Brain Network Classification for Accurate Detection of Alzheimer's Disease via Manifold Harmonic Discriminant Analysis

Hongmin Cai, *Senior Member, IEEE*, Xiaoqi Sheng, Guorong Wu, *Member, IEEE*, Bin Hu, *Fellow, IEEE*, Yiu-Ming Cheung, *Fellow, IEEE*, and Jiazhou Chen, *Member, IEEE*

***Abstract*— Mounting evidence shows that Alzheimer's disease (AD) manifests the dysfunction of the brain network much earlier before the onset of clinical symptoms, making its early diagnosis possible. Current brain network analyses treat high-dimensional network data as a regular matrix or vector, which destroys the essential network topology, thereby seriously affecting diagnosis accuracy. In this context, harmonic waves provide a solid theoretical background for exploring brain network topology. However, the harmonic waves are originally intended to discover neurological disease propagation patterns in the brain, which makes it difficult to accommodate brain disease diagnosis with high heterogeneity. To address this challenge, this article proposes a network manifold harmonic discriminant analysis (MHDA) method for accurately detecting AD. Each brain network is regarded as an instance drawn on a Stiefel manifold. Every instance is represented by a set of orthonormal eigenvectors (i.e., harmonic waves) derived from its Laplacian matrix, which fully respects the topological structure of the brain network. An MHDA method within the Stiefel space is proposed to identify the group-dependent common harmonic waves, which can be used as group-specific references for downstream analyses. Extensive experiments are conducted to demonstrate the effectiveness of the proposed method in stratifying cognitively normal (CN) controls, mild cognitive impairment (MCI), and AD.**

***Index Terms*— Alzheimer's disease (AD), brain network, classification, manifold learning.**

Manuscript received 26 December 2022; revised 24 May 2023; accepted 28 July 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFE0112200; in part by the National Natural Science Foundation of China under Grant 62325204, Grant U21A20520, Grant 62102153, and Grant 62172112; in part by the Science and Technology Project of Guangdong Province under Grant 2022A0505050014; in part by the Key-Area Research and Development Program of Guangzhou City under Grant 202206030009; in part by the Natural Science Foundation of Guangdong Province of China under Grant 2022A1515011162; in part by the China Postdoctoral Science Foundation under Grant 2021M691062 and Grant 2023T160226; and in part by the Alzheimer's Disease Neuroimaging Initiative (ADNI). *(Hongmin Cai and Xiaoqi Sheng contributed equally to this work.) (Corresponding author: Jiazhou Chen.)*

Hongmin Cai, Xiaoqi Sheng, and Jiazhou Chen are with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China (e-mail: csjzchen@scut.edu.cn).

Guorong Wu is with the Department of Psychiatry and Computer Science, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA.

Bin Hu is with the School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China (e-mail: bh@lzu.edu.cn).

Yiu-Ming Cheung is with the Department of Computer Science, Hong Kong Baptist University, Hong Kong, China.

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TNNLS.2023.3301456.

Digital Object Identifier 10.1109/TNNLS.2023.3301456

I. Introduction

ALZHEIMER's disease (AD) is a neurodegenerative disease in the human brain that is often caused by atrophy of the nerve area of brain tissue [1]. In clinical diagnosis, AD patients usually suffer from memory upset, and behavioral and cognitive decline, which seriously affect their quality of life. Broadly, AD is divided into three phases according to disease progression: cognitively normal (CN) controls, mild cognitive impairment (MCI), and AD. MCI is an early stage of AD and is characterized by obvious cognitive impairment without complete cognitive ability loss. Converging evidence shows that MCI has a high risk of turning into AD [2]. The accurate diagnosis of the MCI stage is a crucial step in alleviating the development of AD, because medication might help delay the symptoms during the MCI stage [3], [4]. Patients in the primary stage should be appropriately treated to slow the progression of the disease. However, changes in brain function and anatomical structure in the early stage of AD are subtle [5], making the early diagnosis of AD challenging.

With the rapid development of noninvasive neuroimaging techniques, it becomes possible to capture multimodal brain images from the same sample, which provides an efficient and feasible method for investigating structural brain connectivity [6]. For example, combining diffusion-weighted magnetic resonance imaging (DW-MRI) and magnetic tractography (MEG) techniques offers a noninvasive window to reconstruct the major fiber tracts in the brain and visualize the structural spreading pathways that connect different brain regions, thus helping us better understand the neuropathological propagation underpinnings of neurological disorders [7]. Recent neuroscience research has shown that interactions between brain regions are a critical driver in the analysis of neurodegenerative diseases [8], [9]. Inspired by graph theory, brain networks can be composed of nodes and edges which are used to represent the interactions between brain regions. The development of graph research methods allows us to quantify the complex topological features of brain networks and detect network modularity alterations. Although identifying brain network alterations is beneficial to the early diagnosis of AD, designing a novel brain network analysis method is a difficult problem with the brain network being a high-dimensional irregular data structure.

Recently, a growing number of brain network-based methods have been proposed to study neurological disorders. They can be divided into two categories: 1) studies on specific network models, such as small-world networks [10] and default mode networks (DMNs) [11], and 2) studies on brain network classification based on machine learning methods [12], [13]. For the first category, the backbone of these studies focus on quantifying the topological differences between AD and CN brain networks by applying graph-theory analysis methods to specific brain networks. Although these studies validate the hypothesis of disconnected and abnormal syndromes in AD and MCI patients, brain network alterations in early AD are not restricted to specific brain subnetworks, making it difficult to accurately distinguish between AD and CN in individuals. Regarding the second category, most data-driven machine learning methods are employed to train a classification model to identify neurological disorders. These machine learning methods first extract the most discriminative subset of features from the brain network as initial features. Then, such discriminative features are directly used as feature vectors for subsequent disease diagnosis tasks. Notably, unlike traditional data represented in feature space, vectorization of irregular data structures destroys the essential topological structure properties. Motivated by the above analysis, most studies on structural brain networks focus on specific subnetwork representations and feature learning steps to analyze the differences between structural brain networks, which leads to the feature selection problem in network classification. Recent efforts to investigate the topological properties of brain networks have focused on Grassmann manifolds [14] or Stiefel manifolds [15]. However, most manifold methods explore only connectivity alterations in brain networks and lack discriminative representation. Recently, graph neural networks (GNNs) [16] have become a hot topic of research in analyzing irregular brain networks owing to their powerful ability to analyze graph-structured data. However, brain networks are different from other graphs in the real world, so GNN design needs to be customized to follow the unique nature of brain network data [17]. In addition, GNN-based methods have several major limitations, including poor globality and easy oversmoothing.

One possible solution is to define a novel mathematical representation of the brain network that has sufficient statistical power to: 1) represent the topology profile of the brain network; 2) provide rigorous mathematical characterization to facilitate reasoning; and 3) possess strong discriminative properties for addressing heterogeneity across brain networks. As mentioned earlier, the structural brain network constructed by diffusion MRI is inherently a spatially embedded network with a high-dimensional complex topology [18]. Emerging studies have confirmed that such a complex topological structure in a brain network can be expressed by the eigensystem of its underlying Laplacian matrix, such as modules and hubs [14]. The eigenvectors of the Laplacian matrix constitute the unique bases of the topological structure in the underlying network, which are introduced in detail in Section II-C. Moreover, recent studies have found that the eigenvectors of the Laplacian matrix are a set of orthonormal bases, named harmonic waves, which can be used to characterize the propagation patterns of neuropathological events [19]. Therefore, such eigenvectors behind individual brain networks reside on a high-dimensional Stiefel manifold. In this context, harmonic-based analyses have been employed for brain network analysis based on frequency alterations on the Stiefel manifold [20]. However, these methods focus on analyzing the spreading pathways of neuropathological burdens in neurodegenerative diseases and have not paid sufficient attention to the discrimination ability to enhance features, which in turn leads to poor classification results in terms of the high heterogeneity across brain networks.

To overcome the limitations of the existing brain network analysis methods, by endowing the brain network topology with the well-studied Stiefel manifold, a novel manifold harmonic discriminant analysis (MHDA) method is proposed to stratify CN, MCI, and AD. The backbone of the proposed method applies harmonic waves to approximate the self-organized oscillation patterns of each brain network. It estimates the discriminative group-dependent common harmonic waves on the Stiefel manifold as the group-specific centers for downstream analyses. The MHDA workflow is illustrated in Fig. 1. An adjacency matrix is first constructed for each brain network, and then, its harmonic waves are captured by the eigendecomposition of the Laplacian matrix, as shown in Fig. 1(a). After that, the linear discriminant analysis of Euclidean space is extended to the Stiefel manifold through a series of manifold algebra operations. The group-dependent common harmonic waves with discriminant information are evaluated by minimizing the intragroup and intergroup geodesic distances of individual harmonic waves, as shown in Fig. 1(b). Finally, the test sample label is the one with the shortest geodesic distance with respect to the manifold center, as illustrated in Fig. 1(c). In the literature, several methods [21], [22] have been proposed exploiting the advantages of high-dimensional matrix-valued data for the classification task. These approaches follow the traditional Fisher's linear discriminant analysis (FLDA) to project high-dimensional data to the low-dimensional space to enhance the classification performance. However, this dimensionality reduction process is not appropriate for brain network analysis, as it would destroy the harmonic wave properties across the brain network. Instead, in the proposed approach, group-dependent common harmonic waves are learned on Stiefel manifolds to solve the problem of projection matrices. In addition, two optimization methods based on the Stiefel manifold are proposed to address the computational complexity of MHDA and multigroup common harmonic waves optimization challenges. As a result, this article proposes that the brain network analysis method is built upon two sets of Riemannian manifold optimization, making the model more flexible for high-dimensional brain network analysis problems. We evaluate the proposed method on 155 structural brain networks collected from the ADNI dataset, which contains 55 CN, 44 MCI, and 56 AD subjects. The experiments confirm the effectiveness of the proposed method for diagnosing brain diseases and identifying the regions of the brain responsible for causing these diseases. Furthermore, the scalability and complexity analyses demonstrate that the

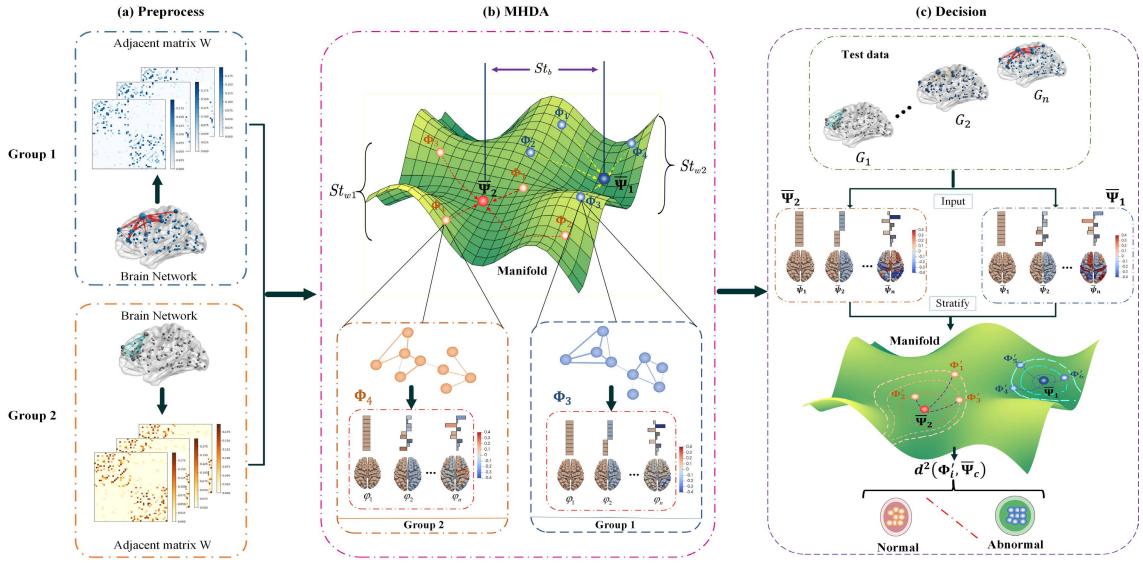

Fig. 1. Overview of the MHDA method in brain network classification tasks. (a) Brain networks are taken as the input, and a set of adjacency matrices is generated (preprocess). (b) Then, the group-dependent common harmonic waves are obtained by MHDA, where $\mathbf{St}_{w1}$ and $\mathbf{St}_{w2}$ represent the intragroup geodesic distance and $\mathbf{St}_b$ represents the intergroup geodesic distance learned over all individual harmonic waves located on the Stiefel manifold. (c) Finally, the learned group-dependent common harmonic waves are used for brain network classification (decision), where the contour lines indicate the geodesic distance from the individual harmonic waves to the group-dependent common harmonic waves.

algorithm is applicable to high-dimensional brain networks of varying scales.

The main contributions of our MHDA method are threefold.

