Deep Manifold Harmonic Network With Dual Attention for Brain Disorder Classification

Xiaoqi Sheng, Jiazhou Chen, *Member, IEEE*, Yong Liu, *Member, IEEE*, Bin Hu, *Member, IEEE*, and Hongmin Cai, *Senior Member, IEEE*

***Abstract*—Numerous studies have shown that accurate analysis of neurological disorders contributes to the early diagnosis of brain disorders and provides a window to diagnose psychiatric disorders due to brain atrophy. The emergence of geometric deep learning approaches provides a new way to characterize geometric variations on brain networks. However, brain network data suffer from high heterogeneity and noise. Consequently, geometric deep learning methods struggle to identify discriminative and clinically meaningful representations from complex brain networks, resulting in poor diagnostic accuracy. Hence, the primary challenge in the diagnosis of brain diseases is to enhance the identification of discriminative features. To this end, this paper presents a dual-attention deep manifold harmonic discrimination (DA-DMHD) method for early diagnosis of neurodegenerative diseases. Here, a low-dimensional manifold projection is first learned to comprehensively exploit the geometric features of the brain network. Further, attention blocks with discrimination are proposed to learn a representation, which facilitates learning of group-dependent discriminant matrices to guide downstream analysis of group-specific references. Our proposed DA-DMHD model is evaluated on two independent datasets, ADNI and ADHD-200. Experimental results demonstrate that the model can tackle the hard-to-capture challenge of heterogeneous brain network topological differences and obtain excellent classifying performance in both accuracy and robustness compared with several existing state-of-the-art methods.**

***Index Terms*—Brain network, classification, deep learning, neurological disorders.**

Manuscript received 28 June 2022; revised 30 September 2022; accepted 2 November 2022. Date of publication 8 November 2022; date of current version 5 January 2023. This work was supported in part by the National Natural Science Foundation of China under Grants U21A20520, 62102153, 62172112, and 62227807, in part by the National Key R & D Program of China under Grant 2019YFA0706200, in part by the Key-Area Research and Development of Guangdong Province under Grant 2020B1111190001, in part by the Natural Science Foundation of Guangdong Province of China under Grant 2022A1515011162, in part by the Key-Area Research and Development Program of Guangzhou City under Grant 202206030009, in part by China Postdoctoral Science Foundation under Grant 2021M691062, and in part by the National Key Research and Development Program of China under Grant 2022YFE0112200. *(Corresponding author: Hongmin Cai.)*

Xiaoqi Sheng, Jiazhou Chen, and Hongmin Cai are with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China (e-mail: shengxiaoqi521@126.com; csjzchen@mail.scut.edu.cn; hmcai@scut.edu.cn).

Yong Liu is with the School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100000, China (e-mail: yongliu@bupt.edu.cn).

Bin Hu is with the School of Information Science and Engineering, Lanzhou University, Lanzhou, Gansu 730000, China (e-mail: bh@bit.edu.cn).

Digital Object Identifier 10.1109/JBHI.2022.3220545

I. Introduction

NEUROLOGICAL disorders are common progressive diseases of the nervous system characterized by selective dysfunction and neuronal loss [1], which might lead to serious problems, such as tremors in the limbs or face, along with memory loss. Given the progressive nature of neurological disorders, brain disorders usually occur gradually, resulting in severe behavioral and cognitive dysfunction. In the clinical context, the treatment of neurological disorders with specific medications and management strategies has emerged to enhance the quality of life. Unfortunately, there is no complete cure for these neurological disorders. In other words, neurological disorders have long been defined and studied, but the etiological bases and neural substrates are still not fully understood. Therefore, effective analysis of progressive brain dysfunction remains challenging, especially when brain diseases are already presented before clinical diagnosis.

The human brain is seen as a complicated network of interconnected neurons whose alterations in connectivity strongly indicate the onset of neurological pathologies [2]. Thus, investigating the organization of brain network connections is essential for understanding brain dysfunction caused by neurological pathologies. The evolution of noninvasive neuroimaging techniques and modern network science has provided exciting opportunities for analyzing the structural and functional connectivity of the brain [3], [4]. In parallel with the above developments, the study of brain network classification techniques is an essential application for understanding the mechanisms of brain connectivity in neuroscientific phenomena [5], which offers a potential methodology for neurological disorders in terms of motor function and social behavior features analysis [6].

