



ELSEVIER

Contents lists available at ScienceDirect

# Information Processing and Management

journal homepage: [www.elsevier.com/locate/infproman](http://www.elsevier.com/locate/infproman)

## Multimodality Alzheimer's Disease Analysis in Deep Riemannian Manifold

Junbo Ma<sup>a,\*</sup>, Jilian Zhang<sup>b,\*</sup>, Zeyu Wang<sup>c</sup><sup>a</sup> Guangxi Key Lab of Multi-source Information Mining and Security, Guangxi Normal University, Guilin, Guangxi, China<sup>b</sup> College of Cyber security, Jinan University, Guangzhou China<sup>c</sup> The College of Information Science and Engineering, Guilin University of Technology, Guilin, China

### ARTICLE INFO

#### Keywords:

Alzheimer's Disease diagnosis  
Multimodality fusion  
Symmetric positive definite matrix  
Riemannian manifold  
cross-modal interaction

### ABSTRACT

Structural Magnetic Resonance Image (sMRI) and functional MRI (fMRI) are two of the most important modalities to unveil brain disorders for Alzheimer's disease (AD) analysis. Comprehensively utilizing both modalities is the way to ensure an accurate AD diagnosis. Currently, the most common computational approach to aid the AD diagnosis is to formulate the sMRI and fMRI into graphs and then extract discriminative features through Graph Neural Networks (GNNs). However, most GNNs rely heavily on the aggregation operation on each node, which exploits the local topological information from the neighborhood nodes but does not fully respect the characteristics of the global graph topology. Also, only a few works addressed the structural and functional coupling problem on the graphs. In this paper, a novel Riemannian manifold-based model, called Cross-Modal Riemannian Network (CMRN), is proposed to solve the above issues, which respects the global topologies and invariant characteristics of the sMRI and fMRI graphs by fully operating on the Riemannian Manifold. Furthermore, a novel cross-modal attention mechanism is proposed to enable the interactions between two modalities on the Riemannian manifold, which helps the model comprehensively utilize both modalities to identify the most discriminative information for AD diagnosis. Extensive experimental results on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset demonstrate the effectiveness of the proposed method.

### 1. Introduction

Alzheimer's disease (AD) is a neurodegenerative dementia disease. AD patients will experience a progressive loss of cognitive abilities such as memory failure and personality changes, which will then be severe enough to interfere with their daily lives (Alzheimer's Association 2021; Ma et al. 2020; Tanveer et al. 2020). A most recent report from the Alzheimer's Association says that more than 6 million Americans suffer from AD, which costs \$355 billion for health care per year (Alzheimer's Association 2021). Unfortunately, the disease progression is irreversible and still has no cure yet (Gonneaud et al. 2021). Thus, early diagnosis is essential to introduce early medical interventions to moderate cognitive decline (Swati, Kumar, and Namasudra 2022).

Magnetic Resonance Imaging (MRI) is one of the most effective tools for AD diagnoses, as brain damages usually happen long before the cognitive symptoms (Pisano et al. 2021). The two most common MRI modalities are structural MRI (sMRI) and functional

\* The corresponding authors

E-mail addresses: [nudt\\_mjb@outlook.com](mailto:nudt_mjb@outlook.com) (J. Ma), [zhangjilian@jnu.edu.cn](mailto:zhangjilian@jnu.edu.cn) (J. Zhang), [Neptuneusyd@gmail.com](mailto:Neptuneusyd@gmail.com) (Z. Wang).

MRI (fMRI). The sMRI provides the anatomic structure of the brain by tracking physical fiber connections between brain regions. The fMRI represents the brain's cognitive activities through the changes of blood-oxygen-level-dependent (BOLD) signals in brain regions. The brain regions are usually divided based on the widely-accepted brain parcellation templates (Destrieux et al. 2010; Gordon et al. 2016). These brain regions will then be presented as nodes on the brain connectivity graph, while the edges between these nodes will be derived from sMRI to produce the structural connectivity (SC) graph or from fMRI to produce the functional connectivity (FC) graph (Wang et al. 2022). Each connectivity graph represents the brain from different perspectives. Thus, comprehensively utilizing both modalities is the key to identifying high-risk subjects (Sreedevi et al. 2022). In practice, because of the complexity of AD, some subjects may cognitively function well but already have brain degenerations, whereas others may not show significant brain damage but already start losing memories (Alzheimer's Association 2021; Gonneaud et al. 2021). Such complexity requires our computation aid diagnosis tools to have the ability to comprehensively process and fully utilize the information from both modalities (Veitch et al. 2019).

With the fast development of deep learning techniques, the Graph Neural Network (GNN) has become the most widely used method to extract information from graphs to perform AD diagnosis. Ma et al. (2020) innovatively proposed an attention-guided deep graph neural network to analyze AD data from the sMRI, which can reveal the most relevant brain regions and the essential time point for AD progression. Zhu et al. (2022) proposed a dynamic graph convolutional network and made it more interpretable with feature learning techniques. Gan et al. (2021) proposed a multi-graph fusion method to explore the common and complementary information between functional brain networks. However, these methods only focused on either SC or FC, which do not fully use the comprehensive information from the other modality.

Recently, several works have been proposed to address the SC and FC coupling problem when analyzing brain networks. Huang et al. (2020) proposed an Attention-Diffusion-Bilinear neural network (ADB-NN) by utilizing the innovative attention diffusion map to integrate information from both SC and FC, which also refine the node representation through both direct and indirect connections. Chu, Parhi, & Lenglet (2018) proposed a joint structural-functional brain network model to recover the under-estimated SC connection through FC and improve the anatomical circuit estimation. However, these GNN based methods still have a major drawback of negligent the global topological information of the graph's geometric structure caused by the node aggregation operation (Kipf and Welling 2017; Yuan et al. 2021). The aggregation operation is one of the essential operations in GNN, which aggregates the information from all the neighborhood nodes of each node on the graph to exploit the graph topology. Thus, the local topological information from each node dominates the extracted information from GNN.

To preserve the global intrinsic geometry of the brain networks, manifold learning has been introduced to extract the topological information from the SC and FC matrices. A manifold is a low-dimensional geometry lying in a high-dimensional Euclidean space (Ke et al. 2021; Pennec 2006). Manifold learning assumes that the data samples are lying on a smooth manifold. Thus, any data operation should respect the intrinsic geometry and the topological properties of the manifold (Huang and Gool 2017). Ke et al. (2021) proposed a deep manifold learning model to analyze the dynamics in the cardiac MRI data. Huang et al. (2021) proposed a geometric deep learning model to detect the change points in the functional brain networks. However, these methods only focus on one modality without considering the comprehensive information from SC and FC coupling.

To solve the above issues, a novel Riemannian manifold-based model, called Cross-Modal Riemannian Network (CMRN), is proposed. Riemannian manifold is a differentiable geometry equipped with a Riemannian metric, which is a positive-definite inner product on the tangent space at each manifold point to formulate the manifold measurements (Huang and Gool 2017; Kim et al. 2021; Lin and Zha 2008). The SC matrices are derived from diffusion tensor magnetic resonance imaging (DTI), which are commonly Symmetric Positive Definite (SPD) matrices lying on the Riemannian manifold (Arsigny et al. 2006; Pennec 2006; Pennec, Fillard, and Ayache 2006). Similarly, the FC matrices are covariance matrices derived from statistical measures such as the Pearson correlation, which are also commonly SPD matrices (Brosch and Tam 2013; Huang et al. 2021). Thus, it is natural to exploit the SPD manifold properties of Riemannian geometry on SC and FC matrices. To fully equip the Riemannian geometry, the proposed CMRN model utilizes the Riemannian network operations proposed in (Huang and Gool 2017), such as the BiMapLayer and the ReEigLayer. Thus, the proposed CMRN model will process the SPD matrix on the Riemannian manifold throughout the model and can be trained by stochastic gradient descent through backpropagation on Stiefel manifolds (Huang and Gool 2017).

To ensure cross-modal interactions between SC and FC, a novel cross-modal attention (CMA) mechanism is proposed for the CMRN. Inspired by the self-attention mechanism (Vaswani et al. 2017), the proposed CMA will query one modality information with the keys from the other modality. Then, the geodesic distance between the query and key is used to estimate how much information should be passed by. The proposed CMA also utilizes the Riemannian network operations in (Huang and Gool 2017) to ensure that the overall model fully exploits the SPD manifold properties of Riemannian geometry.