1) To fully consider the topology of the brain network, a novel MHDA method is proposed to classify each brain network into subtypes. MHDA employs the group-dependent harmonic waves representation strategy to explore the topological differences among brain networks in group comparisons, thereby achieving accurate classification of brain diseases. Compared with our previous manifold analysis method [20], MHDA primarily identifies group-dependent harmonic waves for the classification of brain networks and analysis of topological differences between groups.
2) The changes in brain networks between different groups (such as CN and MCI) are subtle, which causes large intraclass diversity and interclass similarity. To this end, a tailored harmonic-based Fisher's criterion is presented that not only maintains the topology of brain networks but also alleviates the heterogeneity of brain networks. Compared with classical manifold-based discriminant analysis [21], MHDA mainly enhances the discriminative power of brain networks in the original Stiefel manifold space rather than in the low-dimensional Euclidean space, which is more conducive to clinical interpretability in neuroscience.
3) Two optimization algorithms are proposed to solve parameter matrices with orthogonal constraints. The first algorithm provides higher accuracy, but at the expense of increased computational complexity. Alternatively, the second algorithm trades off a small amount of accuracy to achieve improvement of computational efficiency. Extensive experiments on both synthetic and real datasets were conducted to demonstrate that the proposed method has superior classification performance compared to the state-of-the-art approaches. Furthermore, the flexibility and adaptability of the two Riemannian optimization methods in analyzing high- or low-dimensional manifold-valued data from different scenarios has been discussed.

The rest of this article is organized as follows. The theoretical background associated with the study in Section II is briefly reviewed. Then, a detailed description of the MHDA model and its numerical optimization scheme is also provided in Section III. Experiments are conducted to test the performance of the MHDA in Section IV. In Section V, the proposed MHDA model, including discovering important brain regions and limitations, is further discussed. Finally, concluding remarks are provided in Section VI.

II. Notation and Related Works

A. Notation

Throughout this article, a matrix $\boldsymbol{M} \in \mathbb{R}^{n\times p}$ is written as a boldface capital letter, and the ith row vector and the (ij)th element in matrix $\boldsymbol{M}$ are written as boldface lowercase letters $\boldsymbol{m}_i$ and lowercase letters m_{ij}, respectively. The identity matrix is denoted $\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$. $\mathbf{0}_p \in \mathbb{R}^{p\times p}$ denotes the zero matrix. The notation Δ denotes the tangent vector on tangent plane $\mathcal{T}_{\boldsymbol{M}}$ generated on point $\boldsymbol{M}$ in Stiefel manifold space $\mathrm{St}(n, p)$. In addition, $\mathrm{Tr}(\cdot)$ is the trace operation, and $\exp_{\Delta}(\cdot)$ represents the exponential map operation projecting tangent Δ back onto the Stiefel manifold. We use $F_{\boldsymbol{M}}$ to denote the matrix derivative of function F with respect to $\boldsymbol{M}$ and $\nabla_{\boldsymbol{M}} F$ to denote the gradient of F at point $\boldsymbol{M}$, where F is a smooth real-valued function defined on the Stiefel manifold.

B. Stiefel Manifold

The Stiefel manifold $\mathrm{St}(n, p)$ consists of an ordered set of orthonormal n-by-p matrices. Specifically, any matrix (i.e., individual harmonic waves) $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ can be considered a point on the Stiefel manifold if it satisfies $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{I}_p$. In exceptional cases, $\mathrm{St}(n, p)$ can be embedded into the np-dimensional Euclidean space of the n-by-p dimensional matrix [23]. Therefore, the effective approximation to the geodesic distance between $\boldsymbol{X} \in \mathrm{St}(n, p)$ and $\boldsymbol{Y} \in \mathrm{St}(n, p)$ can be defined as follows:

$$d^2(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{2}\mathrm{Tr}\,(\boldsymbol{X}-\boldsymbol{Y})^T(\boldsymbol{X}-\boldsymbol{Y}) = p - \mathrm{Tr}(\boldsymbol{X}^T\boldsymbol{Y}). \quad (1)$$

For a smooth real-valued function F defined on the Stiefel manifold, the gradient of function F at point $\boldsymbol{X} \in \mathrm{St}(n, p)$ is defined as the tangent vector on the tangent space, i.e., $\nabla_{\boldsymbol{X}}F \in \mathcal{T}_{\boldsymbol{X}}F$, where the tangent space $\mathcal{T}_{\boldsymbol{X}}F$ is composed of a set of tangent vectors at $\boldsymbol{X}$. $\nabla_{\boldsymbol{X}}F$ can be defined by

$$\nabla_{\boldsymbol{X}}F = F_{\boldsymbol{X}} - \boldsymbol{X}F_{\boldsymbol{X}}^T\boldsymbol{X} \quad (2)$$

where $F_{\boldsymbol{X}}$ denotes the n-by-p derivative matrix of F with respect to $\boldsymbol{X}$ on the Stiefel manifold. Given a tangent vector $\Delta \in \mathcal{T}_{\boldsymbol{X}}$, directly calculating the intrinsic geodesic distance on the Stiefel manifold emanating from $\boldsymbol{X}$ in direction Δ is computationally complex. Therefore, the matrix exponential method is applied to alleviate the computational challenge [23]. First, the matrices $\boldsymbol{Q} \in \mathbb{R}^{n\times p}$ and $\boldsymbol{R} \in \mathbb{R}^{p\times p}$ are obtained by compact QR decomposition of $(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^T)\Delta$. Therein, $\boldsymbol{Q}$ and $\boldsymbol{R}$ are the orthonormal matrix and upper triangular matrices, respectively. After that, the tangent vector is mapped to the Stiefel manifold through exponential retraction

$$\exp_{\boldsymbol{X}}(\Delta) = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{Q}\boldsymbol{C} \quad (3)$$

with $\boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{p\times p}$ given by

$$\begin{bmatrix}\boldsymbol{B}\\ \boldsymbol{C}\end{bmatrix} = \exp\left(\begin{bmatrix}\boldsymbol{A} & -\boldsymbol{R}^T\\ \boldsymbol{R} & \boldsymbol{0}_p\end{bmatrix}\right)\begin{bmatrix}\boldsymbol{I}_p\\ \boldsymbol{0}_p\end{bmatrix} \quad (4)$$

where $\boldsymbol{A} = \boldsymbol{X}^T\Delta \in \mathbb{R}^{p\times p}$. Equation (4) can be easily computed by solving a $2p$-by-$2p$ skew-symmetric eigenvalue problem using tailored singular value decomposition (SVD) algorithm [23].

The manifold algebra mentioned above provides an effective solution for measuring the difference between individual harmonic waves, thereby facilitating the search for discriminative group-dependent common harmonic waves, which are used for the subsequent classification task.

C. Harmonic Waves Transform on Brain Network

At the macroscale, brain networks are represented by graphs, where each node represents a specific brain region, and the edges represent the connectivity between a pair of nodes. More specifically, given a brain network encoded as a graph $G = (V, \mathcal{E}, \boldsymbol{W})$, it consists of a finite set of nodes $V = \{v_i | i \in 1, \ldots, n\}$ with n nodes and $\mathcal{E} = \{e_{ij}|(v_i, v_j) \in V \times V\}$ representing all possible connections between nodes. $\boldsymbol{W} \in \mathbb{R}^{n\times n}$ is an adjacency matrix with positive weights, where each element $w_{ij} \in \boldsymbol{W}$ represents the connection strength between nodes v_i and v_j. Then, the graph Laplacian matrix $\boldsymbol{L} \in \mathbb{R}^{n\times n}$ is calculated by

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W} \quad (5)$$

where $\boldsymbol{D} = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ is a degree matrix of $\boldsymbol{W}$. Each diagonal element equals the total connectivity degree of the underlying node, i.e., $d_i = \sum_{j=1}^{n} w_{ij}$.

The eigenvectors of the Laplacian matrix are obtained from the close-form solution to the following objective function:

$$\min_{\boldsymbol{\Phi}\in\mathbb{R}^{n\times p}} \mathrm{Tr}(\boldsymbol{\Phi}^T\boldsymbol{L}\boldsymbol{\Phi}), \quad \text{s.t. } \boldsymbol{\Phi}^T\boldsymbol{\Phi} = \boldsymbol{I}_p. \quad (6)$$

In (6), the eigenvectors in $\boldsymbol{\Phi} = [\phi_i]_{i=1}^{p}$ are associated with eigenvalues, so they are sorted according to their eigenvalues in an ascending order. As the eigenvalue increases, the associated eigenvector contains a large amount of high-frequency noise and oscillates violently. Additionally, brain network connectivity is usually better expressed in the low-frequency feature space of the graph spectrum. As a result, for a brain network with n nodes, the first p column ($p < n$) low-frequency eigenvectors are selected to characterize the topology of the brain network.

In the spatial domain, the brain network is a type of irregular graph structure with specific topological characteristics, such as hub nodes, small worlds, and hierarchical modality. The brain network organization is dominated by the Laplacian matrix of each network [14], and its underlying eigenvectors have been widely employed as graph embeddings [19], where each element of the eigenvector characterizes the topological profile of a specific node in the graph. In the frequency domain, a harmonic wave is a spatial expansion on the Fourier basis linked to the brain network structure, with an increasing spatial frequency, which in turn fits well with the anatomical structure of the brain. Each harmonic wave is defined as a self-organized propagation pattern of brain energy [19]. Mathematically, harmonic waves are a fundamental problem in geometric analysis and can be considered points on the Stiefel manifold due to orthogonality [24]. Therefore, harmonic waves not only establish a relation between the spatial structure of brain networks and frequency oscillations but also provide a new window for differential geometric understanding of brain anatomy.

III. Proposed Method

In this section, by exploring the brain network topology using the Stiefel manifold with Fisher's criterion, an MHDA method is presented in Section III-A, followed by the optimization scheme presented in Section III-B. Therein, the MHDA input is the eigenvectors (i.e., individual harmonic waves) obtained from the eigendecomposition of the Laplacian matrix. The goal of MHDA is to identify two sets of group-average eigenvectors from two different populations, where each set of group-average eigenvectors contains group-specific information about the intrinsic topology of the brain network. The derived group-average eigenvectors are the group-dependent common harmonic waves located on the Stiefel manifold, which can be regarded as the harmonic bases of the underlying group-specific brain network template. Besides, group-dependent common harmonic waves

can be used to alleviate the challenge of brain network classification.

A. Manifold Harmonic Discriminant Analysis

Given a set of Laplacian matrices $\boldsymbol{L} = \{(\boldsymbol{L}_1, y_1), (\boldsymbol{L}_2, y_2), \ldots, (\boldsymbol{L}_N, y_N)\}$ with $\boldsymbol{L}_i \in \mathbb{R}^{n\times n}$ and $y_i \in \{1, 2\}$, each individual harmonic waves $\boldsymbol{\Phi}_i \in \mathbb{R}^{n\times p}$ can be obtained by the eigendecomposition of Laplacian matrix $\boldsymbol{L}_i$. C_j $(j = 1, 2)$ is defined as the set of brain networks from the jth group. The goal of MHDA is to learn the group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_j \in \mathbb{R}^{n\times p}$ from all individual harmonic waves $\{\boldsymbol{\Phi}_i\}$ on the Stiefel manifold. Therefore, the geodesic distance between the two group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_j$ $(j = 1, 2)$ is maximized while minimizing the geodesic distance between the group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_j$ and the individual harmonic waves $\boldsymbol{\Phi}_i \in C_j$ within each group by (7). The intragroup distance St_w and intergroup distance St_b on the Stiefel manifold are defined as

$$\begin{aligned}\mathrm{St}_b &= \frac{1}{2}\mathrm{Tr}\left(\overline{\boldsymbol{\Psi}}_1 - \overline{\boldsymbol{\Psi}}_2\right)^T\left(\overline{\boldsymbol{\Psi}}_1 - \overline{\boldsymbol{\Psi}}_2\right) = p - \mathrm{Tr}\left(\overline{\boldsymbol{\Psi}}_1^T\overline{\boldsymbol{\Psi}}_2\right)\\ \mathrm{St}_w &= \frac{1}{2}\sum_{j=1}^{2}\sum_{\boldsymbol{\Phi}_i\in C_j}\mathrm{Tr}\left(\boldsymbol{\Phi}_i - \overline{\boldsymbol{\Psi}}_j\right)^T\left(\boldsymbol{\Phi}_i - \overline{\boldsymbol{\Psi}}_j\right)\\ &= \sum_{j=1}^{2}\sum_{\boldsymbol{\Phi}_i\in C_j}\left(p - \mathrm{Tr}\left(\boldsymbol{\Phi}_i{}^T\overline{\boldsymbol{\Psi}}_j\right)\right).\end{aligned} \tag{7}$$

Generally, traditional FLDA solves (7) with a trace ratio problem of the form $\max_{\boldsymbol{M}} \mathrm{Tr}(\boldsymbol{M}^T\mathrm{St}_b\boldsymbol{M})/\mathrm{Tr}(\boldsymbol{M}^T\mathrm{St}_w\boldsymbol{M})$, where $\boldsymbol{M}$ is the transformation matrix. However, the denominator causes instability while searching for a transformation matrix. Alternatively, the trace difference [25] $\max_{\boldsymbol{M}} \mathrm{Tr}[\boldsymbol{M}^T(\mathrm{St}_b - \lambda\mathrm{St}_w)\boldsymbol{M}]$ can be maximized. As a result, adhering to the idea of FLDA for trace difference problems, (7) is constituted as a discriminant analysis problem on the manifold, as follows:

$$\min_{\{\boldsymbol{\Phi}_i\},\overline{\boldsymbol{\Psi}}_j} St_w - St_b. \tag{8}$$

In the process of minimization, each instance $\boldsymbol{\Phi}_i$ is adjusted toward the potential manifold center $\overline{\boldsymbol{\Psi}}_j$. Once $\overline{\boldsymbol{\Psi}}_j$ is evaluated, it can be utilized to guide the refinement of $\boldsymbol{\Phi}_i$. The final results obtained are group-dependent common harmonic waves located on the Stiefel manifold, which can be considered harmonic bases of the underlying population-specific brain network templates. The complete solution procedure is provided in Section III-B. Unlike FLDA, the low-dimensional projection matrix $\boldsymbol{M}$ is set as an identity matrix, indicating that the individual harmonic waves do not require any transformation operation. Thus, the proposed method respects the geometric relationships within individual brain networks while avoiding the disruption of nonlinear structures through low-dimensional mapping.