Generally, brain network classification methods are divided into traditional machine and deep learning methods. Traditional machine learning methods, such as random forest and AdaBoost methods, have been widely tested [7]. These methods are effective on limited sample sizes. Unfortunately, traditional machine learning methods with shallow architectures utilize only low-level features for recognition, which has limitations in feature representation and brain network classification. Compared to traditional machine learning algorithms, deep learning methods can automatically learn high-level features from complex brain network data, thereby greatly advancing the development of the clinical diagnosis of neurological disorders [8], [9], [10]. Despite the novelty and success of recently reported deep learning methods, they suffer from hindered performance when analyzing

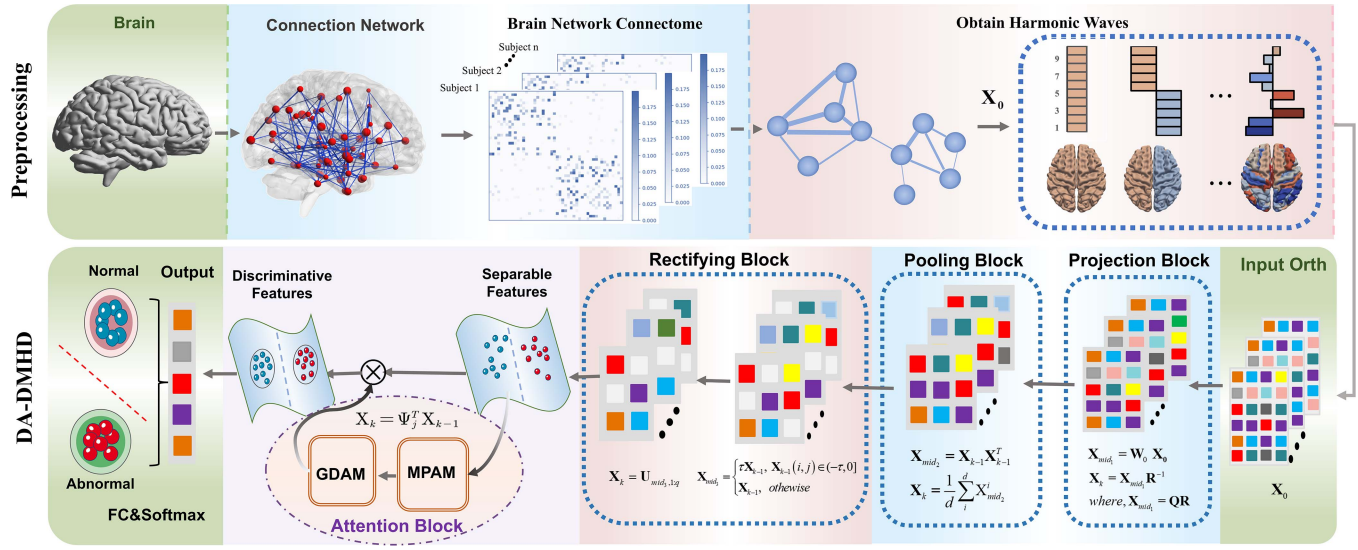

Fig. 1. Schematic diagram of the proposed brain network classification framework. The first row of this figure shows the process of constructing harmonic waves by utilizing graph theory for a given brain network dataset. The second row is the proposed DA-DMHD model, composed of the projection block to generate more discriminative Grassmann matrices via a learnable mapping W. The manifold value features are further integrated and compressed by the pooling block. Then, the nonlinearization of the deep manifold is achieved by the rectifying block. Next, the attention block is designed to capture the group-based geometric data distribution and enhance intergroup dispersion and intragroup compactness. Finally, these low-dimensional and discriminative features are classified by a fully connected layer and softmax layer.

high-dimensional brain network data. As is well-known, the connectivity of the brain network is a potentially non-Euclidean spatial structure [11] that arises naturally from the difficulty of capturing brain network topology. However, traditional deep learning algorithms are limited to data residing in vector space and ignore the brain network structure, which seriously affects classification accuracy. To circumvent this problem, numerous studies have attempted to represent the intrinsic topology of brain networks via Grassmann manifolds [12], [13], [14]. Nonetheless, existing Grassmann manifold brain network methods still focus on constructing kernel methods or metric learning, which have difficulty extracting deeper nonlinear features.

