificity, particularly with models like GAT. Some configurations, such as fMRI+DTI (MLP), perform less effectively, especially in Sensitivity. Overall, the heatmap highlights the importance of model selection and multi-modal integration in improving early Alzheimer's disease classification.

DISCUSSION

Impact of Non-Imaging Information on Classification Performance

In earlier Graph Convolutional Network (GCN) studies, non-imaging information has been identified as an important factor influencing classification performance, where nodes sharing the same attribute are often assigned higher edge weights. For example, the experimental results in showed that incorporating gender and scanner type into GCN construction can lead to a 3% improvement in classification accuracy (ACC) for Alzheimer's Disease (AD) and Autism Spectrum Disorder (ASD) prediction tasks. Similarly, the results in

indicated that these attributes can yield an average ACC improvement of 7.1% for Normal Control (NC) and Mild Cognitive Impairment (MCI) classification. In addition, different data sources, such as those encountered in multi-center studies, also influence prediction performance for ASD

To evaluate the influence of non-imaging covariates, experiments were conducted to examine the classification performance under different combinations of demographic and acquisition information. As shown in Fig. 7, when only one type of non-imaging information is used, the average ACCs across the evaluated classification tasks are 84.2%, 84.9%, 83.8%, and 88.2% for gender, equipment type, data source, and training-label-related grouping information, respectively. The label-related grouping information was used only within the training fold and was not provided as an input for test samples.

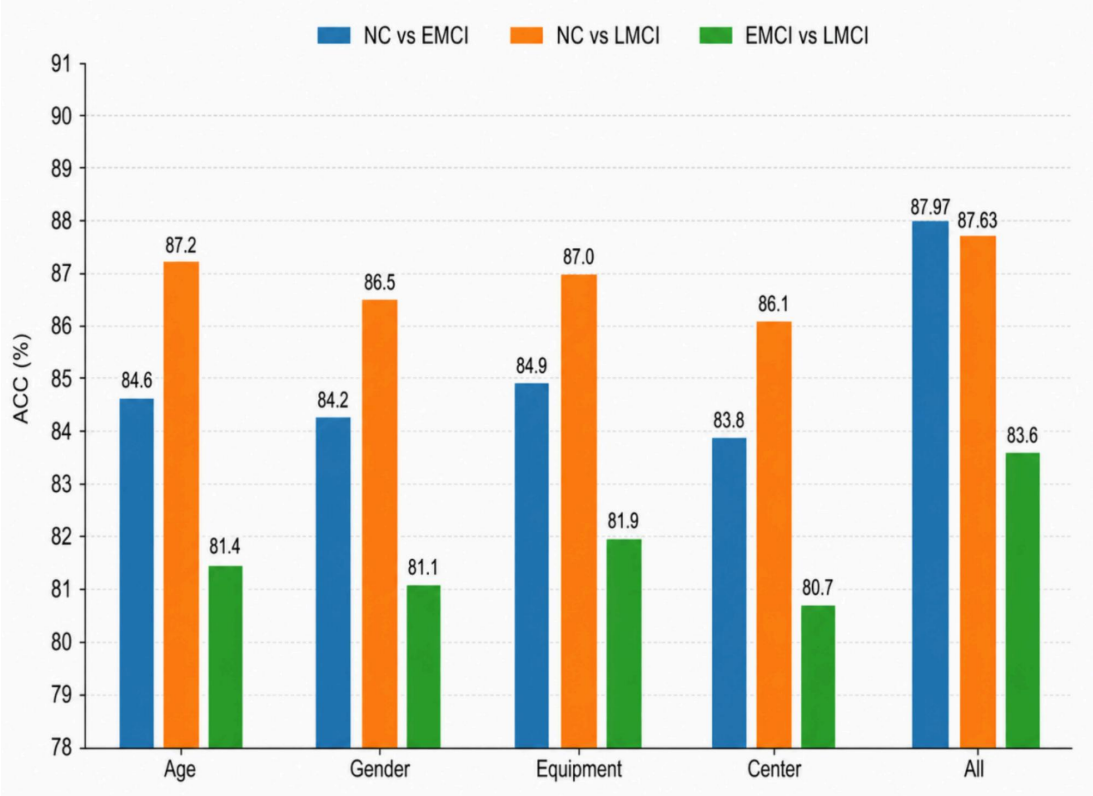

Figure 7. Effect of different non-imaging attribute settings on classification accuracy across MCI-related classification tasks

The corresponding attention coefficients learned for each non-imaging attribute during training are presented in Table 1. These coefficients were optimized jointly with the classification loss and incorporated into the phenotypic similarity matrix through a weighted combination of attribute-level similarities. Therefore, they provide an interpretable description of how demographic and acquisition-related factors modulate the final

population graph. The non-zero residual base term further ensures that the phenotypic graph retains a baseline contribution even when specific categorical attributes are weakly informative.