The construction of individual harmonic waves is influenced not only by the underlying Laplacian matrix but also by the attraction of the group-dependent common harmonic waves at the center of the manifold. Therefore, by combining (6) on the Stiefel manifold and (8), the final objective function is modeled by the following:

$$\begin{aligned}\min_{\{\boldsymbol{\Phi}_i\},\overline{\boldsymbol{\Psi}}_j} &\sum_{i=1}^{N}\mathrm{Tr}(\boldsymbol{\Phi}_i^T\boldsymbol{L}_i\boldsymbol{\Phi}_i)\\ &+\sum_{j=1}^{2}\sum_{\boldsymbol{\Phi}_i\in C_j}\lambda_1\left(p - \mathrm{Tr}\left(\boldsymbol{\Phi}_i{}^T\overline{\boldsymbol{\Psi}}_j\right)\right)\\ &-\lambda_2\left(p - \mathrm{Tr}\left(\overline{\boldsymbol{\Psi}}_1^T\overline{\boldsymbol{\Psi}}_2\right)\right)\\ \text{s.t. }&\boldsymbol{\Phi}_i^T\boldsymbol{\Phi}_i = \boldsymbol{I}_p, \quad \text{for } i = 1, \ldots, N\end{aligned} \tag{9}$$

where λ_1 and λ_2 are nonnegative scalars balancing intragroup and intergroup terms. The first term in (9) is employed to penalize each set of individual harmonic waves $\boldsymbol{\Phi}_i$ characterizing the topology of its own Laplacian matrix $\boldsymbol{L}_i$. The second term is the intragroup distance, and the third term is the intergroup distance. In this way, (9) can effectively address intragroup variability and intergroup heterogeneity while preserving the local structure of the brain network. Specifically, since the trace-norm operation is not an intrinsic metric defined in the Stiefel manifold space, the orthonormal constraint term must be implemented to ensure the orthonormality of individual harmonic waves, i.e., $\boldsymbol{\Phi}_i^T\boldsymbol{\Phi}_i = \boldsymbol{I}_p$, for $i = 1, \ldots, N$, such that all individual harmonic waves $\{\boldsymbol{\Phi}_i\}$ are on the Stiefel manifold. For the group-dependent common harmonic waves $\{\overline{\boldsymbol{\Psi}}_j | j = 1, 2\}$, its orthonormality is not restricted since it is entirely driven by the intrinsic manifold geometry, as demonstrated in Section III-B. Moreover, because the objective function is not invariant to any orthonormal matrix $\boldsymbol{Q}(\boldsymbol{Q} \in \mathbb{R}^{n\times p}, \boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I})$, the common harmonic waves are optimized on the Stiefel manifold instead of other Riemannian manifolds [26].

B. Numerical Scheme to Solve the Manifold Harmonic Discriminant Analysis

One difficulty in solving the trace difference problem in (9) is that there is high complexity in jointly optimizing individual harmonic waves and group-dependent common harmonic waves simultaneously. Accordingly, two manifold gradient descent methods are proposed to solve this problem by alternating direction method of multipliers (ADMM) scheme [20]. The first one iteratively calculates the mean tangent in the tangent space and then maps it back to the Stiefel manifold, termed accurate optimization (AO). The second one directly employs curvilinear search optimization on the manifold, named by fast optimization (FO).

1) Accurate Optimization: The augmented Lagrangian function of (9) is as follows:

$$\begin{aligned}\min_{\{\boldsymbol{\Phi}_i\},\overline{\boldsymbol{\Psi}}_j} &\sum_{i=1}^{N}\{\mathrm{Tr}(\boldsymbol{\Phi}_i^T\boldsymbol{L}_i\boldsymbol{\Phi}_i) + \sum_{j=1}^{2}\sum_{\boldsymbol{\Phi}_i\in C_j}\left\{\lambda_1\left(p - \mathrm{Tr}\left(\boldsymbol{\Phi}_i{}^T\overline{\boldsymbol{\Psi}}_j\right)\right)\right.\\ &\left.+\mathrm{Tr}\left(\boldsymbol{\Lambda}_i^T\left(\boldsymbol{\Phi}_i^T\boldsymbol{\Phi}_i - \boldsymbol{I}_p\right)\right)\right\} - \lambda_2\left(p - \mathrm{Tr}\left(\overline{\boldsymbol{\Psi}}_1^T\overline{\boldsymbol{\Psi}}_2\right)\right)\end{aligned} \tag{10}$$

where $\boldsymbol{\Lambda}_i \in R^{p\times p}$ is the Lagrangian multiplier. The minimization problem in (10) can be decoupled into two subproblems. Each subproblem is solved, and the obtained solution is

updated iteratively until convergence is reached. The whole numerical procedure is summarized in Algorithm 1.

Algorithm 1 AO

Input: Adjacency matrices $\{(\boldsymbol{W}_1, y_1), \cdots, (\boldsymbol{W}_N, y_N)\}$ with $\boldsymbol{W}_i \in \mathbb{R}^{n\times n}$ and $y_i \in \{1, 2\}$.

Output: Group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_{j=1,2}$.

Initializing parameters λ_1, λ_2, and $\boldsymbol{L}_i = \boldsymbol{D}_i - \boldsymbol{W}_i$;
Initialize group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_j$ and individual harmonic waves $\boldsymbol{\Phi}_i \in \mathbb{R}^{n\times p}$;
Calculate the positive definite matrix $\tilde{\boldsymbol{L}}_i = \beta \boldsymbol{I} - \boldsymbol{L}_i$.
while $\varepsilon < threshold$ **do**
  for $i = 1$ to N **do**
    while $\| \boldsymbol{\Phi}_i^{(k+1)} - \boldsymbol{\Phi}_i^{(k)} \| < \varepsilon_1$ **do**
      Update $\boldsymbol{\Theta}_i \in \mathbb{R}^{n\times p} \leftarrow \tilde{\boldsymbol{L}}_i \boldsymbol{\Phi}_i^{(k)} + \lambda_1 \overline{\boldsymbol{\Psi}}_j$.
      Calculate $\mathbf{U}_i \boldsymbol{\Sigma}_i \mathbf{V}_i^T = \boldsymbol{\Theta}_i$ by SVD method.
      Update $\boldsymbol{\Phi}_i^{(k+1)} = \boldsymbol{U}_i \boldsymbol{V}_i^T$.
    end while
  end for
  Randomly set starting point $\overline{\boldsymbol{\Psi}}_j^1 = \boldsymbol{\Phi}_i, \boldsymbol{\Phi}_i \in C_j$.
  while $\|\Delta \overline{\boldsymbol{\Psi}}_j\| < \varepsilon_2$ **do**
    Calculate $\Delta \overline{\boldsymbol{\Psi}}_1^{(k+1)}$ and $\Delta \overline{\boldsymbol{\Psi}}_2^{(k+1)}$ using the Eq. (15) and Eq. (16) respectively.
    Mapping back to manifold space $\overline{\boldsymbol{\Psi}}_j^{(k+1)} = \exp_{\overline{\boldsymbol{\Psi}}_j^{(k)}}(\Delta \overline{\boldsymbol{\Psi}}_j^{(k+1)})$
  end while
  Update $\overline{\boldsymbol{\Psi}}_j = \overline{\boldsymbol{\Psi}}_j^{(k+1)} (j = 1, 2)$.
  Calculate New_{cost} using the Eq. (9).
  $\varepsilon = abs(New_{cost} - Old_{cost})$
  Update $Old_{cost} = New_{cost}$
end while

Subproblem 1 (Solving the Individual Harmonic Waves $\boldsymbol{\Phi}_i$ Intragroup): All the individual harmonic waves $\{\boldsymbol{\Phi}_i\}$ are independent of the group-dependent common harmonic waves; thus, $\boldsymbol{\Phi}_i$ can be evaluated by fixing $\overline{\boldsymbol{\Psi}}_j$. The problem modeled in (10) is a quadratic optimization problem on the Stiefel manifold provided that the matrix $\boldsymbol{L}_i$ is positive definite; thus, $\tilde{\boldsymbol{L}}_i = \beta \boldsymbol{I} - \boldsymbol{L}_i$ is modified to substitute $\boldsymbol{L}_i$, where the parameter β is set as the maximum eigenvalue of $\boldsymbol{L}_i$. By removing the irrelevant variables, the objective function is simplified as

$$\max_{\{\boldsymbol{\Phi}_i\}} \sum_{i=1}^{N} \mathrm{Tr}(\boldsymbol{\Phi}_i^T \tilde{\boldsymbol{L}}_i \boldsymbol{\Phi}_i) + \sum_{j=1}^{2} \sum_{\boldsymbol{\Phi}_i \in C_j} \left\{ \lambda_1 \mathrm{Tr}\left(\boldsymbol{\Phi}_i^T \overline{\boldsymbol{\Psi}}_j\right) - \mathrm{Tr}\left(\boldsymbol{\Lambda}_i^T \left(\boldsymbol{\Phi}_i^T \boldsymbol{\Phi}_i - \boldsymbol{I}_p\right)\right) \right\}. \tag{11}$$

According to the KKT condition

$$\frac{\partial \tilde{F}_{\boldsymbol{\Phi}_i}}{\partial \boldsymbol{\Phi}_i} = 2\tilde{\boldsymbol{L}}_i \boldsymbol{\Phi}_i + \lambda_1 \overline{\boldsymbol{\Psi}}_j - 2\boldsymbol{\Phi}_i \boldsymbol{\Lambda}_i = 0 \quad \forall \boldsymbol{\Phi}_i \in C_j. \tag{12}$$

However, directly solving $\boldsymbol{\Phi}_i$ in (12) through direct inversion may result in high-computational complexity or instability. To address this issue, the generalized power iteration (GPI) method [27] is employed to iteratively update $\boldsymbol{\Phi}_i$, which effectively solves the inverse of $\boldsymbol{\Phi}_i$ and the calculation of the Lagrange multiplier $\boldsymbol{\Lambda}_i$. The process consists of four key steps.

1) Decompose the Laplacian matrix $\boldsymbol{L}_i$ via SVD to obtain the initial individual harmonic waves $\boldsymbol{\Phi}_i$ to satisfy $\boldsymbol{\Phi}_i^T \boldsymbol{\Phi}_i = \boldsymbol{I}_p$.
2) Update $\boldsymbol{\Theta}_i \in \mathbb{R}^{n\times p} \leftarrow \tilde{\boldsymbol{L}}_i \boldsymbol{\Phi}_i^{(k)} + \lambda_1 \overline{\boldsymbol{\Psi}}_j, \forall \boldsymbol{\Phi}_i \in C_j$.
3) Calculate $\boldsymbol{\Phi}_i$ by solving the constrained trace-norm problem as

$$\max_{\boldsymbol{\Phi}_i^T \boldsymbol{\Phi}_i = I_p} \mathrm{Tr}\left(\boldsymbol{\Phi}_i^T \boldsymbol{\Theta}_i\right) \quad \forall \boldsymbol{\Phi}_i \in C_j, \ j \in \{1, 2\}. \tag{13}$$

It has a closed-form solution of $\boldsymbol{\Phi}_i = \boldsymbol{U}_i \boldsymbol{V}_i^T$, where $\boldsymbol{U}_i \in \mathbb{R}^{n\times p}$ and $\boldsymbol{V}_i \in \mathbb{R}^{p\times p}$ are the left and right eigenmatrices of $\boldsymbol{\Theta}_i$, respectively.
4) Iteratively perform steps (2)–(4) until converges.

Subproblem 2 (Solving the Group-Dependent Common Harmonic Waves $\overline{\boldsymbol{\Psi}}_j$ Intergroup): By fixing the set of individual harmonic waves obtained in subproblem 1, the objective function on group-dependent common harmonic waves can be simplified as follows:

$$\min_{\overline{\boldsymbol{\Psi}}_j} \sum_{j=1}^{2} \sum_{\boldsymbol{\Phi}_i \in C_j} \left(p(1 - \lambda_2) - \mathrm{Tr}\left(\boldsymbol{\Phi}_i^T \overline{\boldsymbol{\Psi}}_j\right)\right) + \lambda_2 \mathrm{Tr}\left(\overline{\boldsymbol{\Psi}}_1^T \overline{\boldsymbol{\Psi}}_2\right). \tag{14}$$

Solving (14) can be considered as finding group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_j$ on the Stiefel manifold. This is achieved by minimizing the intragroup geodesic distance between each individual harmonic waves $\boldsymbol{\Phi}_i \in C_j$ and the underlying common harmonic waves $\overline{\boldsymbol{\Psi}}_j$ while maximizing the intergroup geodesic distance between the group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_j$ from different groups. Consequently, the problem is the classical Fréchet mean problem, which can be solved by the Weizfeld algorithm [28]. It consists of three iterative steps until convergence is reached.