One possible solution is to extend traditional deep learning to deep Riemannian manifold space [15], [16], [17]. Notably, the input data of the network should be a Riemannian manifold-valued sample. Achieving this goal requires a new mathematical representation of the brain network that has sufficient statistical power to (1) represent the anatomical structure of brain networks and (2) provide a rigorous mathematical representation of Riemannian space to facilitate reasoning. In [18], the concept of network harmonic waves is proposed for statistical inference in the spectral domain spanned by graph eigenvectors. The study shows that harmonic waves can comprise a set of orthonormal bases from Laplacian eigenfunctions. Because of the orthogonality and independence of harmonic waves, they can be treated as Grassmann manifold-valued data embeddings, which means that harmonic waves provide a new spatial representation for brain network analysis in place of analyzing eigenvectors in Euclidean space. Subsequently, the topology of brain networks can be captured by the deep Riemannian manifold, facilitating the mining of useful brain connectivity network patterns in disease classification. Nevertheless, facing the highly heterogeneous brain network data, it is difficult to improve the diagnostic capability of the existing deep Riemannian manifold approach via mapping to a low-dimensional solution. Considering that the anatomical alterations caused by brain lesions are subtle, only a few brain network regions emerge with structural changes highly correlated with the disease, while other regions have little discriminating information. Therefore, the critical challenge of deep Riemannian manifold-based brain network diagnosis enhances the identification of discriminatory features as a way to address the heterogeneity among brain networks, including (1) significant characteristics among individual brain networks and (2) globally important regions between brain network groups.

To address the foregoing challenges, a dual-attention deep manifold harmonic discrimination (DA-DMHD) model is proposed to facilitate the diagnosis of brain diseases. This approach is a potential analog of a deep network model that can process manifold-valued data and learn a more favorable geometric representation of the brain network to improve the model's classification performance. Specifically, as shown in Fig. 1, the DA-DMHD model consists of two main components: the brain network preprocessing stage and the deep brain network classification stage. For the preprocessing stage (at the top of Fig. 1), given brain network data, the data are encoded by utilizing harmonic waves to obtain the brain network's manifold-valued data. For the deep brain network classification stage (at the bottom of Fig. 1), DA-DMHD focuses on making the features learned by deep Grassmann networks not only separable but also discriminative. First, DA-DMHD can learn separability structural features from the harmonic waves through the projection block and the pooling block. Then, through the rectifying block, the features between the layers are sparse and subsequently mapped back to Grassmann manifolds. Finally, an attention block is developed to learn the discriminative subspace representation of the group correlation. During training, the resulting group-dependent discriminant matrix acts on the output of the previous block to enhance the decision-making of the model on the brain network. The effectiveness of the DA-DMHD model is evaluated on two public brain network datasets (ADNI and ADHD-200).

The experimental results demonstrate that our proposed DA-DMHD method outperforms the state-of-the-art brain disease diagnosis methods in terms of accuracy and overall robustness. Compared to the existing method, our major contributions can be summarized as follows.

- A DA-DMHD method is proposed to improve the diagnosis of neurological disorders, automatically capturing group-dependent global structural features from brain networks and making brain disease-related classification decisions in a unified framework.
- The designed dual attention mechanism focuses on inter-channel relationships of individual features and intragroup feature differences, yielding group-dependent representations of brain network differences to enhance the discriminative ability of abnormal microstructural alterations caused by brain disorders.
- The projection block, pooling block, rectifying block, and attention block in DA-DMHD are integrated into a unified model with collaborative optimization to benefit each other. Extensive experimental results on synthetic and real datasets validate the efficacy of the method for high-dimensional heterogeneous brain network classification.

The rest of this paper is organized as follows: Section II briefly describes the theoretical background related to the study. Then, the DA-DMHD model and its numerical optimization scheme are illustrated in Section III. Next, Section IV presents the study material and the relevant experimental settings for the DA-DMHD model, and the experimental results are compared with those of several state-of-the-art diagnostic methods. Finally, Section V concludes the work.