Influence of Feature Channel Partitioning on Model Accuracy

The selected features may exhibit heterogeneous statistical properties and different sensitivity to noise, which motivates the use of channel-wise processing. To better accommodate this heterogeneity, the selected features are partitioned into several subsets according to their statistical characteristics, and separate graph filters are assigned to each subset. As shown in Fig. 8, classification performance varies with the number of feature channels, indicating that feature partitioning affects graph filtering. As described in Section 3.4, the channel number W was treated as a validation-selected hyperparameter rather than a learnable parameter. Candidate values were set to $W \in \{1,2,3,4,5,6\}$, and $W=5$ was selected as the default setting because it achieved the best mean ACC across the evaluated MCI-related tasks.

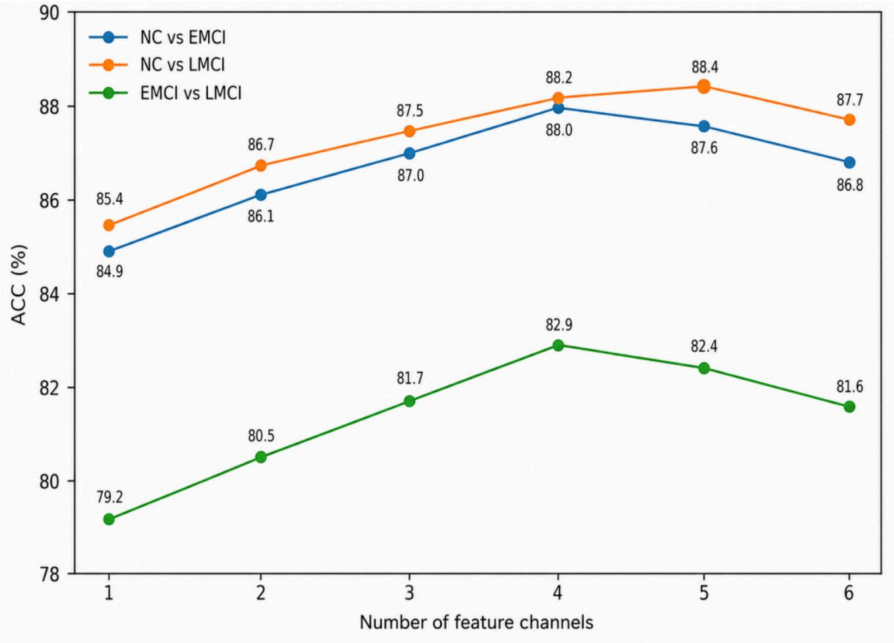

Figure 8. Effect of the number of feature channels on classification accuracy.

As shown in Fig. 8, classification performance varies with the number of feature channels, indicating that feature partitioning affects graph filtering.

In the current setting, a moderate number of channels leads to more stable performance than using too few or too many channels. Specifically, the NC vs. EMCI task reaches its highest ACC when four channels

are used, whereas the NC vs. LMCI and EMCI vs. LMCI tasks achieve their best performance with five channels. Since *W*=5 gives the best mean performance across the evaluated tasks, it was selected as the default channel number in the final model. These results suggest that channel-wise feature partitioning helps balance discriminative feature extraction and model stability. From the perspective of model complexity, the parameter-efficiency claim holds under the disjoint feature partitioning setting. Specifically, the ranked features are divided into non-overlapping subsets, and each feature is assigned to only one be the number of features assigned to the w-th channel, where $\sum_{w=1}^{W} M_w = M$. If each channel-specific graph filter uses the same hidden dimension Q_1 , the total number of channel-specific filtering parameters is:

$$\sum_{w=1}^{W} M_w \times Q_1 = M \times Q_1 \tag{24}$$

This is the same as the parameter count of a conventional single-channel transformation with M input features and Q_1 hidden units. Therefore, the multi-channel design preserves parameter efficiency because each branch operates only on its assigned feature subset rather than on the full feature vector. The concatenation operation itself is parameter-free; any additional projection layer after concatenation is counted separately and kept identical across the compared settings.

In the final classification stage, stratified 10-fold cross-validation was conducted to evaluate the model's performance. The stratified setting was used to preserve the class distribution of NC, EMCI, and LMCI samples in each fold as much as possible, which is particularly important because the LMCI group contained a relatively small number of subjects. For each validation round, the model was trained on nine folds and tested on the remaining fold. The testing fold was kept completely independent from model training and parameter selection. The evaluation metrics included Accuracy (ACC), Sensitivity (SEN), Specificity (SPE), and the Area Under the Receiver Operating Characteristic Curve (AUC). These metrics were reported together to provide a more balanced evaluation under class-imbalanced conditions. Key training parameters were set as follows: dropout rate of 0.1, learning rate of 0.005, and 200 training epochs. Dropout and sparsity-constrained graph construction were used to reduce overfitting, while the relatively compact multi-channel graph attention architecture helped control model complexity.