The main contributions of this paper can be summarized as follows:

- 1) A novel Cross-Modal Riemannian Network (CMRN) is proposed to simultaneously extract the topological information from both SC and FC modalities with respect to their global intrinsic geometries on the Riemannian manifold.
- 2) A novel cross-modal attention mechanism is proposed to enable cross-modal interactions on the Riemannian manifold. To the best of our knowledge, this is the first work that addresses the SC and FC coupling problem on the Riemannian manifold.
- 3) Extensive experimental results and analysis on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset demonstrate the effectiveness and efficiency of the proposed method.

## 2. Related works

### 2.1. Graph Neural Networks

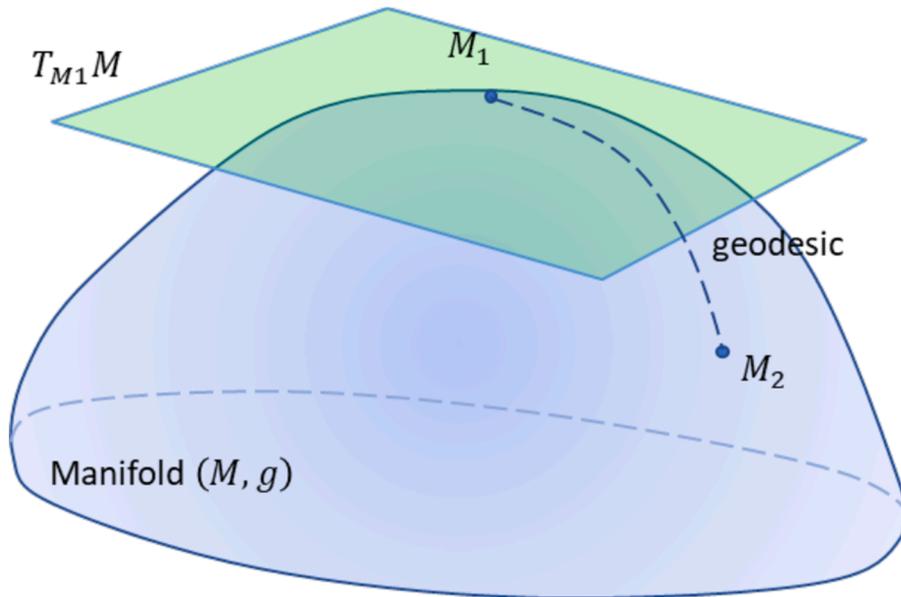
Graph Neural Network (GNN) is a type of deep neural network which aims to process irregular graph data through deep learning techniques (Peng et al. 2022; Yuan et al. 2021). Nowadays, many state-of-the-art methods are all based on GNNs, such as the citation network analysis (Kipf and Welling 2017; Zhu et al. 2017; Zhu, Liu, and Liu 2021), social network analysis (Liao et al. 2022; Wu et al. 2020), and computer aid diagnosis (Ahmedt-Aristizabal et al. 2021; Yu, Wang, and Zhang 2021; Zhu et al. 2022). The GNNs can be roughly classified into two categories (Zhou et al. 2020). One is the spectral-based method, which utilizes the graph Fourier transformation to filter the graph signals (Defferrard, Bresson, and Vandergheynst 2016; Kipf and Welling 2017). The other one is spatial-based methods, which mimic traditional convolutional neural networks by sampling and aggregating node features from their neighborhood (Velickovic et al. 2018). However, the spectral-based methods rely on local filters, and the spatial-based methods rely on local neighborhoods. Both methods are a lake of the ability to handle global graph topologies and their intrinsic geometries.

### 2.2. Deep learning on manifolds

Extending the success from deep learning techniques to the manifold learning field has been a hot research topic in recent years because of the unique geometric characteristics of the manifold (Kim et al. 2021; Z. Zhang et al. 2021). Dai & Hang (2021) proposed a manifold matching approach for generative modeling, which utilizes the metric learning generator to evaluate the matching geometries. Li et al. (2020) proposed a multi-view learning approach on manifolds to retrieve cartoon characters and synthesize new cartoon clips. Li et al. (2021) proposed a Manifold Learning-Based method to solve the semi-supervised Hyperspectral Image Classification problem. One special manifold is defined on the Symmetric Positive Definite (SPD) matrix space, called Riemannian Manifold. Despite many mathematical advantages of SPD algebra, the Riemannian Manifold itself has many advanced properties, such as intrinsic geometry, affine invariance, and similarity invariance, which attract researchers to apply it onto the data naturally being SPD matrices.

### 2.3. Multimodal fusion

Multimodal fusion is a challenging but efficient way to extract comprehensive information from multiple modalities (Hermessi, Mourali, and Zagrouba 2021). The fusion process can be roughly divided into three strategies, called early fusion, middle fusion, and late fusion (Nagrani et al. 2021). The early fusion starts fusing information at the very beginning of the learning process and usually begins right after the feature extraction (Y. Zhang et al. 2021). In contrast, late fusion usually happens at the end of the learning process and right before producing the final results (Uppal et al. 2022). Both strategies are not efficient in extracting the cross-modal information because early fusion mixes all the information at the very beginning, which will introduce unnecessary cross-modal correlations and noises. In contrast, late fusion does not fully utilize the cross-modal information in the feature extracting stage, which can help learning models extract more task-related features from all the modalities (Nagrani et al. 2021). Thus, middle fusion is the most



**Fig. 1.** The Riemannian Manifold. The smooth blue shape represents a Riemannian Manifold  $(M, g)$ . The green plane  $T_{M_1}M$  indicates the tangent space of the manifold at point  $M_1$ . The dotted line is the geodesic between two points  $M_1$  and  $M_2$ , which is the shortest path on the manifold.

efficient way, and many approaches have been proposed to address this problem. However, to the best of our knowledge, the middle fusion on pure Riemannian manifold is still an unsolved problem.

### 3. Method

#### 3.1. Problem formulation and Preliminaries

In this paper, the Alzheimer’s disease (AD) diagnosis is formulated into a multimodal classification problem. Given a subject with two modalities ( $S_0, F_0$ ) and class label  $y \in Y$ , the goal is to predict the class label  $\hat{y}$  with the most discriminative information extracted from both modalities. Here,  $S_0$  is the initial structural connectivity matrix, and  $F_0$  is the initial functional connectivity matrix. As discussed in the introduction section,  $S_0$  and  $F_0$  can be derived from each MRI modality into Symmetric Positive Definite (SPD) matrices,  $S_0, F_0 \in Sym^+$ , where  $Sym^+$  is the SPD space. The rows/columns of the  $S_0, F_0$  matrices represent the brain regions based on a certain brain parcellation template, and the elements inside the matrices represent the corresponding connectivities between brain regions.

Formally, a Riemannian Manifold can be defined as  $(M, g)$ , where  $M \in Sym_d^+$  is the SPD space with  $d \times d$  real matrices,  $g$  is the Riemannian metric defined as the positive-definite inner product on the tangent space  $T_pM$  at each manifold point  $p$ . As shown in Fig. 1, the distance between two points on the Riemannian Manifold is the length of the geodesic connecting them. In this paper, we will focus on the log-Euclidean Riemannian metric proposed in (Arsigny et al. 2006). Two basic matrix operations with log-Euclidean Riemannian metric are logarithmic multiplication  $\odot$  and logarithmic scalar multiplication  $\otimes$ , given by:

$$M_1 \odot M_2 = \exp(\log(M_1) + \log(M_2)), \tag{1}$$

$$\lambda \otimes M = \exp(\lambda \log(M)) = M^\lambda. \tag{2}$$

The distance between points  $M_1$  and  $M_2$  with log-Euclidean Riemannian metric can be calculate as:

$$dist(M_1, M_2) = \|\log(M_1) - \log(M_2)\|^2 = \left( Trace\left(\{\log M_1 - \log M_2\}^2\right) \right)^{\frac{1}{2}}. \tag{3}$$

The proposed Cross-Modal Riemannian Network (CMRN) will use three types of Riemannian network operations introduced in (Huang and Gool 2017), named BiMapLayer, ReEigLayer, and LogEigLayer. The BiMapLayer aims to compress the input SPD matrices to extract the discriminative information, which is done by a bilinear mapping with a weight matrix  $W_l$  at the  $l$  layer, where  $W_l \in \mathbb{R}^{d_l \times d_{l-1}}$ , ( $d_l < d_{l-1}$ ). For example, the SPD matrix in the structural modality will be compressed by Eq. (4):

$$S_l = W_{Sl} S_{l-1} W_{Sl}^T, \tag{4}$$

where  $W_{Sl}$  is the weight matrix of the structural modal at layer  $l$ . Then, the ReEigLayer will take the compressed SPD matrix  $S_l$  as input and do the non-linear transform by Eq. (5).