1) Initialize the group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_j^{(k)}$ as the center points on the current Stiefel manifold.
2) Calculate the gradient $\nabla_{\overline{\boldsymbol{\Psi}}}$ of the energy function in (14) with respect to each individual harmonic waves $\boldsymbol{\Phi}_i$ on the current estimation of the group-dependent manifold center $\overline{\boldsymbol{\Psi}}_j^{(k)}$ through (8). By doing so, the mean tangent $\Delta \overline{\boldsymbol{\Psi}}_j^{(k+1)} \in \mathcal{T}_{\overline{\boldsymbol{\Psi}}_j^{(k)}}$ can be obtained by the following equations:

$$\Delta \overline{\boldsymbol{\Psi}}_1^{(k+1)} = -\sum_{\boldsymbol{\Phi}_i \in C_1} \left\{ \left(\overline{\boldsymbol{\Psi}}_1^{(k)} \boldsymbol{\Phi}_i^T \overline{\boldsymbol{\Psi}}_1^{(k)} - \boldsymbol{\Phi}_i\right) - \lambda_2 \left(\overline{\boldsymbol{\Psi}}_1^{(k)} \left(\overline{\boldsymbol{\Psi}}_2^{(k)}\right)^T \overline{\boldsymbol{\Psi}}_1^{(k)} - \overline{\boldsymbol{\Psi}}_2^{(k)}\right) \right\} \tag{15}$$

$$\Delta \overline{\boldsymbol{\Psi}}_2^{(k+1)} = -\sum_{\boldsymbol{\Phi}_i \in C_2} \left\{ \left(\overline{\boldsymbol{\Psi}}_2^{(k)} \boldsymbol{\Phi}_i^T \overline{\boldsymbol{\Psi}}_2^{(k)} - \boldsymbol{\Phi}_i\right) - \lambda_2 \left(\overline{\boldsymbol{\Psi}}_2^{(k)} \left(\overline{\boldsymbol{\Psi}}_1^{(k)}\right)^T \overline{\boldsymbol{\Psi}}_2^{(k)} - \overline{\boldsymbol{\Psi}}_1^{(k)}\right) \right\}. \tag{16}$$

3) Map the mean tangent $\Delta \overline{\boldsymbol{\Psi}}_j^{(k+1)}$ back to the Stiefel manifold through $\overline{\boldsymbol{\Psi}}_j^{(k+1)} = \exp_{\overline{\boldsymbol{\Psi}}_j^{(k)}}(\Delta \overline{\boldsymbol{\Psi}}_j^{(k+1)})$ to update the group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_j^{(k+1)}$.

2) Fast Optimization: A significant challenge in Riemannian optimization is the computational expense associated with exponential mappings and nonnegative matrix decompositions when optimizing Stiefel manifolds. To this end, the FO is proposed as a tailored manifold optimization to address the high-computational complexity of the AO algorithm. The update strategy of the FO involves a variation of the steepest gradient descent (SGD) method that is adapted for optimization on manifolds. Often, SGD computes the next iteration point as $\mathbf{Y} = \mathbf{X} - \tau \nabla_X F$, where $\tau > 0$ is a step size and $\nabla_X F \in \mathcal{T}_X F$. The issue arises when the newly generated point does not satisfy $\mathbf{Y}^T\mathbf{Y} = \mathbf{I}$. Accordingly, the FO algorithm improves the term $\nabla_X F$ similar to [29] and computes the next iteration point from the equation

$$\hat{\boldsymbol{Y}} = \mathbf{X} - \frac{\tau}{2}\tilde{\boldsymbol{W}}(\mathbf{X} + \hat{\mathbf{X}}) \tag{17}$$

where $\hat{\boldsymbol{Y}}$ is the next point on the manifold. $\mathbf{X}$ and $\hat{\mathbf{X}}$ are the current point and the initialized center point, respectively. Specifically, $(1/2)\tilde{\boldsymbol{W}}\mathbf{X} \approx (1/2)\nabla_X F$. The induced direction $(1/2)\tilde{\boldsymbol{W}}\hat{\mathbf{X}}$ compels the update of $\hat{\boldsymbol{Y}}$ to align with the geodesic, helping the algorithm to follow the natural shape of the manifold. $\tilde{\boldsymbol{W}}$ is a skew-symmetric matrix [26] as $\tilde{\boldsymbol{W}} = \hat{\boldsymbol{W}} - \hat{\boldsymbol{W}}^\top$, where $\hat{\boldsymbol{W}} = F_X\boldsymbol{X}^\top - (1/2)\boldsymbol{X}(\boldsymbol{X}^\top F_X \boldsymbol{X}^\top)$ and F_X is the gradient of the objective function at $\mathbf{X}$. As a result, the new point $\hat{\boldsymbol{Y}}$ is updated by combining the initial direction $(1/2)\tilde{\boldsymbol{W}}\mathbf{X}$ and the induced direction $(1/2)\tilde{\boldsymbol{W}}\hat{\mathbf{X}}$ along the geodesic line. Additionally, a momentum $\hat{\boldsymbol{M}}$ is introduced into the gradient direction via extension [30], [31] to improve the rate of convergence. Therein, the momentum in the kth optimization step is defined as

$$\hat{\boldsymbol{M}}^{(k)} = \beta\hat{\boldsymbol{M}}^{(k-1)} + (1-\beta)F_{\boldsymbol{X}_i}^{(k-1)} \tag{18}$$

where $\beta > 0$ is the balance parameter and the momentum $\hat{\boldsymbol{M}}^{(k-1)} \in \mathcal{T}_X F$. The momentum $\hat{\boldsymbol{M}}^{(k)}$ is the direction of iteration. In this case, $\hat{\boldsymbol{W}}^{(k)} = \hat{\boldsymbol{M}}^{(k)}{\mathbf{X}_i^{(k)}}^T - (1/2)\mathbf{X}_i^{(k)}(\mathbf{X}_i^{(k)T}\hat{\boldsymbol{M}}^{(k)}\mathbf{X}_i^{(k)T})$. The momentum $\hat{\boldsymbol{M}}^{(k)}$ is updated within the tangent space of the current point [30], and can be expressed as $\hat{\boldsymbol{M}}^{(k)} = \tilde{\boldsymbol{W}}^{(k)}\mathbf{X}_i^{(k)}$. In (18), the projection onto the tangent space is a linear map [26] that satisfies $\tau_{\boldsymbol{M}}(\boldsymbol{M}^{(k-1)}) + F_{\boldsymbol{X}_i}^{(k-1)} = \pi_{\mathcal{T}_{\boldsymbol{X}_i}}(\boldsymbol{M}^{(k-1)}) + \pi_{\mathcal{T}_{\boldsymbol{X}_i}}(F_{\boldsymbol{X}_i^{(k-1)}}) = \pi_{\mathcal{T}_{\boldsymbol{X}_i}}(\boldsymbol{M}^{(k-1)} + F_{\boldsymbol{X}_i}^{(k-1)})$. Thus, a linear combination of the gradient $F_{\boldsymbol{X}_i}^{(k-1)}$ and the momentum $\hat{\boldsymbol{M}}^{(k-1)}$ can reasonably be used to generate a new gradient direction $\hat{\boldsymbol{M}}^{(k)}$. Subsequently, the momentum $\hat{\boldsymbol{M}}^{(k)}$ is updated within the tangent space, and the center point $\hat{\mathbf{X}}$ is initialized along this same space following the SGD. Finally, the next point on the Stiefel manifold is updated using (17), which exclusively involves matrix multiplication, circumventing the need for matrix decomposition and exponential retraction. Additionally, this indicates that the FO algorithm updates the manifold points along the gradient direction of the curve as a manifold-to-manifold process. The process of the FO algorithm composes of the following steps.

1) Initialization of the input $\mathbf{X}_i$ and momentum $\hat{\boldsymbol{M}}^0 = \mathbf{0}$, satisfying $\mathbf{X}_i^T\mathbf{X}_i = \boldsymbol{I}_p$.
2) Update $F_{\boldsymbol{X}_i} \in \mathbb{R}^{n\times p}$.
3) Calculate momentum $\hat{\boldsymbol{M}}^{(k)} = \beta\hat{\boldsymbol{M}}^{k-1} + (1-\beta)F_{\boldsymbol{X}_i}$.
4) Calculate the auxiliary matrix by the following equations:

$$\hat{\boldsymbol{W}}^{(k)} \leftarrow \hat{\boldsymbol{M}}^{(k)}{\mathbf{X}_i^{(k)}}^T - \frac{1}{2}\mathbf{X}_i^{(k)}\left(\mathbf{X}_i^{(k)T}\hat{\boldsymbol{M}}^{(k)}{\mathbf{X}_i^{(k)}}^T\right). \tag{19}$$

5) Update $\tilde{\boldsymbol{W}}^{(k)} \leftarrow \hat{\boldsymbol{W}}^{(k)} - (\hat{\boldsymbol{W}}^{(k)})^T$.
6) Update momentum $\hat{\boldsymbol{M}}^{(k)} = \tilde{\boldsymbol{W}}^{(k)}\mathbf{X}_i^{(k)}$.
7) Initialize $\hat{\mathbf{X}}_i = \mathbf{X}_i^{(k)} - \tau\hat{\boldsymbol{M}}^{(k)}$.
8) Update $\mathbf{X}_i^{(k+1)} \leftarrow \mathbf{X}_i^{(k)} - (\tau/2)\tilde{\boldsymbol{W}}^{(k)}(\mathbf{X}_i^{(k)} + \hat{\mathbf{X}}_i)$.
9) Iteratively perform steps (2)–(8) until converges.

In this article, the FO algorithm is employed to update individual harmonic waves and group-dependent common harmonic waves. However, there is a slight difference in the second update step for these two types of harmonic waves. Specifically, $F_{\boldsymbol{\Phi}_i} = \tilde{\boldsymbol{L}}_i\boldsymbol{\Phi}_i + \lambda_1\overline{\boldsymbol{\Psi}}_j$ corresponds to the update of the individual harmonic waves differential, whereas $F_{\boldsymbol{\Psi}_j} = -\boldsymbol{\Phi}_i + \lambda_2\boldsymbol{\Psi}_j$ corresponds to the update of the group-dependent common harmonic waves differential. Detailed information on the optimization procedure can be found in Algorithm 2.

Algorithm 2 FO

Input: Adjacency matrices $\{(\boldsymbol{W}_1, y_1), \cdots, (\boldsymbol{W}_N, y_N)\}$ with $\boldsymbol{W}_i \in \mathbb{R}^{n\times n}$ and $y_i \in \{1, 2\}$.

Output: Group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_{j=1,2}$.

Initializing parameters λ_1, λ_2, β, τ and $\boldsymbol{L}_i = \boldsymbol{D}_i - \boldsymbol{W}_i$; Initialize group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_j$ and individual harmonic waves $\boldsymbol{\Phi}_i \in \mathbb{R}^{n\times p}$; Momentum $\hat{\boldsymbol{M}}^0 = \mathbf{0}$; Calculate the positive definite matrix $\tilde{\boldsymbol{L}}_i = \beta\boldsymbol{I} - \boldsymbol{L}_i$.
while $\varepsilon < threshold$ **do**
  for $i = 1$ to N **do**
    while $\| \boldsymbol{\Phi}_i^{(k+1)} - \boldsymbol{\Phi}_i^{(k)} \| < \varepsilon_1$ **do**
      Update $F_{\boldsymbol{\Phi}_i} \leftarrow \tilde{\boldsymbol{L}}_i\boldsymbol{\Phi}_i^{(k)} + \lambda_1\overline{\boldsymbol{\Psi}}_j$.
      Calculate $\hat{\boldsymbol{M}}^{(k)} = \beta\hat{\boldsymbol{M}}^{k-1} + (1-\beta)F_{\boldsymbol{X}_i}$.
      Calculate $\hat{\boldsymbol{W}}^{(k)}$ by Eq. (19)
      Update $\tilde{\boldsymbol{W}}^{(k)} \leftarrow \hat{\boldsymbol{W}}^{(k)} - (\hat{\boldsymbol{W}}^{(k)})^T$ *and* $\hat{\boldsymbol{M}}^{(k)} \leftarrow \tilde{\boldsymbol{W}}^{(k)}\boldsymbol{\Phi}_i^{(k)}$
      Initialize $\hat{\boldsymbol{\Phi}}_i \leftarrow \boldsymbol{\Phi}_i^{(k)} - \tau\hat{\boldsymbol{M}}^{(k)}$
      Update $\boldsymbol{\Phi}_i^{(k+1)} \leftarrow \boldsymbol{\Phi}_i^{(k)} - \frac{\tau}{2}\tilde{\boldsymbol{W}}^{(k)}(\boldsymbol{\Phi}_i^{(k)} + \hat{\boldsymbol{\Phi}}_i)$.
    end while
  end for
  Randomly set starting point $\overline{\boldsymbol{\Psi}}_j^1 = \boldsymbol{\Phi}_i, \boldsymbol{\Phi}_i \in C_j$.
  while $\|\Delta\overline{\boldsymbol{\Psi}}_j\| < \varepsilon_2$ **do**
    Update $F_{\boldsymbol{\Psi}_j} = -\boldsymbol{\Phi}_i + \lambda_2\boldsymbol{\Psi}_j$.
    Calculate $\hat{\boldsymbol{M}}^{(k)} = \beta\hat{\boldsymbol{M}}^{k-1} + (1-\beta)F_{\boldsymbol{\Psi}_j}$.
    Calculate $\hat{\boldsymbol{W}}^{(k)}$ by Eq. (19)
    Update $\tilde{\boldsymbol{W}}^{(k)} \leftarrow \hat{\boldsymbol{W}}^{(k)} - (\hat{\boldsymbol{W}}^{(k)})^T$ and $\hat{\boldsymbol{M}}^{(k)} \leftarrow \tilde{\boldsymbol{W}}^{(k)}\boldsymbol{\Psi}_j^{(k)}$
    Initialize $\hat{\boldsymbol{\Psi}}_j \leftarrow \boldsymbol{\Psi}_j^{(k)} - \tau\hat{\boldsymbol{M}}^{(k)}$
    Update $\boldsymbol{\Psi}_j^{(k+1)} \leftarrow \boldsymbol{\Psi}_j^{(k)} - \frac{\tau}{2}\tilde{\boldsymbol{W}}^{(k)}(\boldsymbol{\Psi}_j^{(k)} + \hat{\boldsymbol{\Psi}}_j)$.
  end while
  Update $\overline{\boldsymbol{\Psi}}_j = \overline{\boldsymbol{\Psi}}_j^{(k+1)} (j = 1, 2)$.
  Calculate New_{cost} using the Eq. (9).
  $\varepsilon = abs(New_{cost} - Old_{cost})$.
  Update $Old_{cost} = New_{cost}$
end while