To validate the effectiveness of the proposed model, several representative baseline algorithms were implemented for comparison. These include Support Vector Machine (SVM), which achieved an average ACC of 73.3%, and Logistic Regression, which obtained an average ACC of 76.6%.

As shown in Table 4, HMT-GAT achieves the highest ACC in both NC vs. EMCI and NC vs. LMCI tasks among the implemented methods. For the NC vs. EMCI task, HMT-GAT improves ACC from 85.54% obtained by fMRI+DTI-GCN to 87.97%, while also improving SEN from 76.74% to 80.46% and SPE from 90.38% to 91.45%. For the NC vs. LMCI task, HMT-GAT achieves an ACC of 87.63%, which is slightly higher than fMRI+DTI-GCN, and obtains a more balanced performance with SEN of 95.13% and SPE of 94.46%. Although DTI-GCN shows a higher SEN in the NC vs. LMCI task, its ACC and SPE are lower, suggesting that HMT-GAT provides a better overall balance between sensitivity and specificity.

Table 4. Performance Comparison of Different Methods and Modalities on NC vs. EMCI and NC vs. LMCI Classification Tasks

Modality	Method	NC vs EMCI			NC vs LMCI		
		ACC(%)	SEN(%)	SPE(%)	ACC(%)	SEN(%)	SPE(%)
fMRI	GCN	80.56	65.69	90.95	83.33	95.13	71.68
DTI	GCN	78.19	63.53	86.67	73.83	97.69	77.36
fMRI + DTI	GCN	85.54	76.74	90.38	87.35	80.67	84.06
fMRI + DTI	SVM	76.66	36.37	90.39	79.56	74.76	84.85
fMRI + DTI	MLP	72.56	44.28	78.63	80.69	80.45	79.38
fMRI + DTI	GAT	73.97	58.46	82.45	79.46	74.58	84.06
fMRI + DTI	HMT-GAT	87.97	80.46	91.45	87.63	95.13	94.46

Baseline implementation and fair comparison protocol

To ensure a fair comparison, all implemented baseline models were evaluated using the same dataset, preprocessing procedure, subject-level feature representation, and stratified 10-fold cross-validation splits as the proposed HMT-GAT model. The baseline methods included SVM, MLP, GCN, and GAT, because these models represent conventional machine learning, standard fully connected neural networks, graph convolution-based learning, and attention-based graph learning, respectively. For SVM, the regularization parameter and kernel-related parameters were selected through grid search within the training folds [27]. For MLP, GCN, and GAT, the hidden dimension, learning rate, dropout rate, and number of training epochs were tuned using only the training folds. The test fold was kept independent and was not used for model selection or hyperparameter tuning.

For graph-based baselines, the same population graph construction strategy and the same input features were used whenever applicable. The GCN baseline used graph convolution layers with the same fused fMRI–DTI features as input. The GAT baseline replaced graph convolution with graph attention layers while keeping the same data split and training protocol. The proposed HMT-GAT was evaluated under the same cross-validation setting, with dropout rate set to 0.1, learning rate set to 0.005, and 200 training epochs. This unified evaluation protocol was used to reduce bias caused by different data splits or preprocessing settings.

Influence of Edge Pooling Rate on Classification Accuracy

To reduce redundant connections and improve the stability of graph filtering, an edge pooling strategy without additional trainable parameters is introduced. Based on the proposed HMT-GAT framework, experiments are conducted by varying the pooling rate from 0% to 30% to investigate its influence on classification performance. The corresponding results are illustrated in Fig. 9.