$$\tilde{S}_l = U_{Sl} \max(\epsilon I, \Sigma_l) U_{Sl}^T, \tag{5}$$

where  $U_{Sl}$  and  $\Sigma_l$  are aquired by eigenvalue decomposition  $S_l = U_{Sl} \Sigma_l U_{Sl}^T$ ,  $\epsilon$  is a small threshold to replace the eigenvalue that is smaller than it.  $I$  is the identity matrix and  $\max(\epsilon I, \Sigma_{l-1})$  is a diagonal matrix. The diagonal elements of  $\max(\epsilon I, \Sigma_{l-1})$  will be set to  $\epsilon$ , if the corresponding diagonal element  $\Sigma_{l-1}(i, i) \leq \epsilon$ , otherwise it will remain as  $\Sigma_{l-1}(i, i)$ . In this way, the ReEigLayer can prevent the transformed matrices from being near the non-positive SPD matrices (Huang and Gool 2017). The LogEigLayer aims to transform the

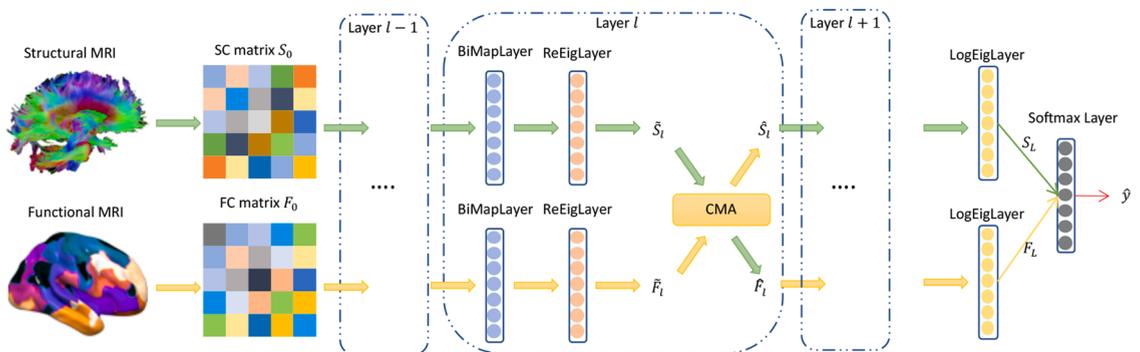


Fig. 2. The proposed Cross-Modal Riemannian Network. The fat arrows are all operated on Riemannian manifold. The thin arrows are operated on the traditional Euclidean space. The green arrows indicate the information coming from the structural modality, while the yellow arrows indicate the information coming from the functional modality. The CMA is the proposed Cross-Modal Attention module, in which the two modalities interact with each other on the Riemannian Manifold.

SPD matrices on the Riemannian manifold back to the Euclidean space with  $S_l = U_{l-1} \log \Sigma_{l-1} U_{l-1}^T$ , where  $\Sigma_{l-1}$  are the eigenvalues calculated by  $S_{l-1} = U_{l-1} \Sigma_{l-1} U_{l-1}^T$  and the logarithm operation:

$$\log \Sigma = U(\text{Diag}(\log \sigma_i))U^T = \sum_{k=1}^{+\infty} \frac{(-1)^{k+1}}{k} (\Sigma - \mathbf{I})^k. \tag{6}$$

After the LogEigLayer,  $S_l$  will become a vector, and the traditional deep learning layers can be used on it. With all these Riemannian operations and layers ready, the proposed CMRN will be introduced in the next section.

### 3.2. Cross-Modal Riemannian Network

Fig. 2 illustrates the proposed Cross-Modal Riemannian Network (CMRN). The MRI image data from each modality will first be preprocessed into SPD matrices,  $S_0$  and  $F_0$ . CMRN will take them as the input and then compress these SPD matrices layer by layer by the Riemannian network operations. In layer  $l$ , the input will be the compressed SPD matrices from the previous layer,  $S_{l-1}$  and  $F_{l-1}$ . For each modality, the CMRN will firstly utilize a BiMapLayer to get the compressed SPD matrix  $S_l$  with Eq. (4). Then, the ReEigLayer will do the non-linear transform with Eq. (5) on compressed SPD matrix  $S_l$ . After the non-linear transformation, the current structural modal matrix  $\hat{S}_l$  will be updated by the proposed Cross-Modal Attention (CMA) mechanism to engage with the information from the functional modality  $\hat{F}_l$ . The details of CMA will be introduced in the next section. The updated  $\hat{S}_l$  and  $\hat{F}_l$  by CMA will be the final output of layer  $l$ . Note that all the operations inside layer  $l$  are based on the Riemannian manifold operations, and all the CMRN layers except the last layer are all Riemannian network layers. Thus, the proposed CMRN model can fully utilize the geometry characteristics of the input SPD matrices at every processing step.

In the last layer  $L$  of the model, the SPD matrices need to be projected back to the Euclidean space to fulfill the classification task, which is done by the LogEigLayer with Eq. (6). The vector  $S_L$  and vector  $F_L$  will then be concatenated into a long vector and passed into a traditional softmax layer to do the final classification. As illustrated in Fig. 2, the Cross-Modal Attention (CMA) mechanism is an essential module to enable the proposed CMRN model to do cross-modal interactions and learn the complementary information from both modalities. In the next section, a detailed introduction of CMA will be presented.

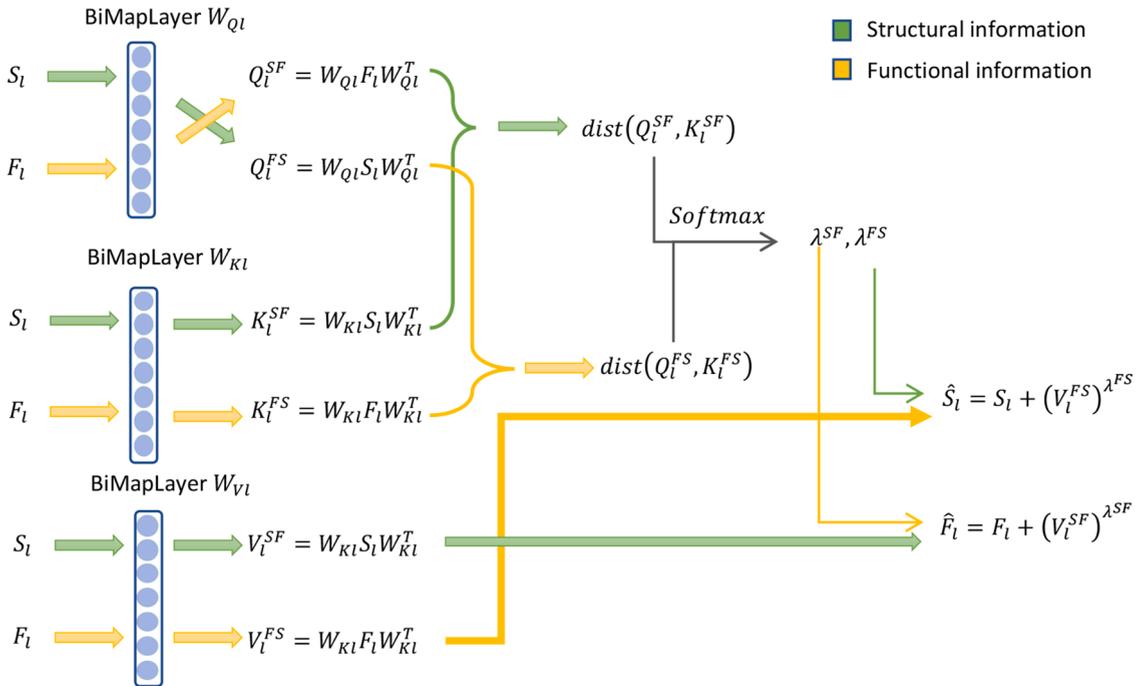


Fig. 3. The Cross-Modal Attention (CMA) module. The fat arrows are all operated on Riemannian manifold.  $S_l$  and  $F_l$  respectively represent the output of each modality at layer  $l$ .  $Q_l^*$ ,  $K_l^*$ , and  $V_l^*$  are the query, key, and value matrices corresponding to the original self-attention mechanism. The superscript  $^{*SF}$  indicates the cross-modal information from structural modality to the functional modality, while the superscript  $^{*FS}$  indicates the opposite cross-modal information. The  $dist()$  function is calculating the distance with log-Euclidean Riemannian metrics. The final outputs are the updated information with the other modality.