By the above alternate optimization strategy, both the individual harmonic waves and group-dependent common harmonic waves are eventually obtained for early diagnosis of AD and AD-related brain region identification. In our subsequent experiments, we will use the MHDA and MHDA-M to refer to the AO and FO algorithms, respectively. We discuss the effectiveness of both optimization algorithms in Section IV-F. It is worth noting that to meet the requirement of high accuracy in predicting the early stage of AD, we primarily apply the MHDA algorithm to analyze the limited brain network data in most of the experiments.

IV. Experiments and Results

A. Comparison Approaches

14 popular methods are borrowed from seven major types for performance comparison. They are briefly introduced as follows.

1) *Baselines:* FLDA, support vector machine with linear kernel (L-SVM), AdaBoost, and random forest.
2) *GNNs:* GCNautoencoder [16] and BrainGB [17].
3) *Spectral Graph Analysis Methods:* Spectral regression discriminant analysis (SRDA) [32], and spectral regression kernel discriminant analysis (SRKDA) [33] with multicluster sparse feature selection.
4) *Manifold Regularized Method:* Manifold discriminative feature selection (MDFS) [34] with linear kernel SVM classifier, i.e., M-SVM.
5) *Multiple Kernel Method:* Scalable multiple kernel learning (EasyMKL) [35].
6) *Higher-Order Tensor Method:* Higher-order SVD with sparse logistic regression (S-HOSVD) [36].
7) *Manifold Metric and Matrix Linear Discriminant Methods:* Projection metric learning (PML) [21], manifold harmonic analysis (MHA) [20], and matrix linear discriminant analysis (MLDA) [22].

The goal of this work is to learn group-dependent common harmonic waves with discriminative ability by combining harmonic waves and the Stiefel manifolds. The classification performance of the proposed method is compared with four classic classifiers, as well as several other brain network classification methods. For the GNN approach, the GCNautoencoder learns a low-dimensional embedding for brain network classification through a symmetric GNN. BrainGB is a benchmark tailored to brain networks using GNNs. Spectral graph analysis methods, including SRDA and SRKDA, are constructed as supervised brain network classification algorithms, where SRKDA utilizes the MCFS algorithm for feature selection. The manifold regularization method M-SVM leverages low-dimensional manifold projections to discover the local correlations of the brain network. For the multikernel method, easyMKL attempts to classify the brain network data by a set of predefined kernel matrices. The higher order tensor method S-HOSVD is a higher order feature extraction classification model proposed for structural brain networks computed from diffusion MRI. Two manifold methods, PML and MHA, are geometric-aware dimensionality reduction and structural brain network MHA methods, respectively. Matrix discriminant analysis is a method for classifying high-dimensional matrices.

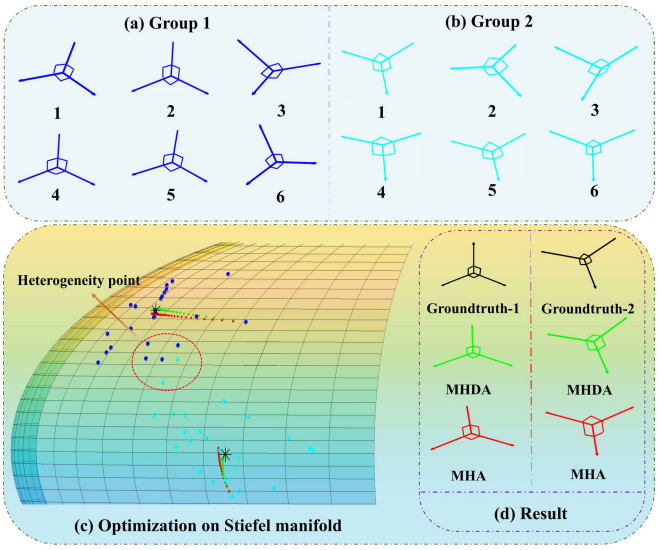

Fig. 2. Comparison of MHA and MHDA for finding the group-dependent common centers on the Stiefel manifold. Top: Two groups of samples (a) Group-1 and (b) Group-2 are generated by different rotations of the ground truth. Bottom: (c) Optimization process of MHA (trajectory in red) and MHDA (trajectory in green) on Stiefel manifolds. (d) Group-dependent common center identification results of MHDA and MHA.

To ensure experimental fairness, the parameters of the comparison methods were empirically tuned according to the original works.

To evaluate the reasonableness of the group-dependent harmonic waves obtained by the MHDA algorithm, the performance of our group-dependent common harmonic waves $\bar{\boldsymbol{\Psi}}_j$ optimized on the Stiefel manifold is compared with the performances of two other Euclidean average harmonic waves, that is: 1) $\boldsymbol{\Psi}_j$ by averaging all individual harmonic waves and 2) $\widetilde{\boldsymbol{\Psi}}_j$ by first averaging all adjacency matrices and then applying SVD decomposition to the averaged adjacency matrix. In the following, $\boldsymbol{\Psi}_j$ and $\widetilde{\boldsymbol{\Psi}}_j$ are the arithmetic mean harmonic waves and pseudo-harmonic waves, respectively.

B. Experiment on Synthetic Data

In this section, low-dimensional synthetic data are employed to illustrate the robustness of MHDA in terms of the heterogeneity of brain networks. Because the proposed method finds group-dependent common centers on the Stiefel manifold, it is reasonable to generate a series of low-dimensional synthetic data with orthonormality. Motivated by Fletcher et al. [37], a set of 3-D orthonormal rotation matrices is generated as individual harmonic waves located on the Stiefel manifold. This synthetic dataset consists of 48 random 3-D rotation matrices with 24 samples for each group. Specifically, the ground truth (common harmonic waves) for each group is displayed in black in the first row of Fig. 2(d). Given the rotation axes $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, the rotation angles θ_1 and θ_2 are randomly generated from a zero-mean Gaussian distribution with standard deviation $\sigma = \pi/18$ and a Gaussian distribution with mean $\pi/4$ and $\sigma = \pi/20$, respectively. After that, the two groups of rotation matrices centered at the ground truth are obtained. Fig. 2(a) and (b) shows six rotation matrices for each group. On the Stiefel manifold, the optimization

TABLE I
DEMOGRAPHIC INFORMATION ON TRAINING DATA IN ADNI DATABASE

Gender	Number	Age Range	Mean Age	CN	MCI	AD
Male	70.0	55.6 ∼ 89.0	74.6	24.0	29.0	32.0
Female	85.0	55.0 ∼ 90.3	73.0	31.0	15.0	24.0
Total	**155.0**	**55.0 ∼ 90.3**	**73.8**	**55.0**	**44.0**	**56.0**

processes of the MHA and MHDA are shown with red and green trajectories, respectively. The ground truth (manifold center) of each group is displayed as black "*" in Fig. 2(c).

For fairness, the MHA and MHDA are set to start from the same initial point to the stable point corresponding to the convergence of each method during this experiment. Specifically, for the MHA method, the common center is identified group by group by solving the Fréchet mean. The MHA and MHDA are applied to evaluate the group-dependent common centers from two groups of random perturbed 3-D rotation matrices. The MHA and MHDA results are shown in Fig. 2(c) and (d), and the group-dependent common centers identified by the proposed MHDA method are closer to the ground truth than those identified by the MHA method regarding the influence of some heterogeneous points [shown in the red dashed circle in Fig. 2(c)]. This is mainly because the MHDA with the introduction of the Fisher criterion is more reasonable in evaluating the group-dependent common centers on the Stiefel manifold than the MHA method by simply finding the manifold center group by group. Notably, there is no significant difference across the group-dependent common harmonic waves initialized with different starting points, as demonstrated in Section IV-C1.

C. Experiment on the ADNI Dataset

A total of 155 subjects were collected from the ADNI database for AD classification tasks, including 55 CN subjects, 44 MCI subjects, and 56 AD subjects. Each subject was scanned by T1-weighted MRI and diffusion-weighted MRI (DTI). The cortical surface was divided into 148 cortical regions based on T1-weighted MRI according to a Destrieux atlas [38]. Then, surface seed-based probabilistic fiber tractography [38] was applied to diffusion tensor imaging data to generate a 148×148 connectivity matrix. The demographic information is indicated in Table I.

To understand the effectiveness of the proposed method, the performance of the group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_j$ identified by the proposed MHDA method was evaluated on the ADNI dataset. A series of experiments were designed to compare the performance of the proposed MHDA method with the performances of the two categories of methods, including two common harmonic waves-based methods (arithmetic mean harmonic waves $\boldsymbol{\Psi}_j$ and pseudo-harmonic waves $\widetilde{\boldsymbol{\Psi}}_j$) and 14 representative brain network classification approaches. Specifically, the replicability of the common harmonic waves ($\overline{\boldsymbol{\Psi}}_j$, $\boldsymbol{\Psi}_j$, and $\widetilde{\boldsymbol{\Psi}}_j$) was evaluated by resampling tests. Then, the diagnostic capability of group-dependent common harmonic waves was compared with that of all comparison methods. To test the scalability of the proposed method on large-scale samples, the classification accuracy of the MHDA was evaluated on resting-state functional magnetic resonance imaging (rs-fMRI) images of 465 subjects from the ADNI database. Finally, the hyperparameter sensitivity, computational complexity, and convergence analysis were discussed.

1) Evaluation of Replicability: To verify the robustness of the group-dependent common harmonic waves obtained by MHDA, the replicability of the three common harmonic waves (i.e., $\overline{\boldsymbol{\Psi}}_j$, $\boldsymbol{\Psi}_j$ and $\widetilde{\boldsymbol{\Psi}}_j$) was tested between CN and AD. Specifically, the resampling procedure below was applied to generate 50 test/retest datasets from the CN and AD brain network datasets. First, a total of 30 CN networks were randomly sampled from 55 CN networks, and 30 AD networks were randomly sampled from 56 AD networks. Second, two sets of networks were sampled from the remaining 25 CN networks, each with three networks. Similarly, the same process was applied to AD networks. Third, combining the networks sampled in steps 1) and 2) above, CN and AD can form two paired cohorts with 33 networks for each cohort. After that, the MHDA algorithm was used on these two paired cohorts to obtain $\overline{\boldsymbol{\Psi}}_j$. Since two paired cohorts only have 9.1%(3/33) differences in each group (CN or AD), the replicability of the proposed method could be evaluated by testing whether there is a significant difference at each element in the group-dependent common harmonic waves via the paired t-test. Fewer significant elements imply that the algorithm has better replicability.

Because each row of the harmonic wave matrix relates to a brain region, the significance ($p < 0.01$) of these harmonic waves can be mapped to the cortical surface in Fig. 3(a)–(f). Obviously, the group-dependent common harmonic waves corresponding to the CN and AD groups obtained by the proposed algorithm have stronger replicability than pseudoharmonic waves $\widetilde{\boldsymbol{\Psi}}_j$. In addition, the arithmetic mean harmonic waves $\boldsymbol{\Psi}_j$ [in Fig. 3(a) and (b)] have replicability performance similar to that of the group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_j$, but the arithmetic mean harmonic waves $\boldsymbol{\Psi}_j$ do not guarantee orthonormality [20] and thus destroy the topology of the brain network. This nonorthogonality influences the classification performance of the brain networks, which is demonstrated in Section IV-C2.