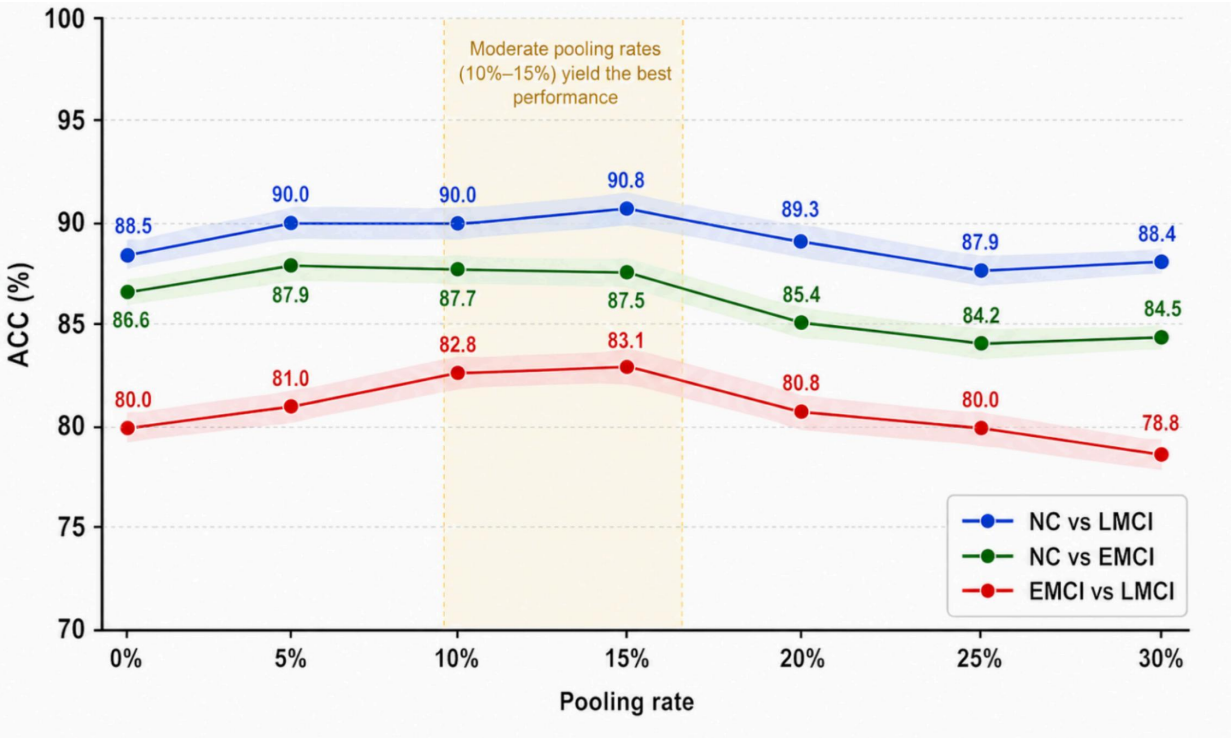

Figure 9. Effect of the pooling rate on classification accuracy across MCI-related classification tasks

Experimental findings indicate that the edge pooling rate has a noticeable influence on classification accuracy. As shown in Fig. 9, the NC vs. LMCI task achieves the highest ACC when the pooling rate is set to 15%, while

the NC vs. EMCI task reaches its best performance at 5%. For the EMCI vs. LMCI task, the ACC increases as the pooling rate rises from 0% to 15%, and then gradually decreases when the pooling rate is further increased. Overall, a moderate pooling rate, especially around 10%–15%, tends to preserve useful graph structure while reducing redundant connections. In contrast, excessive pooling may remove informative edges or nodes, leading to a decline in classification performance.

Overall, the average accuracy across the evaluated MCI-related classification tasks is highest when the pooling rate is around 15%. This result suggests that a moderate edge pooling rate can reduce redundant or noisy connections while preserving informative graph structure.

Discriminative Connectivity Features and Related ROIs

To interpret the effectiveness of the proposed framework from a biological perspective, an analysis of the discriminative features is conducted following the classification stage. The proposed model achieves the highest classification accuracy across all tasks, and a key contributing factor lies in the dual-modality brain connectivity network, which captures more robust and biologically relevant interactions by fusing rs-fMRI and DTI data. This fusion facilitates the identification of brain ROIs that are closely associated with AD and MCI, thus enhancing the model's interpretability and diagnostic value.

In the fused connectivity network, several well-established AD/MCI-related ROIs were consistently identified [28]. Representative regions included the right inferior temporal gyrus (ITG.R), right amygdala (AMYG.R), right insula (INS.R), left olfactory cortex (OLF.L), left angular gyrus (ANG.L), and right precuneus (PCUN.R). These regions are broadly consistent with previous evidence that AD/MCI-related alterations involve distributed cortical and subcortical networks [29]. Among them, the ITG.R is especially notable for its involvement in advanced cognitive functions and emerges as the most significant region across multiple tasks.

Table 5 lists the top 10 ROIs most strongly associated with each classification task. When ROIs are extracted from single-modality connectivity networks (based on either fMRI or DTI), only a portion of them are found to be closely related to early-stage AD. In contrast, most of the top 10 ROIs derived from the proposed dual-modality fusion network show strong and consistent associations with early AD pathology. For instance, from fMRI-derived features, regions such as the hippocampus (HIP.L), putamen (PUT.L), amygdala (AMYG.L), and middle occipital gyrus (MOG.R) are identified all of which are known to be implicated in early AD [30]. Similarly, in the DTI-based features, the insula (INS.R), postcentral gyrus (PoCG.R), precuneus (PCUN.R), and

supplementary motor area (SMA.R) are highlighted as structurally affected in the early stages of neurodegeneration.