### 3.3. Cross-Modal Attention Mechanism

The Cross-Modal Attention (CMA) mechanism aims to update the current modality with the cross-modal information from the other modality. The proposed Cross-Modal Attention (CMA) mechanism is inspired by the self-attention mechanism introduced in (Vaswani et al. 2017). The Eq. (4) shows the original self-attention mechanism.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (4)$$

where  $Q$ ,  $K$ ,  $V$  are the query, key, and value matrices, respectively.  $d_k$  is the dimensionality of the key vectors in  $K$ , and  $\frac{1}{\sqrt{d_k}}$  is the scaling factor. A softmax function is applied to obtain the weights on the values. The weight associated with each value vector is calculated by querying with the corresponding keys. The output of the self-attention is a weighted value on the value matrix  $V$ .

Similarly, the proposed cross-modal attention is defined as  $C^{SF} = \text{CMA}(Q^{SF}, K^{SF}, V^{SF})$  and  $C^{FS} = \text{CMA}(Q^{FS}, K^{FS}, V^{FS})$ , where  $C^{SF}$  is the cross-modal information from structural modality to the functional modality and  $C^{FS}$  is the cross-modal information from functional modality to structural modality. Obviously, the self-attention equation is built on the Euclidean space and cannot handle the SPD matrices. Thus, the Riemmanian manifold network operations are introduced into the CMA module.

As shown in Fig. 3, to update the current structural modality SPD matrix  $S_l$  with the functional modality SPD matrix  $F_l$ , a query from  $F_l$  is perform to  $S_l$ , where  $S_l$  is the key to answering this query. To enhance the generalization ability, a transformation is performed firstly with a BiMapLayer with weight  $W_{Ql}$  and  $W_{Kl}$ . Both weights  $W_{Ql}$  and  $W_{Kl}$  are set to the same target dimension to ensure the transformed SPD matrices are on the same manifold. Formally, this transformation can be written as:  $Q_l^{SF} = W_{Ql}F_lW_{Ql}^T$  and  $K_l^{SF} = W_{Kl}S_lW_{Kl}^T$ . Simultaneously, a query from the structural modality  $S_l$  to the functional modality  $F_l$  is also performed, with  $Q_l^{FS} = W_{Ql}S_lW_{Ql}^T$  and  $K_l^{FS} = W_{Kl}F_lW_{Kl}^T$ . Note that  $Q_l^{SF}$  and  $Q_l^{FS}$  are sharing the same weight  $W_{Ql}$ ,  $K_l^{SF}$  and  $K_l^{FS}$  are sharing the same weight  $W_{Kl}$ . Thus, these four SPD matrices will lie on the same Riemannian manifold. Then, the distances between the query and the key matrices are calculated on this manifold to estimate the similarities between them. Specifically, the distance scalars are calculated based on the log-Euclidean Riemannian metric in Equation (3) :

$$\text{dist}(Q_l^{SF}, K_l^{SF}) = \|\log(Q_l^{SF}) - \log(K_l^{SF})\|^2 = \left(\text{Trace}\left(\{\log Q_l^{SF} - \log K_l^{SF}\}^2\right)\right)^{\frac{1}{2}},$$

$$\text{dist}(Q_l^{FS}, K_l^{FS}) = \|\log(Q_l^{FS}) - \log(K_l^{FS})\|^2 = \left(\text{Trace}\left(\{\log Q_l^{FS} - \log K_l^{FS}\}^2\right)\right)^{\frac{1}{2}}.$$

We then normalize both distances with

$$\text{softmax}(\text{dist}(Q_l^{SF}, K_l^{SF}), \text{dist}(Q_l^{FS}, K_l^{FS})).$$

Let  $\lambda^{SF}$  and  $\lambda^{FS}$  denote the normalized distance scalars, where  $\lambda^{SF} + \lambda^{FS} = 1$ . In this way,  $\lambda^{SF}$  and  $\lambda^{FS}$  can be considered as the percentage of information to be passed from one modality to the other. Before we update the current modality matrix with the cross-modal information, a transformation of the values are also done by the BiMapLayer, with  $V_l^{SF} = W_{Vl}S_lW_{Vl}^T$  and  $V_l^{FS} = W_{Vl}F_lW_{Vl}^T$ . Finally, the cross-modal attention can be written as follows:

$$C^{SF} = \text{CMA}(Q^{SF}, K^{SF}, V^{SF}) = \text{CMA}\left(W_{Ql}F_lW_{Ql}^T, W_{Kl}S_lW_{Kl}^T, W_{Vl}S_lW_{Vl}^T\right)$$

$$C^{FS} = \text{CMA}(Q^{FS}, K^{FS}, V^{FS}) = \text{CMA}\left(W_{Ql}S_lW_{Ql}^T, W_{Kl}F_lW_{Kl}^T, W_{Vl}F_lW_{Vl}^T\right)$$

The structure modality will be updated by functional cross-modal information  $C^{FS}$  with:

$$\widehat{S}_l = S_l + \lambda^{FS} \otimes V_l^{FS} = S_l + \exp(\lambda^{FS} \log V_l^{FS}) = S_l + (V_l^{FS})^{\lambda^{FS}},$$

where  $\otimes$  is the logarithmic scalar multiplication defined in Eq. (2). And the functional modality will be updated by the structural cross-modal information  $C^{SF}$  with:

$$\widehat{F}_l = F_l + \lambda^{SF} \otimes V_l^{SF} = F_l + \exp(\lambda^{SF} \log V_l^{SF}) = F_l + (V_l^{SF})^{\lambda^{SF}}.$$

Intuitively, the proposed CMA will introduce less cross-modal information when the cross-modal similarity is high, which avoids redundant cross-cross modal information being passed. This is achieved by  $\lambda^{SF}$  and  $\lambda^{FS}$  in  $(V_l^{SF})^{\lambda^{SF}}$  and  $(V_l^{FS})^{\lambda^{FS}}$ . Since  $\lambda^{SF}$  and  $\lambda^{FS}$  are percentages of the similarity comparison, which are real values between 0 to 1 and  $\lambda^{SF} + \lambda^{FS} = 1$ . The smaller  $\lambda^{SF}$  and  $\lambda^{FS}$  means the more similar the query and key pairs. Correspondingly, the  $(V_l^{SF})^{\lambda^{SF}}$  and  $(V_l^{FS})^{\lambda^{FS}}$  will also be small.

As discussed in Section 2, the proposed CMA is a middle fusion approach, which can be plugged into any Riemannian layers in CMRN. Thus, the model can focus on extracting more modality-specific information in the early layers while utilizing cross-modal information in the later layers. When to start cross-modal information sharing is a hyperparameter that needs to be tuned to specific tasks. We will discuss how this hyperparameter impacts the model performance in the Experiment section.

## 4. Experiments

### 4.1. Data and Experiment setup

The data set used in this paper is a partition of the well-known Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset (Jack Jr. et al. 2008), which has a total of 202 sMRI and fMRI pairs. There are 51 Alzheimer’s Disease (AD) samples, 43 Mild Cognitive Impairment converters (MCIc), 56 MCI non-converters (MCI<sub>n</sub>), and 52 Normal Controls. All the sMRI and fMRI images are first preprocessed to correct the spatial distortion, strip the skull, remove the cerebellum parts, and then parcellate them into 148 cortical regions (Destrieux et al. 2010). The 148 brain parcellation template is attached in the supplementary material. For sMRI, fiber tractography is applied to construct the structural connectivity matrices. For fMRI, the Pearson correlations are calculated for each brain region pair to construct the functional connectivity matrices. Finally, the disparity filtering method is applied to each matrix to guarantee the sparsity and full connection.