2) Evaluate the Diagnostic Performance of Group-Dependent Common Harmonic Waves: In this experiment, the classification performance of the MHDA method is compared with the performances of different harmonic waves ($\boldsymbol{\Psi}_j$, $\widetilde{\boldsymbol{\Psi}}_j$, and $\overline{\boldsymbol{\Psi}}_j$) and 14 representative machine learning algorithms. Specifically, since baseline, spectral graph, manifold regularization, and multikernel methods based on vectors as input are challenging to implement for high-dimensional brain network classification, simple preprocessing is applied to the adjacency matrix. Here, because the 148×148 adjacency matrix is symmetric, the upper or lower triangular matrix is vectorized as a sample, and all samples are combined into a matrix, where each row represents a sample and each column represents the features. Compared with the above methods, the higher order tensor, GNN and matrix discriminant analysis methods work directly on the adjacency matrix of the brain network without additional preprocessing methods. The manifold metric methods take the $n \times p$ dimensional harmonic waves

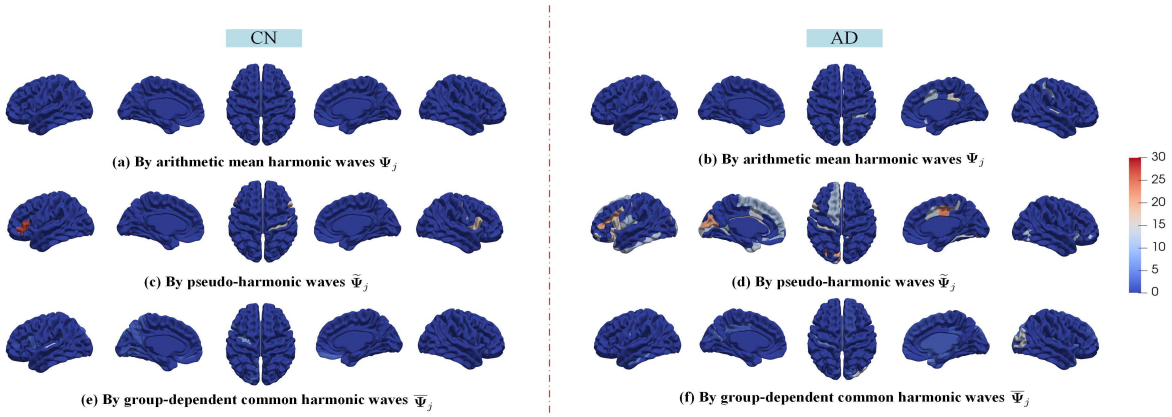

Fig. 3. Replicability test results of (a) and (b) arithmetic mean harmonic waves $\boldsymbol{\Psi}_j$, (c) and (d) pseudo-harmonic waves $\tilde{\boldsymbol{\Psi}}_j$, and (e) and (f) group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_j$ on CN (left) and AD (right) data. Among them, the color on the cortical surface reflects the number of times the replicability tests failed.

TABLE II
CLASSIFICATION RESULTS FOR DIFFERENT HARMONIC WAVES. ACC: ACCURACY, SEN: SENSITIVITY, SPE: SPECIFICITY

Data	Method	ACC	SEN	SPE	F-score
CN/MCI	$\tilde{\boldsymbol{\Psi}}_j$	0.502±0.031	0.472±0.041	0.527±0.038	0.462±0.033
	$\boldsymbol{\Psi}_j$	0.555±0.036	0.501±0.049	0.573±0.045	0.505±0.041
	$\overline{\boldsymbol{\Psi}}_j$	**0.642±0.030***	**0.633±0.044***	**0.651±0.040***	**0.615±0.032***
AD/MCI	$\tilde{\boldsymbol{\Psi}}_j$	0.480±0.034	0.467±0.043	0.491±0.040	0.446±0.035
	$\boldsymbol{\Psi}_j$	0.589±0.035	0.555±0.046	0.627±0.043	0.561±0.042
	$\overline{\boldsymbol{\Psi}}_j$	**0.685±0.029***	**0.656±0.042***	**0.709±0.037***	**0.653±0.034***
CN/AD	$\tilde{\boldsymbol{\Psi}}_j$	0.548±0.035	0.581±0.041	0.514±0.038	0.568±0.036
	$\boldsymbol{\Psi}_j$	0.591±0.038	0.618±0.042	0.561±0.046	0.607±0.038
	$\overline{\boldsymbol{\Psi}}_j$	**0.776±0.023***	**0.836±0.039***	**0.714±0.034***	**0.793±0.035***

derived from the Laplacian matrix as input. In addition to using harmonic waves as input, a testing sample can also be stratified as the label of the group-dependent manifold center that has the shortest geodesic distance to this testing sample. The performance of all methods is evaluated by calculating the classification accuracy, sensitivity, specificity, and F-score based on fivefold cross-validation. Finally, the classification results based on different common harmonic waves are summarized in Table II, and the classification results compared with the 14 approaches are summarized in Table III. Specifically, the bolded and underlined values in Tables II and III are the optimal and suboptimal values.

As shown in Table II, our group-dependent common harmonic waves $\overline{\boldsymbol{\Psi}}_j$ consistently outperform the arithmetic mean harmonic waves $\boldsymbol{\Psi}_j$ and the pseudo-harmonic waves $\tilde{\boldsymbol{\Psi}}_j$ in the classification tasks of CN/MCI, AD/MCI, and CN/AD. Table II and Fig. 3(c) and (d) show that pseudo-harmonic waves $\tilde{\boldsymbol{\Psi}}_j$ fail to achieve satisfactory classification performance in all three classification tasks, mainly because of their poor replicability and greater sensitivity to noise. Although arithmetic mean harmonic waves $\boldsymbol{\Psi}_j$ in Fig. 3(a) and (b) have replicability similar to that of our algorithm, they destroy the topological structure of the brain network when utilizing a simple arithmetic average, thereby resulting in poor classification performance.

Table III presents the performances of all methods for the three classification tasks on the ADNI dataset. According to the above experimental results, the proposed algorithm further improves the diagnostic performance of AD compared with its competitors. This can be attributed to the ability of the proposed method to preserve the spatial information of the brain network and to obtain group-dependent common harmonic waves of the Stiefel manifold to enhance the separability of different classes. First, the three discriminative-based methods are compared. The MHDA and MLDA methods based on high-dimensional matrix-valued data outperform the low-dimensional FLDA method for the three classification tasks of ADNI. Additionally, the accuracy and sensitivity of MHDA substantially improved compared with those of the PML and MHA methods. These results show that the MHDA is more effective than PML and MHA in measuring the geodesic distance between any two Stiefel matrices. Notably, MHDA-M achieves a slightly lower accuracy than MDHA; however, it has higher computational efficiency (at least four times improvement) than MDHA, as shown in Fig. 7(a) of Section IV-F. In addition, Fig. 7(b) demonstrated that MHDA-M is more efficient than MHDA on high-dimensional brain network datasets.

Next, the comparison of GCNAutoencoders, BrainGCN, and the proposed MHDA is notable. The GCNAutoencoders method has a slightly lower classification performance than the MHDA on the classification tasks of the brain network. These experiments show that GCNAutoencoders is not as effective as the proposed model when dealing with high-dimensional small-scale brain network datasets. Moreover, the performance of BrainGB based on GNN is comparable to that of the proposed algorithm and higher than that of GCNAutoencoders. This suggests that although GNNs are suitable for graph data, the design of GNN models in practical applications should be tailored to follow the unique nature of brain network data. As a result, the method proposed in this article can be compared with BrainGN for small-scale brain network data, mainly considering the intrinsic topological properties of brain network data.

Third, vector-based brain network classification methods, such as M-SVM, L-SVM, and SRKDA, destroy the topology of the brain network during data processing, resulting in poor performance. However, the proposed method has good classification performance, indicating that the MHDA method

TABLE III

CLASSIFICATION RESULTS COMPARED WITH 14 REPRESENTATIVE METHODS. ACC: ACCURACY, SEN: SENSITIVITY, SPE: SPECIFICITY

Data	Method	ACC	SEN	SPE	F1-score
CN/MCI	FLDA	0.619±0.032	0.376±0.077	0.633±0.064	0.635±0.038
	MLDA	0.633±0.039	0.519±0.111	0.705±0.102	0.669±0.046
	L-SVM	0.619±0.070	0.407±0.105	0.711±0.094	0.659±0.068
	Adaboost	0.609±0.062	0.458±0.132	**0.842±0.101**	**0.687±0.054**
	RandomForest	0.544±0.037	0.441±0.095	0.686±0.125	0.592±0.064
	SRDA	0.563±0.037	0.484±0.056	0.642±0.100	0.563±0.062
	SRKDA	0.556±0.038	0.546±0.048	0.681±0.057	0.532±0.059
	M-SVM	0.583±0.046	0.546±0.094	0.630±0.088	0.618±0.042
	EasyMKL	0.560±0.013	0.484±0.052	0.620±0.047	0.497±0.036
	S-HOSVD	0.533±0.029	0.435±0.088	0.661±0.153	0.508±0.047
	PML	0.552±0.031	0.586±0.061	0.508±0.078	0.598±0.039
	MHA	0.573±0.025	0.506±0.029	0.692±0.035	0.566±0.035
	GCNAutoencoders	0.633±0.047	0.601±0.097	0.670±0.145	0.664±0.053
	BrainGB	**0.645±0.053**	0.631±0.106	0.646±0.088	0.636±0.048
	MHDA-M	0.635±0.048	0.607±0.085	0.657±0.069	0.661±0.078
	MHDA	0.642±0.030	**0.633±0.044**	0.651±0.040	0.615±0.032
AD/MCI	FLDA	0.613±0.092	0.380±0.033	0.769±0.078	0.631±0.029
	MLDA	0.662±0.056	0.484±0.078	0.764±0.072	0.712±0.041
	L-SVM	0.650±0.050	0.481±0.103	0.727±0.055	0.664±0.058
	Adaboost	0.613±0.039	0.351±0.065	**0.820±0.060**	0.683±0.036
	RandomForest	0.537±0.032	0.358±0.072	0.788±0.077	0.628±0.045
	SRDA	0.520±0.023	0.524±0.091	0.501±0.066	0.438±0.053
	SRKDA	0.550±0.032	0.582±0.079	0.513±0.082	0.510±0.046
	M-SVM	0.623±0.036	0.521±0.076	0.682±0.032	0.648±0.023
	EasyMKL	0.607±0.012	0.493±0.060	0.715±0.066	0.538±0.048
	S-HOSVD	0.561±0.059	0.493±0.057	0.639±0.069	0.538±0.053
	PML	0.593±0.025	0.508±0.037	0.659±0.044	0.647±0.028
	MHA	0.613±0.026	0.517±0.023	0.738±0.038	0.602±0.028
	GCNAutoencoders	0.679±0.048	0.676±0.071	0.665±0.104	0.702±0.083
	BrainGB	0.685±0.071	**0.696±0.108**	0.667±0.136	0.701±0.086
	MHDA-M	0.683±0.045	0.651±0.069	0.702±0.074	**0.721±0.050**
	MHDA	**0.685±0.029**	0.656±0.042	0.709±0.037	0.653±0.034
CN/AD	FLDA	0.638±0.054	0.775±0.097	0.585±0.088	0.592±0.052
	MLDA	0.729±0.047	0.816±0.039	0.650±0.081	0.689±0.073
	L-SVM	0.681±0.065	0.743±0.077	0.700±0.087	0.669±0.074
	Adaboost	0.601±0.044	0.770±0.083	0.608±0.110	0.564±0.068
	RandomForest	0.660±0.037	0.779±0.078	0.582±0.104	0.590±0.066
	SRDA	0.663±0.041	0.780±0.066	0.549±0.130	0.627±0.098
	SRKDA	0.682±0.046	0.792±0.036	0.538±0.080	0.626±0.075
	M-SVM	0.693±0.049	0.749±0.062	0.625±0.088	0.648±0.066
	EasyMKL	0.642±0.023	0.596±0.025	0.692±0.059	0.588±0.052
	S-HOSVD	0.671±0.028	0.553±0.028	**0.788±0.079**	0.627±0.015
	PML	0.624±0.039	0.635±0.068	0.612±0.047	0.627±0.048
	MHA	0.665±0.030	0.647±0.037	0.682±0.028	0.658±0.032
	GCNAutoencoders	0.753±0.053	0.716±0.078	0.759±0.106	0.743±0.086
	BrainGB	**0.784±0.083**	0.808±0.076	0.738±0.130	**0.796±0.084**
	MHDA-M	0.766±0.046	0.775±0.056	0.760±0.055	0.747±0.048
	MHDA	0.776±0.023	**0.836±0.039**	0.714±0.034	0.793±0.035

can learn brain network representations with more discriminative information, ultimately improving the classification performance.

Finally, the proposed algorithm obtains high accuracy and sensitivity and is more convincing in the diagnosis of AD than other algorithms. High sensitivity is indispensable for clinical applications because the costs of misclassifying a normal person as a patient and misclassifying a patient as a healthy person [39] are different. Therefore, the manifold discriminant analysis method constructed by brain network harmonic waves is advantageous for brain network classification.