Table 5. Top 10 Discriminative Features extracted by HMT-GAT framework

DTI		fMRI		Ours	
ROI Index	ROI abbr	ROI Index	ROI abbr	ROI Index	ROI abbr
86	MTG.R	24	SFGmed.R	90	ITG.R
87	TPOmid.L	37	HIP.L	30	INS.R
59	SPG.L	71	CAU.L	21	OLF.L
30	INS.R	61	IPL.L	61	IPL.L
78	THA.R	73	PUT.L	65	ANG.L
58	PoCG.R	41	AMYG.L	42	AMYG.R
68	PCUN.R	84	TPOsup.R	68	PCUN.R
75	PAL.L	87	TPOmid.L	40	PHG.R
56	FFG.R	36	PCG.R	74	PUT.R
20	SMA.R	52	MOG.R	78	THA.R

Figure 10 summarizes, for each classification task, the 20 most discriminative connectivity patterns together with the 10 most influential ROIs. These results reveal that distinct connectivity patterns and ROIs are emphasized across different diagnostic group comparisons, supporting the biological heterogeneity of AD progression.

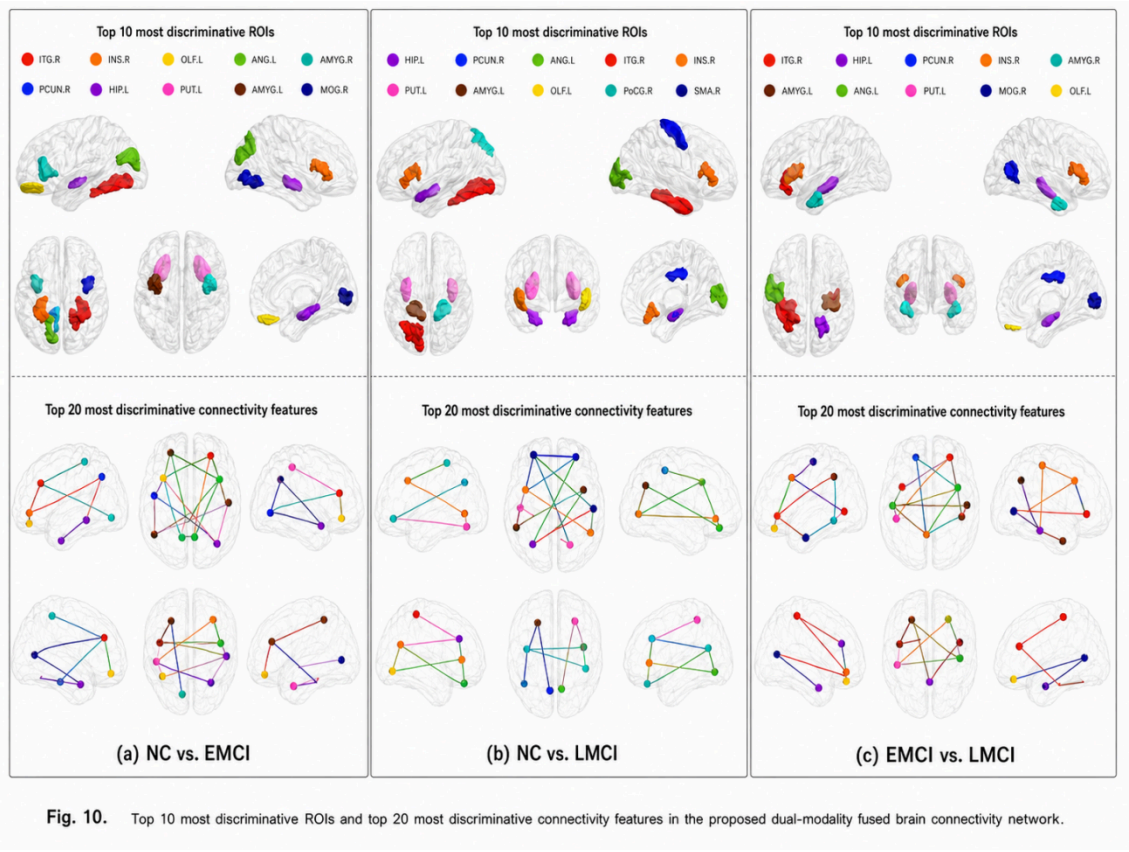

Figure 10. Discriminative ROIs and connectivity patterns identified from the dual-modality fused brain network

Overall, the incorporation of dual-modality fusion not only improves classification accuracy but also enables the discovery of clinically meaningful biomarkers, demonstrating the potential of the proposed model for precision diagnosis and early detection of neurodegenerative diseases.