TABLE I
LIST OF SYMBOLS USED AND THEIR CORRESPONDING EXPLANATIONS

Symbols	Explanation
$x, \boldsymbol{x}, \boldsymbol{X}$	Scalar,vector and matrix
k	The number of layers of this network
$\boldsymbol{X}_{k-1}$	The orthonormal input matrix of the $(k-1)$-th layer
$\boldsymbol{X}_k$	The output Grassmann matrix of the k-th layer
$\boldsymbol{X}_0$	The input to the model
$\boldsymbol{X}_{mid}$	The middle variables of the block
$\boldsymbol{Q}$	The orthonormal matrix consisting of the first p columns
$\boldsymbol{R}$	The invertible upper-triangular matrix
$\boldsymbol{W}_k$	The weight of the k-th layer
$\boldsymbol{U}_{mid_3,1:q}$	The first q largest eigenvectors of $\boldsymbol{X}_{mid_3}$
τ	The activation threshold
d	The number of instances for $\boldsymbol{X}_{mid_2}$
N	Sample size of brain network dataset
$\boldsymbol{\Psi}_j^T$	The group-dependent discriminant matrix
$tr(\cdot)$	Trace operation
$\Delta, \mathcal{T}_{\boldsymbol{X}}$	Tangent vector and tangent space at $\boldsymbol{X}$
$\mathcal{F}_{\boldsymbol{X}}$	Matrix derivative of function $\mathcal{F}$ with respect to $\boldsymbol{X}$
$\nabla_{\boldsymbol{X}}\mathcal{F}$	Gradient of $\mathcal{F}$ at point $\boldsymbol{X}$ in manifold space

II. PRELIMINARIES

This section introduces the primary theoretical background for extending the traditional deep learning (TDL) framework to Grassmann manifolds. First, the theoretical background of brain networks and the data representation are described. Then, the geometry of Grassmann manifolds is briefly introduced. The projection block and pooling block of previous works [16], which are employed in the proposed DA-DMHD model, are briefly introduced. The symbols used in this article and their corresponding explanations are provided in Table I.

A. Brain Network Representation via Harmonic Waves

A brain network can be encoded as a graph $G = (\mathcal{V}, \mathcal{E}, \boldsymbol{E})$ consisting of nodes $V = \{v_i | i \in 1, \ldots, n\}$ and edges $\mathcal{E} = \{e_{ij} | (v_i, v_j) \in V \times V\}$. Nodes divide the brain into different brain regions based on anatomical and functional criteria. Edges correspond to anatomical connections or statistical dependencies between brain regions [19]. $\boldsymbol{E} \in \mathbb{R}^{n \times n}$ is an adjacency matrix with positive weights.

Consider the brain networks with adjacency matrix $\{(\boldsymbol{E}_1, y_1), (\boldsymbol{E}_2, y_2), \ldots, (\boldsymbol{E}_N, y_N)\}$, where $\boldsymbol{E}_i \in \mathbb{R}^{n \times n}$ represents the i-th individual network and its label $y_i \in \{1, 2\}$. $C_j\ (j = 1, 2)$ is defined as the set of brain networks from the j-th group. Each brain network $(\boldsymbol{E}_i, y_i)$ can be represented by the eigenvector $\boldsymbol{\Phi}_i \in \mathbb{R}^{n \times n}$ of the Laplacian matrix $\boldsymbol{L}_i = \boldsymbol{D}_i - \boldsymbol{E}_i$, where $\boldsymbol{D}_i = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ is a degree matrix of $\boldsymbol{E}_i$. Since the eigendecomposition operator is a harmonic functional, the resulting eigenvector $\boldsymbol{\Phi}_i$ can be referred to as individual harmonic waves. Essentially, harmonic waves, which are an extension of the classic Fourier transform, offer a new mathematical basis for analyzing spatial patterns of brain self-organization [18]. Due to the orthogonality of harmonic waves, each set of individual harmonic waves $\boldsymbol{\Phi}_i$ can be reasonably represented as an instance on the Grassmann manifold for the downstream analyses of brain networks. However, the harmonic waves associated with high frequencies may contain substantial noise; thus, only the first p harmonic waves in each $\boldsymbol{\Phi}_i \in \mathbb{R}^{n \times p}$ is taken [19]. Importantly, harmonic waves are a spatial expansion on a Fourier basis, in which spatial frequencies are highly consistent with the anatomical structure of the brain network [20]. Meanwhile, the Grassmannian manifold representation of the brain network is obtained by harmonic waves and has fewer dimensions than the original brain network. Thus, harmonic waves tailor a new spatial analysis of brain networks for brain disease diagnosis.