D. Scalability Analysis

According to previous studies, the effectiveness of the proposed MHDA method in the brain network classification of small-scale samples is observed. To show that the proposed algorithm can also handle the large-scale brain network dataset, rs-fMRI images of 465 subjects, including 282 CN subjects and 183 MCI subjects from the ADNI dataset, were collected. The functional brain network generation method is based on [40]. Here, five representative methods, including FLDA, MLDA, MHA, GCNAutoencoders, and BrainGB, are selected for comparison. The experimental results are shown in Fig. 4. From Fig. 4, one can observe that the classification results of the GNN approach are not consistently better than the proposed MHDA. Meanwhile, no significant differences in accuracy ($p = 0.110$), F-score ($p = 0.052$) and sensitivity ($p = 0.068$) are found between MHDA and BrainGB. Specificity score of MHDA is significant higher ($p = 0.048$) than BrainGB. Additionally, MHDA has a higher accuracy, F1-score, and sensitivity than the state-of-the-art MLDA and MHA. The results show that the proposed MHDA model can learn topological properties of brain networks with differentiation in large-scale data.

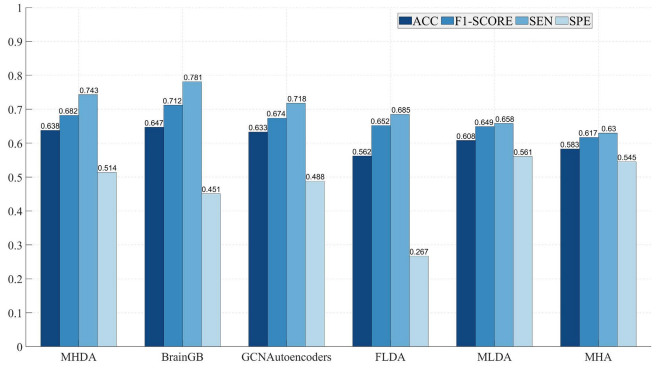

Fig. 4. Scalability analysis in the ADNI re-fMRI dataset.

E. Hyperparameter Sensitivity

The hyperparameters λ_1 and λ_2 in our objective function (9) play essential roles in balancing the weights of intragroup and intergroup geometric distances. In this experiment, the effects of different hyperparameters λ_1 and λ_2 on our final classification results are investigated. Specifically, CN/AD, AD/MCI, and CN/MCI classification experiments are performed on the ADNI data, and then, the grid search method is applied to find optimal hyperparameters λ_1 and λ_2 corresponding to the highest ACC score in a set $[0.001, 0.0015, 0.01, 0.015, 0.1]$. From the experimental results shown in Fig. 5, the MHDA has a relatively stable ACC score within a specific range.

F. Computational Complexity and Convergence Analysis

This section is dedicated to discussing the convergence and complexity of both AO algorithm and FO algorithm separately.

According to AO algorithm, the computational complexity of the MHDA is determined by two factors: 1) computing the individual harmonic wave $\mathbf{\Phi}_i$ intragroup, which is $\mathcal{O}(Nt_1(n^2p + p^3))$, where t_1 is the number of iterations, and 2) computing the group-dependent common harmonic waves $\overline{\mathbf{\Psi}}_j$ intergroup. Here, the computational complexity of (15) and (16) is $\mathcal{O}(n^2p)$. The corresponding complexities of QR decomposition and exponential mapping are $\mathcal{O}(np^2)$.

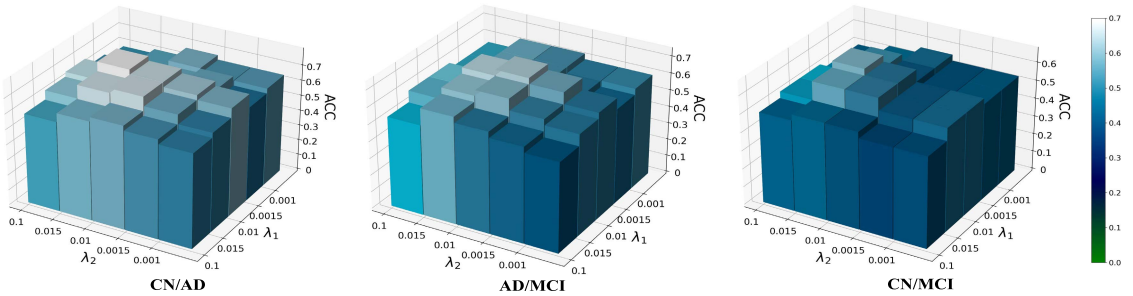

Fig. 5. ACC performance of the proposed MHDA with respect to the varying hyperparameters λ_1 and λ_2 on ADNI.

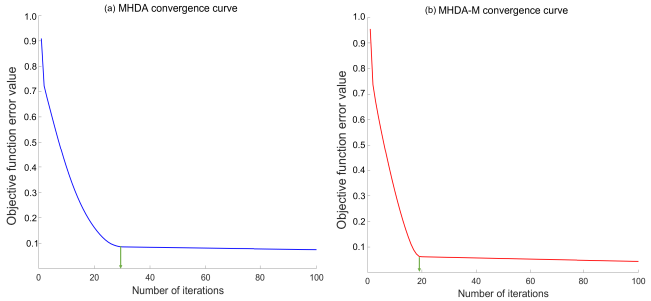

Fig. 6. Convergence curve of the proposed algorithm. (a) MHDA convergence curve. (b) MHDA-M convergence curve.

Clearly, the computational complexity of calculating $\overline{\mathbf{\Psi}}_j$ is $\mathcal{O}(t_2np(n+p))$, with t_2 being the number of iterations. Therefore, the primary time complexity of the proposed algorithm is $\mathcal{O}(Nt_1(n^2p+p^3)+t_2np(n+p))$. In the experiments, both t_1 and t_2 are set to 100. In contrast, FO algorithm has a computational complexity of $\mathcal{O}((t_2+Nt_1)n^2p)$. The complexity of FO algorithm is evidently decreased by $\mathcal{O}(Nt_1p^3+t_2np^2))$ in comparison to AO algorithm. Consequently, FO algorithm shows obvious advantages when processing high-dimensional manifold values.

According to [27] and [28], the convergence of AO algorithm is theoretically proven in the Supplementary Material S1. Experimentally, the convergence curves of AO algorithm and FO algorithm with respect to (9) are shown in Fig. 6(a) and (b), respectively. It can be observed that both AO algorithm and FO algorithm exhibit good convergence, where FO algorithm has faster convergence than AO algorithm.

To further compare the performance of the two proposed optimization algorithms on Stiefel manifold, we record the training time and accuracy of MHDA and MHDA-M on the ADNI dataset. For each optimization algorithm, 100 training iterations were conducted on a CPU using MATLAB 2022b. As shown in Fig. 7(a), FO algorithm significantly reduces the computational complexity by updating the manifold points simply by matrix multiplication. Furthermore, we generate 100 simulated brain networks with two classification labels, where the dimension (number of nodes) of each brain network varies from 100 to 1000 at intervals of 100, to test the influence of different brain network dimensions on the computational times of these two optimization algorithms. From Fig. 7(b), it is obvious that as the dimension of the brain

(a) Classification performance of MHDA and MHDA-M on ADNI data

Data	Method	ACC	Training time(s)
CN/MCI	MHDA	0.642	395.12
	MHDA-M	0.635	103.84
AD/MCI	MHDA	0.685	462.02
	MHDA-M	0.683	106.49
CN/AD	MHDA	0.776	560.88
	MHDA-M	0.766	118.98

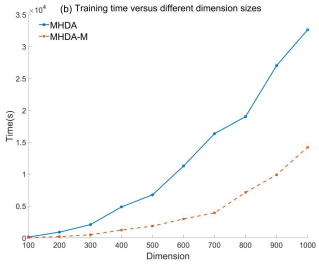

Fig. 7. Time cost of two optimization algorithms. (a) Classification performance of MHDA and MHDA-M on ADNI data. (b) Training time versus different dimension sizes.

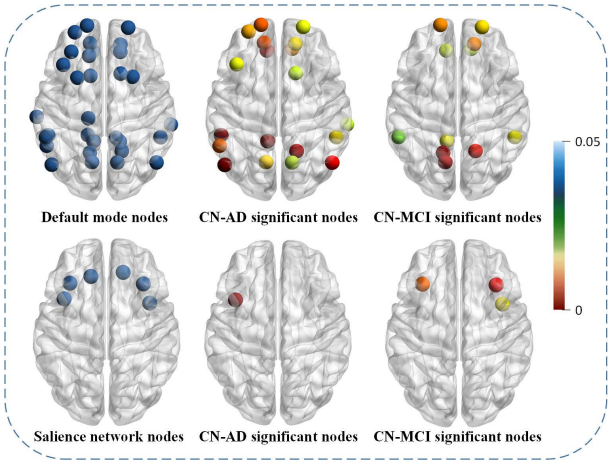

Fig. 8. Spatial alignment with identified harmonic waves to the DMN and SNs. Left: Nodes belonging to the DMN and SN. Middle: In the context of CN/AD significant nodes. Right: In the context of CN/MCI significant nodes.

network increases, FO algorithm manifests stronger computational efficiency than AO algorithm in terms of train time. In a word, for extremely high-dimensional and large-scale brain network datasets, we recommend FO algorithm because it is relatively simple and has less computational time. In the case of medium-sized datasets, AO algorithm is recommended since it is more accurate and the computational complexity is acceptable.

V. Discussion

A. Discovering Significant Brain Regions in AD

AD is commonly associated with the DMN, while behavioral variant frontotemporal dementia (bvFTD) primarily manifests in the salience network (SN) [41]. These findings provide an important option to employ structural connectivity

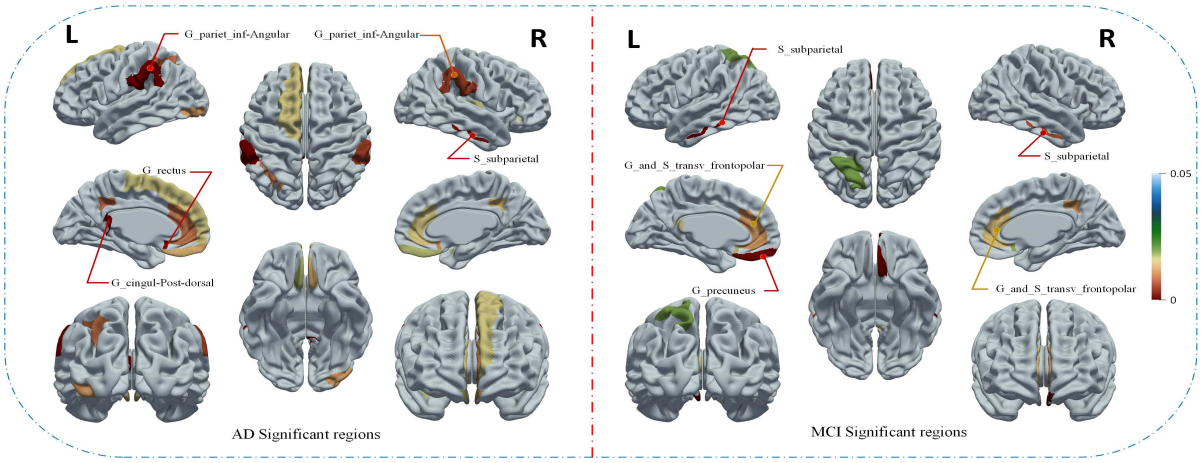

Fig. 9. Spatial locations of the top-five significant brain regions corresponding to the CN/AD test (left) and CN/MCI test (right).

to identify the significant brain regions in AD. Herein, since the individual harmonic waves are adjusted by minimizing the intragroup geodesic distance, it is reasonable to detect the significant difference between two groups of individual harmonic waves through statistical methods.

First, all learned individual harmonic waves are divided into two groups (CN/MCI or CN/AD). Then, the statistical multivariate t-test method is applied to find the significant brain regions related to disease in three steps: 1) two groups (i.e., CN/AD or CN/MCI) of optimized individual harmonic waves are obtained by the MHDA; 2) multivariate Hotelling's T-squared test [42] and false discovery rate (FDR) correction are applied to these two groups to detect underlying brain regions showing significant group differences (FDR-adjusted p-value < 0.05); and 3) spatial alignment is performed to obtain the spatial distribution of significant brain regions in the DMN network and SN network, as shown in Fig. 8 with the color bar representing the FDR-adjusted p-value. The CN/AD and CN/MCI groups had 19 and 11 significant nodes aligned with the DMN, respectively, while only one and three significant nodes were aligned with the SN. Furthermore, the CN/AD test has more matching points than the CN/MCI test on the DMN and less on the SN, which is consistent with the findings of the AD study.

Next, whether the top significant brain regions identified by our methods are closely relevant to the progression of the neurodegenerative disorder is investigated. The top-five significant brain regions found by the above method are mapped onto the cortical surface, as shown in Fig. 9. In Fig. 9, the top-five significant nodes of AD are observed mainly in the angular, subparietal-sulcus, postdorsal, and rectus regions, where bilateral angular gyri (core DMN hubs) are thought to be engaged in the global integration of information in the brain and to participate in the close interaction of large-scale brain networks [43], while both the postdorsal and rectus gyri play essential roles in recognition, cognition, and application [44]. The top-five MCI brain regions are mainly concentrated in the subparietal-sulcus, frontopolar, and precuneus regions. Among them, the frontopolar and precuneus regions have been reported to be associated with early signs of AD [45], [46]. The subparietal region is a highly significant brain region in both AD and MCI, and this region plays a crucial role in face recognition and related scene memory [47].

B. Limitations and Future Work

The proposed MHDA method aims to address the heterogeneity across brain networks by generating group-dependent common harmonic waves. Nonetheless, limitations exist in the approach.