Limitations and generalizability under limited LMCI samples

Although the proposed HMT-GAT framework achieved promising performance in MCI-related classification tasks, the relatively small number of LMCI samples remains an important limitation of this study. In the current dataset, the LMCI group contained only 21 subjects, which may reduce statistical power and increase the risk of unstable model estimation. To mitigate this issue, stratified 10-fold cross-validation was adopted to preserve the class distribution of NC, EMCI, and LMCI samples as much as possible in each fold. In addition, dropout regularization, sparsity-constrained brain network construction, and a compact multi-channel graph attention structure were used to reduce overfitting. Multiple evaluation metrics, including ACC, SEN, SPE, and AUC, were also reported to provide a more balanced assessment under class-imbalanced conditions. Nevertheless, the results involving LMCI should still be interpreted with caution. Future studies will further validate the

proposed framework on larger, more balanced, and independent multi-center cohorts, and external validation will be introduced to better assess the generalizability of the model.

Methodological comparison with other GCN variants

Several improved GCN variants, including HGNN, Graph U-Net, GraphSAGE, GAT, and AGNN, have been proposed to enhance graph representation learning from different perspectives. In this study, these models are discussed as representative graph neural network architectures rather than as directly re-implemented experimental baselines, unless otherwise specified in table 4. The quantitative comparison in this manuscript is limited to the implemented methods listed in table 4, including SVM, MLP, GCN, GAT, and the proposed HMT-GAT, all of which were evaluated under the same data split and validation protocol.

HGNN extends conventional pairwise graph modeling by introducing hyperedges to capture high-order relationships among multiple nodes. Graph U-Net introduces graph pooling and unpooling operations to learn hierarchical graph representations. GraphSAGE learns aggregation functions from local neighborhoods and is useful for inductive graph learning. AGNN dynamically updates node interactions according to feature similarity, while GAT uses attention coefficients to assign different weights to neighboring nodes. These methods provide valuable directions for improving graph representation learning. However, they either require additional architectural components, introduce extra trainable parameters, or focus mainly on structural propagation without explicitly modeling multimodal brain connectivity and non-imaging covariates in the same framework.

Compared with these representative GCN variants, the proposed HMT-GAT focuses on phenotype-aware multimodal brain network learning for MCI identification. Its main difference lies in the combination of three components: structurally guided fMRI–DTI fusion, non-imaging information-driven population graph construction, and multi-channel graph attention learning. Therefore, the purpose of this subsection is to clarify the methodological relationship between HMT-GAT and existing GCN variants, rather than to claim direct quantitative superiority over methods that were not re-implemented in the present experiments.

CONCLUSION

This study addresses the need for accurate early identification of MCI, a prodromal stage of AD, by proposing the HMT-GAT framework. Conventional approaches often process fMRI and DTI separately or fuse them at a later stage, thereby failing to exploit their synergistic potential. Moreover, inter-site heterogeneity and the oversmoothing effect inherent in deep GCNs further hinder classification accuracy.

A structural connectivity–guided fusion strategy is introduced to constrain the construction of functional connectivity networks using DTI-derived structural information. This biologically grounded integration at the graph level substantially improves the network's discriminative capacity in distinguishing MCI from NC. Comparative experiments indicate that the proposed fusion strategy improves classification performance compared with the implemented baseline settings under the same evaluation protocol.

The impact of demographic and acquisition-related covariates, including acquisition site, gender, scanner type, and training-fold label-related grouping information, was also assessed. The label-related grouping information was used only within the training folds and was not provided as an input feature for test samples. Notably, embedding disease status into graph construction yields the most substantial performance gain, followed by multi-center attributes, thus confirming the impact of site-specific heterogeneity. The proposed attention mechanism effectively incorporates these covariates into the graph architecture, boosting model robustness and generalization.

Experimental findings also underscore the efficacy of the multi-channel GAT design, which enables fine-grained filtering by assigning dedicated channels to statistically distinct feature subsets. This architecture minimizes noise propagation and enhances the representation of discriminative features. Additionally, a pooling mechanism based on disease-state similarity refines the graph structure by excluding diagnostically irrelevant nodes, leading to an observed accuracy gain of up to 2.3%.

In summary, HMT-GAT substantially improves feature filtering and classification outcomes, as validated by consistently superior accuracy, sensitivity, and specificity metrics. Beyond the context of MCI detection, the proposed framework provides a robust, interpretable, and transferable solution for multi-modal graph learning in neuroimaging-based diagnostic applications.

Ethics Approval and Consent to Participate

This study was approved by the Ethics Committee of Shanghai Maritime University, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

Author Contributions

Conceptualization – Su P; methodology – Su P and Kong W; formal analysis – Su P; investigation – Kong W and Wang S; resources – Su P; writing-original draft preparation – Su P and Kong W; writing-review and editing

– Kong W and Wang S; visualization – Su P; supervision – Kong W. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This research received no external funding.