B. Geometry of Grassmann Manifolds

Consider an orthonormal group $\boldsymbol{O}_n$ consisting of orthonormal matrices $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{n \times n}$. Mathematically, a Grassmann manifold $\mathcal{G}(n, p)$ is defined as the set of all p-dimensional subspaces of the Euclidean space $\mathbb{R}^n$. Every element is identifiable on the Grassmann manifold $\mathcal{G}(n, p) = \boldsymbol{O}_n / (\boldsymbol{O}_p - \boldsymbol{O}_{n-p})$ by an equivalence class of orthonormal basis matrices $\boldsymbol{X} = \boldsymbol{\mathcal{Q}}(:, 1:p)$ of size $n \times p$ spanning the same subspace.

The points of the Grassmann manifold correspond to a unique projection matrix $\boldsymbol{X}\boldsymbol{X}^T$ with size $n \times n$ and rank p [21]. Each point on the Grassmann manifold, in this way, has a corresponding unique matrix. Accordingly, the inner product corresponds to the projection mapping $\Upsilon(\boldsymbol{X})$ that efficiently represents the linear subspaces and approximates the true Grassmann geodesic distances, denoted $\langle \boldsymbol{X}_1, \boldsymbol{X}_2 \rangle_\Upsilon = tr(\Upsilon(\boldsymbol{X}_1)^T \Upsilon(\boldsymbol{X}_2))$. Then, a geodesic distance measure called the projection metric is derived

as follows [22]:

$$d^2\left(\boldsymbol{X}_1, \boldsymbol{X}_2\right) = p - tr(\boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{X}_2^T\boldsymbol{X}_1^T) \tag{1}$$

For a smooth real-valued function $\mathcal{F}$ defined on the Grassmann manifold, the gradient of the differentiable function $\mathcal{F}$ at point $\boldsymbol{X}_1 \in \mathcal{G}(n,p)$ can be found according to the formula in [21], that is, $\nabla_{\boldsymbol{X}_1}\mathcal{F} = \mathcal{F}_{\boldsymbol{X}_1} - \boldsymbol{X}_1\boldsymbol{X}_1^T\mathcal{F}_{\boldsymbol{X}_1}$, where $\mathcal{F}_{\boldsymbol{X}_1} \doteq \frac{\partial \mathcal{F}}{\partial \boldsymbol{X}_1}$. Accordingly, (1) becomes:

$$\nabla_{\boldsymbol{X}_1} d^2\left(\boldsymbol{X}_1, \boldsymbol{X}_2\right) = -2(I_m - \boldsymbol{X}_1\boldsymbol{X}_1^T)\boldsymbol{X}_2\boldsymbol{X}_2^T\boldsymbol{X}_1 \tag{2}$$

Notably, the gradient calculation on the tangent space occupies an essential position on the nonlinear manifold, which offers an effective solution for the gradient descent direction of the function $\mathcal{F}$ on the manifold [23]. Now, the tangent vector $\Delta \in \mathcal{T}_{\boldsymbol{X}_1}$ is mapped to the Grassmann manifold via exponential mapping:

$$\exp_{\boldsymbol{X}_1}(\Delta) = [\boldsymbol{X}_1\boldsymbol{V}\cos(\boldsymbol{\Sigma}) + \boldsymbol{U}\sin(\boldsymbol{\Sigma})]\boldsymbol{V}^\top \tag{3}$$

where $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ is the compact singular value decomposition (SVD) of the tangent vector. Clearly, Grassmann manifolds provide a reasonable medium for the geometric analysis and topological preservation of brain networks.