Four binary classification tasks are set up to test the model performance, e.g., AD vs. NC, AD vs. MCI, NC vs. MCI, MCIc vs. MCI<sub>n</sub>. The MCI here refers to the total samples from both MCIc and MCI<sub>n</sub>. The MCI samples are considered as the high-risk groups, which have already experienced some cognitive ability loss such as memory or language but still be able to take care of themselves independently (Alzheimer’s Association 2021). Thus, the MCIc vs. MCI<sub>n</sub> will be the most challenging task, as they will have very similar patterns in the data. To further test the model’s potential, a multiple class classification task is set up as AD vs. MCIc vs. MCI<sub>n</sub> vs. NC, which will be the most difficult task in the experiments.

The competing method is called Attention-Diffusion-Bilinear Neural Network (ADB-NN), which is the state-of-the-art method addressing the SC and FC coupling problem for brain network analysis (Huang et al. 2020). ADB-NN is a Graph Neural Network based method and relies on the proposed attention diffusion map to guide the diffusion process, which can integrate SC and FC and refine the node representations from both direct and indirect connections (Huang et al. 2020). However, the ADB-NN still relies heavily on the local topological information from the neighborhood nodes, which is an excellent competing method to demonstrate the effectiveness of our proposed method.

The implementation of ADB-NN follows similar configurations in the Epilepsy Classification task (Huang et al. 2020) and is written with PyTorch. In the proposed CMRN model, we gradually compress the input matrix by 10% of the initial dimension at each layer. Specifically, the input matrices in our experiments are  $148 \times 148$ . Thus, we choose to reduce the matrix dimension by 15 in each layer, which means the matrices are compressed into 133, 118, 103, 88, and so on. We stop the manifold operation on the 8<sup>th</sup> layer, which means the output matrices are  $28 \times 28$ . Then, the LogEigLayer is used to project the SPD matrices back to the Euclidean space and a conventional softmax layer is used to perform the final classification. The proposed CMRN model is implemented by Geomstats package (Miolane et al. 2020) with PyTorch (Paszke et al. 2017) backend. All the models are trained on a deep learning server equipped with 4 RTX-TITAN GPUs.

### 4.2. Experimental Results

The diagnosis accuracy on four binary classification tasks are present in Table 1. Both the competing method and our proposed method are tested with single modality settings and multimodality settings. The (sMRI) means the model is only trained with the structural connectivity matrices, so as (fMRI). The (sMRI+fMRI) means the model is utilizing both SC and FC modalities information to perform classification, which will present the full potential of the model. The results in Table 1. are produced by the best hyperparameter settings of each model with five-fold cross-validation applied to each task. The hyperparameter settings are discussed in the next section.

As shown in Table 1, both models work well on the AD vs. NC task. The performance gains from both modalities is very limited. This is because the AD subjects normally have significant structural brain damage and consequently experience cognitive losses, whereas the NC subjects normally do not have these problems. For the AD vs MCI and NC vs MCI tasks, both models’ performance has dropped. However, the proposed CMRN model still works better with both modalities. This may be because the proposed CMRN model has the ability to exchange modality information in a global way, whereas the ADB-NN model can only rely on the local topological information from neighborhoods and the attention-guided diffusion process. The ADB-NN works slightly better when using only one modality on the NC vs. MCI task. This may be because the brain damage is mild for the MCI subjects and most of them can still perform daily activities independently. Thus, there may not be significant global damages compared to NC subjects, and it would be better to look locally to find the most discriminative information. Both models’ performance significantly dropped on the MCIc vs. MCI<sub>n</sub> task,

**Table 1**  
Performance comparison of the competing methods on different tasks.

Methods	AD vs. NC	AD vs. MCI	NC vs. MCI	MCIc vs. MCI <sub>n</sub>	AD vs. MCIc vs. MCI <sub>n</sub> vs. NC
ADB-NN (sMRI)	92.31±7.73	85.78±8.29	86.81±9.74	73.14±9.21	83.41±5.93
ADB-NN (fMRI)	92.58±8.16	85.50±6.63	85.30±6.88	72.57±10.76	82.63±8.01
ADB-NN (sMRI + fMRI)	93.43±7.86	86.69±7.68	87.04±9.22	79.91±9.49	84.15±6.57
CMRN(sMRI)	92.39±7.41	86.98±10.33	85.92±8.57	78.38±6.31	84.09±7.14
CMRN (fMRI)	92.87±7.06	87.29±8.76	85.13±6.41	77.83±8.40	83.78±5.20
CMRN (sMRI + fMRI)	93.89±6.32	88.62±8.81	89.83±7.00	84.97±10.11	86.51±7.42

which is the most challenging one of the four tasks. However, our proposed model still works consistently better than the competing method. This may be because the proposed Cross-Modal Attention mechanism can help distinguish very similar modality information.

#### 4.3. Ablation Study of Cross-Modal Attention

Table 1 also demonstrated the effectiveness of the proposed Cross-Modal Attention (CMA) module. Without the CMA module to exchange information across modalities, the proposed model can only extract information from a single modality, sMRI or fMRI, which results in lower classification accuracies than the model with the CMA module. This performance-boosting with the CMA module is consistent across all five tasks. In the most difficult binary classification task MCIC vs. MCIN, this performance-boosting is more significant than the other three binary classification tasks. The significant performance-boosting is also observed in the multiple class classification task, which is a much harder task than the binary classification tasks. Furthermore, compared with the SC and FC coupling method in the ADB-NN model in the third line of Table 1, our proposed model achieves much higher overall accuracies across all five tasks, which further demonstrates the effectiveness of the CMA module.

#### 4.4. Starting Fusion Layer

One very important hyperparameter is which layer to start fusing multimodal information. To investigate its impacts, the proposed model is fixed with 8 Riemannian layers and 1 traditional softmax layer. And the model is tested on the most challenging task, MCIC vs. MCIN. When the starting fusion layer  $L_f = 0$ , it is the so-called early fusion strategy, whereas, when  $L_f = 8$ , it is called late fusion. Here,  $L_f$  controls how the modality information flows across the whole network. As shown in Fig. 4, the model achieves the best performance when  $L_f = 5$ , which indicates that it is a better strategy to let early layers learn the modality specialized features before mixing the cross-modal information. In this way, the model will have a chance to filter out the less useful information in each modality and avoid the noisy coupling information between modalities in the later layers.

#### 4.5. Model interpretation

To interpret how the model acts on the brain networks, a visualization of the weight activation on the first layer is performed. According to Eq. (4) the weight matrix  $W_{S_0} \in \mathbb{R}^{d_0 \times d_1}$ , ( $d_1 < d_0$ ) will compress the initial input matrix into a smaller dimension  $d_1$ . Thus,  $W_{S_0}^T W_{S_0}$  will get a  $d_0 \times d_0$  matrix as the initial input matrix, which can be interpreted as the importance of the corresponding edge in the initial matrix to the final classification. The diagonal elements of matrix  $W_{S_0}^T W_{S_0}$  can be considered as the importance of the

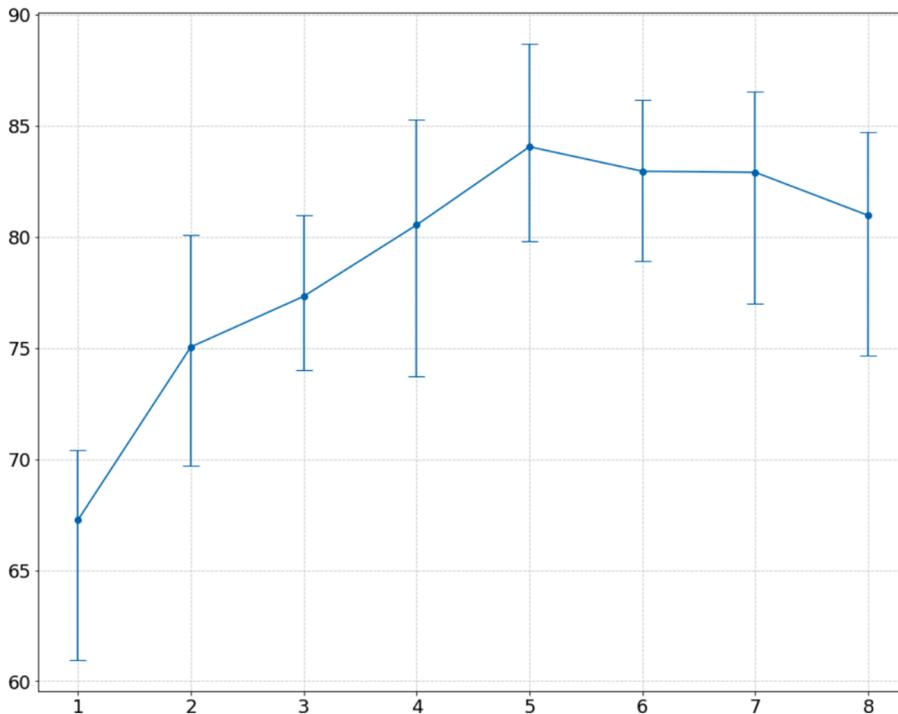


Fig. 4. Investigate the impact of starting fusion layer. The horizon axis is the starting fusion layer  $L_f$ , which means the CMA module will be applied into this layer and all the following layers. The vertical axis is the model accuracy. The markers indicate the average accuracy and the bars indicate the highest and lowest accuracy in the cross-validation folds.