Acknowledgements

Not applicable.

Data Sharing Agreement

Availability of Data and Materials The authors appreciate the Alzheimer's Disease Neuroimaging Initiative (ADNI) for contributing data (The study included 133 participants, with data obtained from both rf-MRI and DTI modalities.) (https://adni.loni.usc.edu).

REFERENCES

[1] Kim J, Jeong M, Stiles WR, Choi HS. Neuroimaging Modalities in Alzheimer's Disease: Diagnosis and Clinical Features. Int J Mol Sci. 2022;23(11):6079. Published 2022 May 28. doi:10.3390/ijms23116079.

[2] Dang C, Wang Y, Li Q, Lu Y. Neuroimaging modalities in the detection of Alzheimer's disease-associated biomarkers. Psychoradiology. 2023;3:kkad009. Published 2023 Jun 22. doi:10.1093/psyrad/kkad009.

[3] Ibrahim B, Suppiah S, Ibrahim N, Mohamad M, Hassan HA, Nasser NS, Saripan MI. Diagnostic power of resting-state fMRI for detection of network connectivity in Alzheimer's disease and mild cognitive impairment: A systematic review. Hum Brain Mapp. 2021;42(9):2941-2968. Published 2021 May 4. doi:10.1002/hbm.25369.

[4] Knudsen LV, Gazerani P, Duan Y, Michel TM, Vafaee MS. The role of multimodal MRI in mild cognitive impairment and Alzheimer's disease. J Neuroimaging. 2022;32(1):148-157. doi:10.1111/jon.12940.

[5] Cui Y, Liu C, Wang Y, Xie H. Multimodal magnetic resonance scans of patients with mild cognitive impairment. Dement Neuropsychol. 2023;17:e20230017. Published 2023 Dec 15. doi:10.1590/1980-5764-DN-2023-0017.

[6] Xu S, Fan Y, Mao C, Hu Z, Yang Z, Qu L, Xu Y, Yu L, Zhu X. Multimodal magnetic resonance imaging analysis of early mild cognitive impairment. J Alzheimers Dis. 2025;104(4):1013-1027. doi:10.1177/13872877251321187.

[7] Ai M, Liu Y, Liu D, Yan C, Wang X, Chen X. Research progress in predicting the conversion from mild cognitive impairment to Alzheimer's disease via multimodal MRI and artificial intelligence. Front Neurol. 2025;16:1596632. doi:10.3389/fneur.2025.1596632.

[8] Zhou H, He L, Chen BY, Shen L, Zhang Y. Multi-Modal Diagnosis of Alzheimer's Disease Using Interpretable Graph Convolutional Networks. IEEE Trans Med Imaging. 2025;44(1):142-153. doi:10.1109/TMI.2024.3432531.

[9] Zhang Y, He X, Chan YH, Teng Q, Rajapakse JC. Multi-modal graph neural network for early diagnosis of Alzheimer's disease from sMRI and PET scans. Comput Biol Med. 2023;164:107328. Published 2023 Aug 7. doi:10.1016/j.compbiomed.2023.107329.

[10] Zhang S, Yang J, Zhang Y, Zhong J, Hu W, Li C, Jiang J. The Combination of a Graph Neural Network Technique and Brain Imaging to Diagnose Neurological Disorders: A Review and Outlook. Brain Sci. 2023;13(10):1462. Published 2023 Oct 16. doi:10.3390/brainsci13101462.

[11] Gao J, Liu J, Xu Y, Peng D, Wang Z. Brain age prediction using the graph neural network based on resting-state functional MRI in Alzheimer's disease. Front Neurosci. 2023;17:1222751. Published 2023 Jun 30. doi:10.3389/fnins.2023.1222751.

[12] Liu L, Li Y, Yang K. Dynamically weighted graph neural network for detection of early mild cognitive impairment. PLoS One. 2025;20(6):e0323894. doi:10.1371/journal.pone.0323894.

[13] Qu Z, Yao T, Liu X, Wang G. A Graph Convolutional Network Based on Univariate Neurodegeneration Biomarker for Alzheimer's Disease Diagnosis. IEEE J Transl Eng Health Med. 2023;11:405-416. doi:10.1109/JTEHM.2023.3293081.

[14] Sun X, Li J, Yan G, Han R. ADMGCN: Graph Convolutional Network for Alzheimer's Disease Diagnosis with a Meta-learning Paradigm. Bioinformatics. 2025. doi:10.1093/bioinformatics/btaf580.