C. Deep Network on Grassmann Manifolds

1) Projection Block: This block is composed of a full-rank mapping (FRMap) layer and a reorthonormalization (ReOrth) layer. Specifically, a nontrivial goal of the FRMap layer is to generate a more compact and efficient feature matrix. In what follows, individual harmonic waves $\boldsymbol{X}_0 = \boldsymbol{\Phi}_i$ are used as initial inputs and converted into a new matrix by the designed FRMap layer f_{FRMap} as follows:

$$\boldsymbol{X}_{mid_1} = f_{FRMap}\left(\boldsymbol{X}_{k-1}; \boldsymbol{W}_k\right) = \boldsymbol{W}_k\boldsymbol{X}_{k-1} \tag{4}$$

where $\boldsymbol{X}_{k-1} \in \mathcal{G}(n,p)$ is the orthonormal input matrix of the $(k-1)$-th layer, $\boldsymbol{W}_k \in \mathbb{R}^{n_k \times n_{k-1}}(n_k \leq n_{k-1})$ is the connection weight matrix, and $\boldsymbol{X}_{mid_1} \in \mathbb{R}^{n_{k-1} \times p}$ is the output result matrix of f_{FRMap}. For each FRMap layer, one can employ multiple projections $\{\boldsymbol{W}_k^1, \ldots, \boldsymbol{W}_k^m\}$ on each orthonormal matrix $\boldsymbol{X}_{k-1}$, where m is the number of connected matrices. The multiple projection operation on $\boldsymbol{X}_{k-1}$ yields m channels, resulting in wealthier features. The intermediate matrix $\boldsymbol{X}_{mid_1}$ may be unorthonormal. To ensure its orthogonality, a QR decomposition is applied in the ReOrth layer, i.e., $\boldsymbol{X}_{mid_1} = \boldsymbol{Q}_k\boldsymbol{R}_k$ with $\boldsymbol{Q}_k \in \mathbb{R}^{n_{k-1} \times p}$ and $\boldsymbol{R}_k \in \mathbb{R}^{p \times p}$. Then, the ReOrth layer aims to normalize the column vectors of the matrix $\boldsymbol{X}_k$:

$$\boldsymbol{X}_k = f_{ReOrth}\left(\boldsymbol{X}_{mid_1}\right) = \boldsymbol{X}_{mid_1}\boldsymbol{R}_k^{-1} = \boldsymbol{Q}_k \tag{5}$$

2) Pooling Block: To maintain a balance between calculation time and accuracy, a mean pooling layer is applied to the Grassmann manifold-valued data. Specifically, the manifold-valued data $\boldsymbol{X}_{k-1}$ are first reduced to a flat space by the projection mapping (ProjMap) layer $f_{ProjMap}$ as follows:

$$\boldsymbol{X}_{mid_2} = f_{ProjMap}(\boldsymbol{X}_{k-1}) = \boldsymbol{X}_{k-1}\boldsymbol{X}_{k-1}^T \tag{6}$$

Specifically, m Euclidean projection matrices $\{\boldsymbol{X}_{mid_2}^i \mid 1 \leq i \leq m\}$ of size $n_{k-1} \times n_{k-1}$ can be obtained from (6). A sliding mean filter is then applied to each patch within each projection matrix (please refer to [16] for details), *i.e.*, the projection pooling (ProjPool) layer can be defined as

$$\boldsymbol{X}_k = f_{ProjPool}\left(\{\boldsymbol{X}_{mid_2}^1, \ldots, \boldsymbol{X}_{mid_2}^d\}\right) = \frac{1}{d}\sum_i^d \boldsymbol{X}_{mid_2}^i \tag{7}$$

$\boldsymbol{X}_{mid_2}^i$ is a channel feature for each subject X. The number d of instances in this paper is set to 4.