1) The method solely relies on using individual harmonic waves extracted from brain networks for the early diagnosis of neurodegenerative diseases, without utilizing other auxiliary information to enhance the discriminative power of the model, such as empirical biomarkers like amyloid deposition. Future work will investigate the use of joint empirical biomarkers and brain networks for more precise diagnosis of neurodegenerative diseases.
2) Two hyperparameters λ_1 and λ_2 exist in the model. As a result, a grid search process is necessary for different classification tasks, which increases the algorithm time complexity. In future work, an adaptive variable parameter will be designed to facilitate quantitative analysis of intragroup and intergroup differences without considering hyperparameters.

VI. Conclusion

This article proposes a model for brain network classification called the MHDA, which provides a new solution for the early diagnosis of neurodegenerative diseases. The core of the proposed method extends the linear discriminant analysis from Euclidean space to the Stiefel manifold space through a series of manifold algebra operations on brain network harmonic waves. Each brain network can be represented by harmonic waves fully retaining its topological information. The intergroup and intragroup geodesic distances between different groups of individual harmonic waves are constructed through the manifold optimization scheme to discover the group-dependent common harmonic waves. Unlike the existing brain network analysis methods, the harmonic waves learned by the proposed method can effectively reflect the specific characteristics of the brain network and quantify the differences in neuropathological disease between individuals.

Meanwhile, two Riemannian optimization algorithms are presented in this article to enhance the generalizability of the proposed model in diverse scenarios. The power of the proposed method in stratifying CN, MCI, and AD is evaluated. Higher classification accuracy, sensitivity, specificity, and F-score are achieved by the proposed method compared with two common harmonic-based methods and 14 existing machine learning algorithms on brain network data. Furthermore, the learned harmonic waves reveal some putative biomarkers associated with AD development at the preclinical stage.

REFERENCES

[1] L. Du et al., "Identifying associations among genomic, proteomic and imaging biomarkers via adaptive sparse multi-view canonical correlation analysis," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 102003.

[2] R. C. Petersen, B. Caracciolo, C. Brayne, S. Gauthier, V. Jelic, and L. Fratiglioni, "Mild cognitive impairment: A concept in evolution," *J. Internal Med.*, vol. 275, no. 3, pp. 214–228, Mar. 2014.

[3] A. Association, "2019 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 15, no. 3, pp. 321–387, Mar. 2019.

[4] Y. Shi, H.-I. Suk, Y. Gao, S.-W. Lee, and D. Shen, "Leveraging coupled interaction for multimodal Alzheimer's disease diagnosis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 1, pp. 186–200, Jan. 2020.

[5] W. Yu, B. Lei, M. K. Ng, A. C. Cheung, Y. Shen, and S. Wang, "Tensorizing GAN with high-order pooling for Alzheimer's disease assessment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4945–4959, Sep. 2022.

[6] Y. Pan, M. Liu, C. Lian, Y. Xia, and D. Shen, "Spatially-constrained Fisher representation for brain disease identification with incomplete multi-modal neuroimages," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2965–2975, Sep. 2020.

[7] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Med. Image Anal.*, vol. 43, pp. 157–168, Jan. 2018.

[8] B. Jie, D. Zhang, W. Gao, Q. Wang, C.-Y. Wee, and D. Shen, "Integration of network topological and connectivity properties for neuroimaging classification," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 576–589, Feb. 2014.

[9] J. Ji, A. Zou, J. Liu, C. Yang, X. Zhang, and Y. Song, "A survey on brain effective connectivity network learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 4, pp. 1879–1899, Apr. 2023.

[10] X. Liao, A. V. Vasilakos, and Y. He, "Small-world human brain networks: Perspectives and challenges," *Neurosci. Biobehavioral Rev.*, vol. 77, pp. 286–300, Jun. 2017.

[11] T. A. Pascoal et al., "Aβ-induced vulnerability propagates via the brain's default mode network," *Nature Commun.*, vol. 10, no. 1, pp. 1–13, Jun. 2019.

[12] B. Lei, P. Yang, T. Wang, S. Chen, and D. Ni, "Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1102–1113, Apr. 2017.

[13] W. Yu et al., "Morphological feature visualization of Alzheimer's disease via multidirectional perception GAN," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 4401–4415, Aug. 2022.

[14] D. Yang et al., "Group-wise hub identification by learning common graph embeddings on Grassmannian manifold," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 8249–8260, Nov. 2022.

[15] C. Mantoux, B. Couvy-Duchesne, F. Cacciamani, S. Epelbaum, S. Durrleman, and S. Allassonnière, "Understanding the variability in graph data sets through statistical modeling on the Stiefel manifold," *Entropy*, vol. 23, no. 4, p. 490, Apr. 2021.

[16] J. Park, M. Lee, H. J. Chang, K. Lee, and J. Y. Choi, "Symmetric graph convolutional autoencoder for unsupervised graph representation learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6518–6527.

[17] H. Cui et al., "BrainGB: A benchmark for brain network analysis with graph neural networks," *IEEE Trans. Med. Imag.*, vol. 42, no. 2, pp. 493–506, Feb. 2023.

[18] D. S. Bassett and O. Sporns, "Network neuroscience," *Nature Neurosci.*, vol. 20, no. 3, pp. 353–364, Feb. 2017.

[19] S. Atasoy, I. Donnelly, and J. Pearson, "Human brain networks function in connectome-specific harmonic waves," *Nature Commun.*, vol. 7, no. 1, pp. 1–10, Jan. 2016.

[20] J. Chen et al., "Learning common harmonic waves on Stiefel manifold—A new mathematical approach for brain network analyses," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 419–430, Jan. 2021.

[21] Z. Huang, R. Wang, S. Shan, and X. Chen, "Projection metric learning on Grassmann manifold with application to video based face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 140–149.

[22] W. Hu, W. Shen, H. Zhou, and D. Kong, "Matrix linear discriminant analysis," *Technometrics*, vol. 62, no. 2, pp. 196–205, Apr. 2020.

[23] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, Jan. 1998.

[24] J. Jost, *Riemannian Geometry and Geometric Analysis*, vol. 42005. Berlin, Germany: Springer, 2008.

[25] J. Wang, L. Wang, F. Nie, and X. Li, "A novel formulation of trace ratio linear discriminant analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 10, pp. 5568–5578, Oct. 2022.

[26] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton NJ, USA: Princeton Univ. Press, 2009.

[27] F. Nie, R. Zhang, and X. Li, "A generalized power iteration method for solving quadratic problem on the Stiefel manifold," *Sci. China Inf. Sci.*, vol. 60, no. 11, Nov. 2017.

[28] K. Aftab, R. Hartley, and J. Trumpf, "Generalized Weiszfeld algorithms for Lq optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 4, pp. 728–745, Apr. 2015.

[29] Z. Wen and W. Yin, "A feasible method for optimization with orthogonality constraints," *Math. Program.*, vol. 142, nos. 1–2, pp. 397–434, Dec. 2013.

[30] F. Huang and S. Gao, "Gradient descent ascent for minimax problems on Riemannian manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 8466–8476, Jul. 2023.

[31] E. Ghadimi, H. R. Feyzmahdavian, and M. Johansson, "Global convergence of the heavy-ball method for convex optimization," in *Proc. Eur. Control Conf. (ECC)*, Jul. 2015, pp. 310–315.

[32] D. Cai, X. He, and J. Han, "SRDA: An efficient algorithm for large-scale discriminant analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 1, pp. 1–12, Jan. 2008.

[33] D. Cai, X. He, and J. Han, "Speed up kernel discriminant analysis," *VLDB J.*, vol. 20, no. 1, pp. 21–33, Feb. 2011.

[34] J. Zhang, Z. Luo, C. Li, C. Zhou, and S. Li, "Manifold regularized discriminative feature selection for multi-label learning," *Pattern Recognit.*, vol. 95, pp. 136–150, Nov. 2019.

[35] F. Aiolli and M. Donini, "EasyMKL: A scalable multiple kernel learning algorithm," *Neurocomputing*, vol. 169, pp. 215–224, Dec. 2015.

[36] L. Zhan et al., "Boosting brain connectome classification accuracy in Alzheimer's disease using higher-order singular value decomposition," *Frontiers Neurosci.*, vol. 9, p. 257, Jul. 2015.

[37] P. T. Fletcher, S. Venkatasubramanian, and S. Joshi, "The geometric median on Riemannian manifolds with application to robust atlas estimation," *NeuroImage*, vol. 45, no. 1, pp. S143–S152, Mar. 2009.

[38] C. Destrieux, B. Fischl, A. Dale, and E. Halgren, "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature," *NeuroImage*, vol. 53, no. 1, pp. 1–15, Oct. 2010.

[39] B. Jie, D. Zhang, C.-Y. Wee, and D. Shen, "Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification," *Hum. Brain Mapping*, vol. 35, no. 7, pp. 2876–2897, Jul. 2014.

[40] L. Zhang, L. Wang, and D. Zhu, "Predicting brain structural network using functional connectivity," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102463.

[41] J. Zhou and W. W. Seeley, "Network dysfunction in Alzheimer's disease and frontotemporal dementia: Implications for psychiatry," *Biol. Psychiatry*, vol. 75, no. 7, pp. 565–573, Apr. 2014.

[42] H. Hotelling, "The generalization of student's ratio," in *Breakthroughs in Statistics*. Berlin, Germany: Springer, 1992, pp. 54–65.

[43] D. Vatansever, A. E. Manktelow, B. J. Sahakian, D. K. Menon, and E. A. Stamatakis, "Angular default mode network connectivity across working memory load," *Hum. Brain Mapping*, vol. 38, no. 1, pp. 41–52, Jan. 2017.

[44] N. V. Gulyaeva, "Functional neurochemistry of the ventral and dorsal hippocampus: Stress, depression, dementia and remote hippocampal damage," *Neurochemical Res.*, vol. 44, no. 6, pp. 1306–1322, Jun. 2019.

[45] D. Yates, "Frontal cortex biopsy samples can predict Alzheimer disease," *Nature Rev. Neurol.*, vol. 7, no. 1, p. 5, Jan. 2011.

[46] J. Kim, Y.-H. Kim, and J.-H. Lee, "Hippocampus-precuneus functional connectivity as an early sign of Alzheimer's disease: A preliminary study using structural and functional magnetic resonance imaging data," *Brain Res.*, vol. 1495, pp. 18–29, Feb. 2013.

[47] O. Woolnough, P. S. Rollo, K. J. Forseth, C. M. Kadipasaoglu, A. D. Ekstrom, and N. Tandon, "Category selectivity for face and scene recognition in human medial parietal cortex," *Current Biol.*, vol. 30, no. 14, pp. 2707–2715, Jul. 2020.

Hongmin Cai (Senior Member, IEEE) received the B.S. and M.S. degrees in mathematics from the Harbin Institute of Technology, Harbin, China, in 2001 and 2003, respectively, and the Ph.D. degree in applied mathematics from The University of Hong Kong, Hong Kong, in 2007.

He is currently a Professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His research interests include biomedical image processing and omics data integration.

Xiaoqi Sheng received the B.S. and M.S. degrees from the School of Information Science and Engineering, Jiangxi University of Science and Technology, Ganzhou, China, in 2017 and 2020, respectively. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China.

His current research interests include medical imaging, brain network analysis, and machine learning.

Guorong Wu (Member, IEEE) received the Ph.D. degree in computer science and engineering from Shanghai Jiao Tong University, Shanghai, China, in 2007.

He is currently an Assistant Professor with the Department of Psychiatry and Computer Science, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. His research interests include fast and robust analysis of large population data, computer-assisted diagnosis, and image-guided radiation therapy.

Dr. Wu was a recipient of the Independent Career Award from NIH for the contribution on noninvasive neurobiological basis for early diagnosis of Alzheimer's disease in 2016.

Bin Hu (Fellow, IEEE) received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences in 1998.

He is currently a Professor and the Dean of the School of Medical Technology, Beijing Institute of Technology, Beijing, China. He has coauthored more than 400 publications. His research interests include using computational approaches to decode emotion, mind, and behavior.

Dr. Hu is an AAIA and Institution of Engineering and Technology (IET) Fellow. He serves as the Editor-in-Chief for the IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS.

Yiu-Ming Cheung (Fellow, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, in 1998.

He is currently a Chair Professor with the Department of Computer Science, Hong Kong Baptist University, Hong Kong. His current research interests include machine learning, pattern recognition, and visual computing, as well as their application.

Dr. Cheung is a fellow of the American Association for the Advancement of Science (AAAS), IET, and British Computer Society (BCS). He is the Founding Chairman for the Computational Intelligence Chapter of the IEEE Hong Kong Section. He is the Editor-in-Chief of the IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE. He also serves as an Associate Editor for the IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS from 2014 to 2020, *Pattern Recognition*, *Knowledge and Information Systems*, and *Neurocomputing*, just to name a few.

Jiazhou Chen (Member, IEEE) received the Ph.D. degree in computer science and technology from the South China University of Technology (SCUT), Guangzhou, China, in 2020.

During his period at SCUT, he studied at the Department of Psychiatry, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, as a Visiting Scholar, for one year, in 2019. He is currently a Post-Doctoral Researcher with the School of Computer Science and Engineering, SCUT. His current research interests include medical image analysis and brain network analysis.