[15] Fu Y, Jiang L, Detre JA, Wang Z. Alzheimer's disease classification using mutual information generated graph convolutional network for functional MRI. J Alzheimers Dis. 2025;106(3):1021-1035. Published 2025 Jul 15. doi:10.1177/13872877251350306.

[16] Feng J, Zhao X, Liu Z, Ding Y, Wang F. A multi-view multimodal deep learning framework for Alzheimer's disease diagnosis. Front Neurosci. 2025;19:1658776. Published 2025 Oct 1. doi:10.3389/fnins.2025.1658776.

[17] Adarsh V, Gangadharan GR, Fiore U, Zanetti P. Multimodal classification of Alzheimer's disease and mild cognitive impairment using custom MKSCDDL kernel over CNN with transparent decision-making for explainable diagnosis. Sci Rep. 2024;14(1):1774. Published 2024 Jan 20. doi:10.1038/s41598-024-52185-2.

[18] Liu J, Tan G, Lan W, Wang J. Identification of early mild cognitive impairment using multi-modal data and graph convolutional networks. BMC Bioinformatics. 2020;21(Suppl 6):123. doi:10.1186/s12859-020-3437-8.

[19] He J, Wang P, He J, Sun C, Xu X, Zhang L, Wang X, Gao X. Utilizing graph convolutional networks for identification of mild cognitive impairment from single modal fMRI data: a multiconnection pattern combination approach. Cereb Cortex. 2024;34(3):bhae065. doi:10.1093/cercor/bhae065.

[20] Wee CY, Liu C, Lee A, Poh JS, Ji H, Qiu A; Alzheimer's Disease Neuroimaging Initiative. Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations. Neuroimage Clin. 2019;23:101929. doi:10.1016/j.nicl.2019.101929.

[21] Fang J, Lin D, Schulz SC, Xu Z, Calhoun VD, Wang YP. Joint sparse canonical correlation analysis for detecting differential imaging genetics modules. Bioinformatics. 2016;32(22):3480-3488. doi:10.1093/bioinformatics/btw485.

[22] Moon S, Lee H. JDSNMF: Joint Deep Semi-Non-Negative Matrix Factorization for Learning Integrative Representation of Molecular Signals in Alzheimer's Disease. J Pers Med. 2021;11(8):686. Published 2021 Jul 21. doi:10.3390/jpm11080686.

[23] Deng J, Zeng W, Luo S, Kong W, Shi Y, Li Y, Zhang H. Integrating multiple genomic imaging data for the study of lung metastasis in sarcomas using multi-dimensional constrained joint non-negative matrix factorization. Inf Sci. 2021;576:24-36. doi:10.1016/j.ins.2021.06.033.

[24] Liu W, Liu L, Cheng X, Ge H, Hu G, Xue C, Qi W, Xu W, Chen S, Gao R, Rao J, Chen J. Functional Integrity of Executive Control Network Contributed to Retained Executive Abilities in Mild Cognitive Impairment. Front Aging Neurosci. 2021;13:710172. Published 2021 Nov 29. doi:10.3389/fnagi.2021.710172.

[25] Song Y, Xu W, Chen S, Hu G, Ge H, Xue C, Qi W, Lin X, Chen J. Functional MRI-Specific Alterations in Salience Network in Mild Cognitive Impairment: An ALE Meta-Analysis. Front Aging Neurosci. 2021;13:695210. Published 2021 Jul 23. doi:10.3389/fnagi.2021.695210.

[26] Boughorbel S, Jarray F, El-Anbari M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLoS One. 2017;12(6):e0177678. Published 2017 Jun 2. doi:10.1371/journal.pone.0177678.

[27] Jitsuishi T, Yamaguchi A. Searching for optimal machine learning model to classify mild cognitive impairment subtypes using multimodal MRI data. Sci Rep. 2022;12(1):4284. Published 2022 Mar 11. doi:10.1038/s41598-022-08231-y.

[28] Wang Y, Li Q, Yao L, He N, Tang Y, Chen L, Long F, Chen Y, Kemp GJ, Lui S, Li F. Shared and differing functional connectivity abnormalities of the default mode network in mild cognitive impairment and Alzheimer's disease. Cereb Cortex. 2024;34(3):bhae094. doi:10.1093/cercor/bhae094.

[29] Kwak K, Giovanello KS, Bozoki A, Styner M, Dayan E. Subtyping of mild cognitive impairment using a deep learning model based on brain atrophy patterns. Cell Rep Med. 2021;2(12):100467. doi:10.1016/j.xcrm.2021.100467.

[30] de Jong LW, van der Hiele K, Veer IM, Houwing JJ, Westendorp RGJ, Bollen ELEM, de Bruin PW, Middelkoop HAM, van Buchem MA, van der Grond J. Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. Brain. 2008;131(Pt 12):3277-3285. doi:10.1093/brain/awn278.