corresponding brain regions. Thus, we first selected the top 15 nodes from the diagonal elements, which is about 10% of the total 148 nodes from the brain parcellation template. Then, from all the positive connections between these 15 nodes, we selected the top 25 edges from the upper triangular matrix of  $W_{S0}^T W_{S0}$ , since  $W_{S0}^T W_{S0}$  is a symmetric matrix. Finally, the selected top 15 nodes and the top 25 edges are projected back to the brain parcellation template and visualized in Fig 5. Similarly, we can interpret the  $d_0 \times d_0$  matrix  $W_{F0}^T W_{F0}$  from the functional modal, as illustrated in Fig 6.

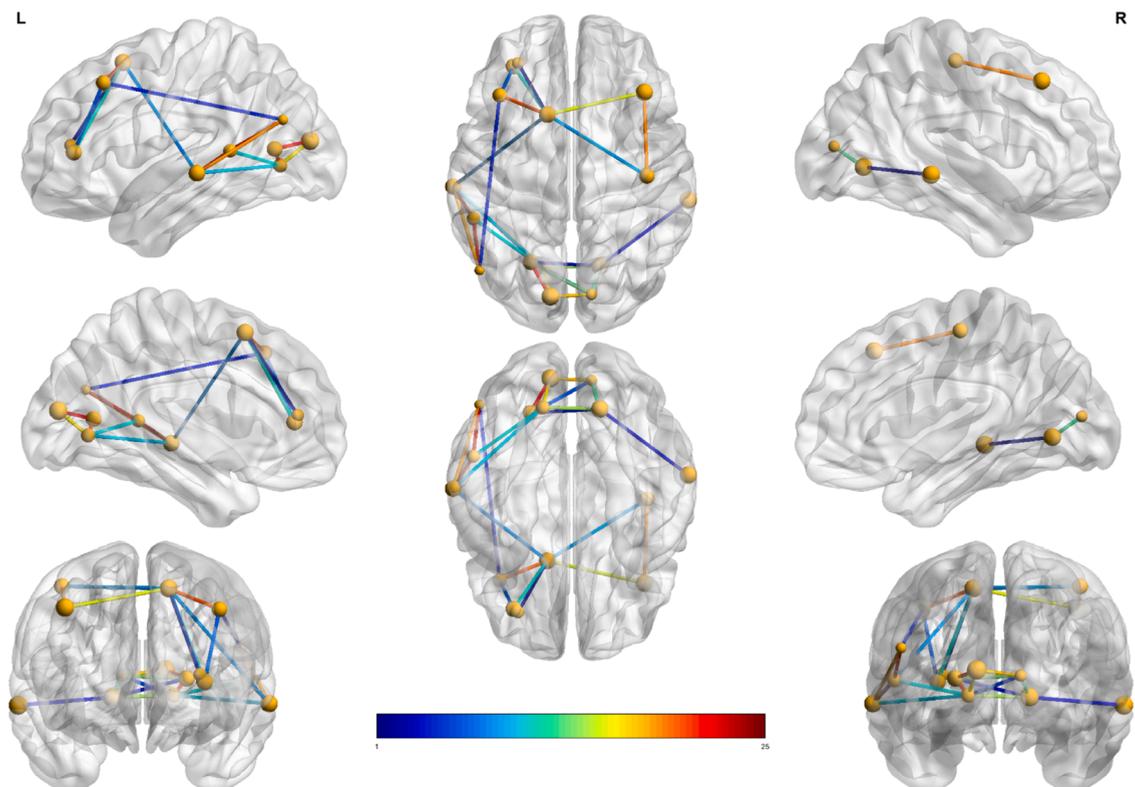
The region names of nodes in the brain parcellation can be found in the supplementary file. From Fig. 5 and Fig 6, we can see that the selected top nodes and top edges are all from the cognitive area and vision area, which align with the current medical findings in the AD diagnosis (Alzheimer’s Association 2021). Comparing Fig 5 and Fig 6, although the actual node connections and node importance are different between the two modalities, the selected nodes in the frontal lobes and occipital lobes are very similar to each other, such as parcellation labels ‘G\_cuneus’, ‘G\_front\_middle’, ‘G\_front\_sup’, and ‘G\_orbital’, which further demonstrates the coupling of the two modalities is successfully identified by our Cross-Modal Attention module.

## 5. Conclusion

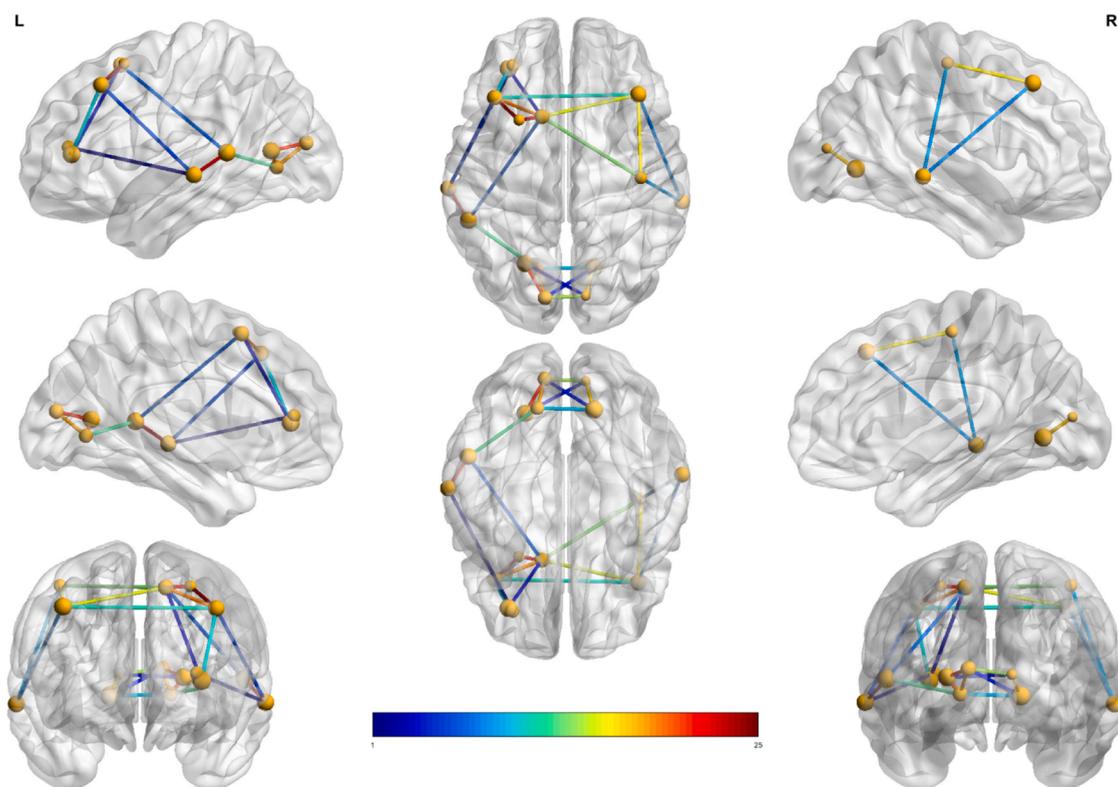
A novel Cross-Modal Riemannian Network (CMRN) is proposed in this paper to solve the multimodality coupling problem on the brain networks for Alzheimer’s Disease analysis. With respect to the global topological geometries, the proposed model works fully on the Riemannian Manifold with Riemannian operations to utilize the manifold properties. Moreover, the Cross-Modal Attention (CMA) mechanism is proposed to allow the proposed model to perform cross-modal interactions, which helps the model comprehensively utilize multimodal information and extract the most discriminative matrices. The proposed CMA also works fully on the Riemannian Manifold to cooperate with the proposed model and exploit the manifold properties. Furthermore, the CMA module can be plugged into any Riemannian layers to perform middle fusion, which is more flexible in determining when to share the cross-modal information. Experimental results on the Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset demonstrate the effectiveness and efficiency of the proposed method.

### CRedit authorship contribution statement

**Junbo Ma:** Conceptualization, Methodology, Validation, Writing – original draft, Visualization, Supervision. **Jilian Zhang:** Data curation, Writing – review & editing, Validation. **Zeyu Wang:** Writing – review & editing, Resources, Validation.



**Fig. 5.** The interpretation of the structural modality. The nodes represent the brain regions corresponding to the brain parcellation template. The size of the nodes represents the importance of the nodes: the bigger, the more important. The lines represent the selected top 25 edges. The color of the lines represents the importance of each edge. The color map is from 1 to 25, with 25 (dark red) showing the most important edge.



**Fig. 6.** The interpretation of the functional modality. The nodes represent the brain regions corresponding to the brain parcellation template. The size of the nodes represents the importance of the nodes: the bigger, the more important. The lines represent the selected top 25 edges. The color of the lines represents the importance of each edge. The color map is from 1 to 25, with 25 (dark red) showing the most important edge.

### Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.ipm.2022.102965](https://doi.org/10.1016/j.ipm.2022.102965).

### References

- Ahmed-Aristizabal, David, Armin, Mohammad Ali, Denman, Simon, Fookes, Clinton, & Petersson, Lars (2021). Graph-Based Deep Learning for Medical Diagnosis and Analysis: Past, Present and Future. *Sensors*, 21(14), 4758. <https://doi.org/10.3390/s21144758>
- Alzheimer's Association. (2021). 2021 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia*, 17(3), 327–406. <https://doi.org/10.1002/alz.12328>
- Arsigny, Vincent, Fillard, Pierre, Pennec, Xavier, & Ayache, Nicholas (2006). Log-Euclidean Metrics for Fast and Simple Calculus on Diffusion Tensors. *Magnetic Resonance in Medicine*, 56(2), 411–421. <https://doi.org/10.1002/mrm.20965>
- Brosch, Tom, & Tam, Roger (2013). Manifold Learning of Brain MRIs by Deep Learning. In K. Mori, I. Sakuma, Y. Sato, C. Barillot, & N. Navab (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013, Lecture Notes in Computer Science* (pp. 633–640). Berlin, Heidelberg: Springer.
- Chu, Shu-Hsien, Parhi, Keshab, K., & Lenglet, Christophe (2018). Function-specific and Enhanced Brain Structural Connectivity Mapping via Joint Modeling of Diffusion and Functional MRI. *Scientific Reports*, 8(4741). <https://doi.org/10.1038/s41598-018-23051-9>
- Dai, Mengyu, and Haibin Hang. 2021. "Manifold Matching via Deep Metric Learning for Generative Modeling." Pp. 6587–97 in.
- Defferrard, Michaël, Bresson, Xavier, & Vandergheynst, Pierre (2016). Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Vol. 29.
- Destrieux, Christophe, Fischl, Bruce, Dale, Anders, & Halgren, Eric (2010). Automatic Parcellation of Human Cortical Gyri and Sulci Using Standard Anatomical Nomenclature. *NeuroImage*, 53(1), 1–15. <https://doi.org/10.1016/j.neuroimage.2010.06.010>
- Gan, Jiangzhang, Peng, Ziwen, Zhu, Xiaofeng, Hu, Rongyao, Ma, Junbo, & Wu, Guorong (2021). Brain functional connectivity analysis based on multi-graph fusion. *Medical Image Analysis*, 71(102057). <https://doi.org/10.1016/j.media.2021.102057>
- Gonneaud, Julie, Baria, Alex T., Binette, Alexa Pichet, Gordon, Brian A., Chhatwal, Jasmeer P., Cruchaga, Carlos, Jucker, Mathias, Levin, Johannes, Salloway, Stephen, Farlow, Martin, Gauthier, Serge, Benzinger, Tammie L. S., Morris, John C., Bateman, Randall J., Breitner, John C. S., Poirier, Judes, Vachon-Preseau, Etienne, & Villeneuve, Sylvia (2021). Accelerated Functional Brain Aging in Pre-Clinical Familial Alzheimer's Disease. *Nature Communications*, 12(1), 5346. <https://doi.org/10.1038/s41467-021-25492-9>
- Gordon, Evan M., Laumann, Timothy O., Adeyemo, Babatunde, Huckins, Jeremy F., Kelley, William M., & Petersen, Steven E. (2016). Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral Cortex*, 26(1), 288–303. <https://doi.org/10.1093/cercor/bhu239>
- Hermessi, Haithem, Mourali, Olfa, & Zagrouba, Ezzeddine (2021). Multimodal Medical Image Fusion Review: Theoretical Background and Recent Advances. *Signal Processing*, 183, Article 108036. <https://doi.org/10.1016/j.sigpro.2021.108036>
- Huang, Zhiwu, & Gool, Luc Van (2017). A Riemannian Network for SPD Matrix Learning. In *Thirty-First AAAI Conference on Artificial Intelligence*.

- Huang, Jiashuang, Zhou, Luping, Wang, Lei, & Zhang, Daoqiang (2020). Attention-Diffusion-Bilinear Neural Network for Brain Network Analysis. *IEEE Transactions on Medical Imaging*, 39(7), 2541–2552. <https://doi.org/10.1109/TMI.2020.2973650>
- Huang, Zhuobin, Cai, Hongmin, Dan, Tingting, Lin, Yi, Laurienti, Paul, & Wu, Guorong (2021). Detecting Brain State Changes by Geometric Deep Learning of Functional Dynamics on Riemannian Manifold. In M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, & C. Essert (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, Lecture Notes in Computer Science* (pp. 543–552). Cham: Springer International Publishing.
- Jack Jr, Clifford R., Bernstein, Matt A., Fox, Nick C., Thompson, Paul, Alexander, Gene, Harvey, Danielle, Borowski, Bret, Britson, Paula J., Whitwell, Jennifer L., Ward, Chadwick, Dale, Anders M., Felmlee, Joel P., Gunter, Jeffrey L., Hill, Derek L. G., Killiany, Ron, Schuff, Norbert, Fox-Bosetti, Sabrina, Lin, Chen, Studholme, Colin, DeCarli, Charles S., Krueger, Gunnar, Ward, Heidi A., Metzger, Gregory J., Scott, Katherine T., Mallozzi, Richard, Blezek, Daniel, Levy, Joshua, Debbins, Josef P., Fleisher, Adam S., Albert, Marilyn, Green, Robert, Bartzokis, George, Glover, Gary, Mugler, John, & Weiner, Michael W. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods. *Journal of Magnetic Resonance Imaging*, 27(4), 685–691. <https://doi.org/10.1002/jmri.21049>
- Ke, Ziwen, Zhuo-Xu Cui, Wenqi Huang, Jing Cheng, Sen Jia, Haifeng Wang, Xin Liu, Hairong Zheng, Leslie Ying, Yanjie Zhu, and Dong Liang. 2021. "Deep Manifold Learning for Dynamic MR Imaging." *ArXiv:2104.01102 [Cs, Eess]*.
- Kim, Kwang-Rae, Dryden, Ian L., Le, Huiling, & Severn, Katie E. (2021). Smoothing Splines on Riemannian Manifolds, with Applications to 3D Shape Space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(1), 108–132. <https://doi.org/10.1111/rssb.12402>
- Kipf, Thomas N., & Welling, Max (2017). Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations (ICLR)*.
- Li, Wei, Hu, Huosheng, Tang, Chao, & Song, Yuping (2020). Manifold Multi-View Learning for Cartoon Alignment. *International Journal of Computer Applications in Technology*, 62(2), 91–101. <https://doi.org/10.1504/IJCAT.2020.104690>
- Li, Zhengying, Hong Huang, Zhen Zhang, and Yinsong Pan. 2021. "Manifold Learning-Based Semisupervised Neural Network for Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 1–12. doi: 10.1109/TGRS.2021.3083776.
- Liao, Guoqiong, Deng, Xiaobin, Wan, Changxuan, & Liu, Xiping (2022). Group Event Recommendation Based on Graph Multi-Head Attention Network Combining Explicit and Implicit Information. *Information Processing & Management*, 59(2), Article 102797. <https://doi.org/10.1016/j.ipm.2021.102797>
- Lin, Tong, & Zha, Hongbin (2008). Riemannian Manifold Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5), 796–809. <https://doi.org/10.1109/TPAMI.2007.70735>
- Ma, Junbo, Zhu, Xiaofeng, Yang, Defu, Chen, Jiazhou, & Wu, Guorong (2020). Attention-Guided Deep Graph Neural Network for Longitudinal Alzheimer's Disease Analysis. In A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, & L. Joskowicz (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, Lecture Notes in Computer Science* (pp. 387–396). Cham: Springer International Publishing.
- Miolane, Nina, Alice Le Brigant, Johan Mathe, Benjamin Hou, Nicolas Guigui, Yann Thanwerdas, Stefan Heyder, Olivier Peltre, Niklas Koep, Hadi Zaatiti, Hatem Hajri, Yann Cabanes, Thomas Gerald, Paul Chauchat, Christian Shewmake, Bernhard Kainz, Claire Donnat, Susan Holmes, and Xavier Pennec. 2020. "Geomstats: A Python Package for Riemannian Geometry in Machine Learning." *ArXiv:2004.04667 [Cs]*.
- Nagrani, Arsha, Shan Yang, Anurag Arnab, Aren Jansen, Cordelia Schmid, and Chen Sun. 2021. "Attention Bottlenecks for Multimodal Fusion." *ArXiv:2107.00135 [Cs]*.
- Paszke, Adam, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. "Automatic Differentiation in PyTorch."
- Peng, Liang, Rongyao Hu, Fei Kong, Jiangzhang Gan, Yujie Mo, Xiaoshuang Shi, and Xiaofeng Zhu. 2022. "Reverse Graph Learning for Graph Neural Network." *IEEE Transactions on Neural Networks and Learning Systems* PP. doi: 10.1109/tnnls.2022.3161030.
- Pennec, Xavier, Fillard, Pierre, & Ayache, Nicholas (2006). A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision*, 66(1), 41–66. <https://doi.org/10.1007/s11263-005-3222-z>
- Pennec, Xavier. (2006). Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal of Mathematical Imaging and Vision*, 25(1), 127. <https://doi.org/10.1007/s10851-006-6228-4>
- Pisano, Thomas J., Dhanerawala, Zahra M., Kislin, Mikhail, Bakshinskaya, Dariya, Engel, Esteban A., Hansen, Ethan J., Hoag, Austin T., Lee, Junuk, Oude, Nina L.de, Venkataraju, Kannan Umadevi, Verpeut, Jessica L., Hoebeek, Freek E., Richardson, Ben D., Boele, Henk-Jan, & Wang, Samuel S. H. (2021). Homologous Organization of Cerebellar Pathways to Sensory, Motor, and Associative Forebrain. *Cell Reports*, 36(12). <https://doi.org/10.1016/j.celrep.2021.109721>
- Sreedevi, A. G., Nitya Harshitha, T., Sugumaran, Vijayan, & Shankar, P. (2022). Application of Cognitive Computing in Healthcare, Cybersecurity, Big Data and IoT: A Literature Review. *Information Processing & Management*, 59(2), Article 102888. <https://doi.org/10.1016/j.ipm.2022.102888>
- Swati, Shipra, Kumar, Mukesh, & Namasudra, Suyel (2022). Early Prediction of Cognitive Impairments Using Physiological Signal for Enhanced Socioeconomic Status. *Information Processing & Management*, 59(2), Article 102845. <https://doi.org/10.1016/j.ipm.2021.102845>
- Tanveer, M., Richhariya, B., Khan, R. U., Rashid, A. H., Khanna, P., Prasad, M., & Lin, C. T. (2020). Machine Learning Techniques for the Diagnosis of Alzheimer's Disease: A Review. *ACM Trans. Multimedia Comput. Commun. Appl.*, 16(1s). <https://doi.org/10.1145/3344998>
- Uppal, Shagun, Bhagat, Sarthak, Hazarika, Devamanyu, Majumder, Navonil, Poria, Soujanya, Zimmermann, Roger, & Zadeh, Amir (2022). Multimodal Research in Vision and Language: A Review of Current and Emerging Trends. *Information Fusion*, 77, 149–171. <https://doi.org/10.1016/j.inffus.2021.07.009>
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." *ArXiv:1706.03762 [Cs]*.
- Veitch, Dallas P., Weiner, Michael W., Aisen, Paul S., Beckett, Laurel A., Cairns, Nigel J., Green, Robert C., Harvey, Danielle, Jack, Clifford R., Jagust, William, Morris, John C., Petersen, Ronald C., Saykin, Andrew J., Shaw, Leslie M., Toga, Arthur W., & Trojanowski, John Q. (2019). Understanding Disease Progression and Improving Alzheimer's Disease Clinical Trials: Recent Highlights from the Alzheimer's Disease Neuroimaging Initiative. *Alzheimer's & Dementia*, 15(1), 106–152. <https://doi.org/10.1016/j.jalz.2018.08.005>
- Velickovic, Petar, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2018. "GRAPH ATTENTION NETWORKS." 12.
- Wang, Lujing, Yuan, Weifeng, Zeng, Lu, Xu, Jie, Mo, Yujie, Zhao, Xinxiang, & Peng, Liang (2022). Dementia Analysis from Functional Connectivity Network with Graph Neural Networks. *Information Processing & Management*, 59(3), Article 102901. <https://doi.org/10.1016/j.ipm.2022.102901>
- Wu, Yongji, Defu Lian, Yiheng Xu, Le Wu, and Enhong Chen. 2020. "Graph Convolutional Networks with Markov Random Field Reasoning for Social Spammer Detection." *Proceedings of the AAAI Conference on Artificial Intelligence*34(01):1054–61. doi: 10.1609/aaai.v34i01.5455.
- Yu, Xiang, Wang, Shui-Hua, & Zhang, Yu-Dong (2021). CGNet: A Graph-Knowledge Embedded Convolutional Neural Network for Detection of Pneumonia. *Information Processing & Management*, 58(1), Article 102411. <https://doi.org/10.1016/j.ipm.2020.102411>
- Yuan, Changan, Zhong, Zhi, Lei, Cong, Zhu, Xiaofeng, & Hu, Rongyao (2021). Adaptive Reverse Graph Learning for Robust Subspace Learning. *Information Processing & Management*, 58(6), Article 102733. <https://doi.org/10.1016/j.ipm.2021.102733>
- Zhang, Yifei, Sidibé, Désiré, Morel, Olivier, & Mériaudeau, Fabrice (2021). Deep Multimodal Fusion for Semantic Image Segmentation: A Survey. *Image and Vision Computing*, 105, Article 104042. <https://doi.org/10.1016/j.imavis.2020.104042>
- Zhang, Zhao, Zhang, Yan, Xu, Mingliang, Zhang, Li, Yang, Yi, & Yan, Shuicheng (2021). A Survey on Concept Factorization: From Shallow to Deep Representation Learning. *Information Processing & Management*, 58(3), Article 102534. <https://doi.org/10.1016/j.ipm.2021.102534>
- Zhou, Jie, Cui, Gangu, Hu, Shengding, Zhang, Zhengyan, Yang, Cheng, Liu, Zhiyuan, Wang, Lifeng, Li, Changcheng, & Sun, Maosong (2020). Graph Neural Networks: A Review of Methods and Applications. *AI Open*, 1, 57–81. <https://doi.org/10.1016/j.aiopen.2021.01.001>
- Zhu, Xiaofeng, Li, Xuelong, Zhang, Shichao, Ju, Chunhua, & Wu, Xindong (2017). Robust Joint Graph Sparse Coding for Unsupervised Spectral Feature Selection. *IEEE Transactions on Neural Networks and Learning Systems*, 28(6), 1263–1275. <https://doi.org/10.1109/TNNLS.2016.2521602>
- Zhu, Wenhao, Liu, Shuang, & Liu, Chaoming (2021). Learning Multimodal Word Representation with Graph Convolutional Networks. *Information Processing & Management*, 58(6), Article 102709. <https://doi.org/10.1016/j.ipm.2021.102709>
- Zhu, Yonghua, Ma, Junbo, Yuan, Changan, & Zhu, Xiaofeng (2022). Interpretable Learning Based Dynamic Graph Convolutional Networks for Alzheimer's Disease Analysis. *Information Fusion*, 77, 53–61. <https://doi.org/10.1016/j.inffus.2021.07.013>