III. Attention-Guided Deep Manifold Learning on Brain Network Data

A. Network Structures on Grassmann Manifolds

As mentioned in Section II, the topological pattern analysis of brain networks is treated as a problem of learning representations on Grassmann manifolds and construct our model based on Grassmann manifolds. Consequently, the DA-DMHD model (shown in Fig. 1) is designed to learn a Grassmann manifold representation with both separability and discriminability. The DA-DMHD model consists of four key blocks: the projection block (i.e., feature representation learning for brain networks described in Section II-C1), the pooling block (i.e., the pooling operation of the feature matrix described in Section II-C2), the rectifying block (i.e., nonlinear operation between layers described in Section III-B) and the attention block (i.e., the discriminative attention mechanism described in Section III-C).

B. Nonlinear Operations Between Layers

Similar to the block in [17], a rectifying block is designed to achieve nonlinear feature learning between the layers. The designed rectifying layer is achieved by $\boldsymbol{X}_{mid_3} = \phi_\tau(\boldsymbol{X}_{k-1})$, where the activation function with the hard threshold is defined as

$$\phi_\tau(\boldsymbol{X}_{k-1}) = \begin{cases} \tau\boldsymbol{X}_{k-1}, & \text{if } \boldsymbol{X}_{k-1}(i,j) \in (-\tau, 0] \\ \boldsymbol{X}_{k-1}, & \text{otherwise} \end{cases} \tag{8}$$

where τ is an activation threshold. For (8), different degrees of nonlinearity are achieved by fine-tuning τ, which effectively prevents nonlinearity from resulting in the over-sparsity of the feature matrix. Specifically, since the resulting eigenmatrix of the rectifying layer is Euclidean, an OrthMap layer $f_{OrthMap}$ is then applied to convert the result of (8) back to an orthonormal matrix, which is denoted $\boldsymbol{X}_k = f_{OrthMap}(\boldsymbol{X}_{mid_3}) = \boldsymbol{U}_{mid_3,1:q}$. Therein, $\boldsymbol{U}_{mid_3,1:q}$ is the top-q largest eigenvectors obtained from the eigenvalue decomposition of the input projection matrix $\boldsymbol{X}_{mid_3} = \boldsymbol{U}_{mid_3}\boldsymbol{\Sigma}_{mid_3}\boldsymbol{U}_{mid_3}^T$. Here, q is set to 10 [16].

C. Discriminative Dual Attention

Abnormal brain lesions tend to occur in a few localized regions, especially in the early stages of neurological disorders. To enhance the discrimination power to deal with complex background, a Grassman manifold attention is designed to improve its performance. Here, two different attention schemes are embedded into two consecutive learning stages: multiprojection attention mechanism (MPAM) and group-dependent discriminative attention mechanism (GDAM) stages. Both MPAM and GDAM are designed to work on the Grassmann manifolds. An intuitive illustration on the two manifold attention blocks is shown in Fig. 2. Detailed descriptions of MPAM and GDAM are provided below.

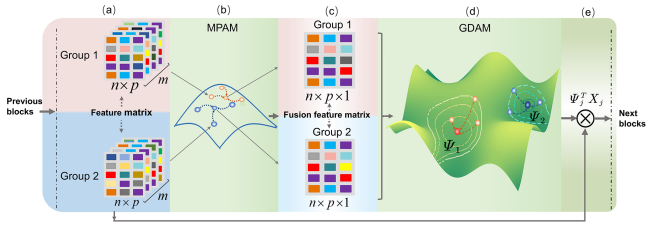

Fig. 2. Illustration of the execution details of MPAM and GDAM.

1) Using the MPAM Block to Enhance Interindividual Representations: Here, the MPAM block generates a multi-projection matrix attention by exploiting the interspatial relationship between features to enhance the inter-individual significance characteristics of brain networks. It is done by computing $Fréchet$ averaging along the channel axis of the m orthonormal projection matrix $\{\boldsymbol{X}_{k-1}^i \mid 1 \le i \le m\}$ to generate a centralized representation; thus noise immunization occurs in manifold space, as shown in Fig. 2(a)–(c). It can be formulated as follows:

$$\boldsymbol{X}_k = \arg\min_{\boldsymbol{X}_k} \sum_{i=1}^{m} (q - tr((\boldsymbol{X}_{k-1}^i)^T \boldsymbol{X}_k \boldsymbol{X}_k^T \boldsymbol{X}_{k-1}^i)) \quad (9)$$

The solution in (9) along the channel axis yields a result that is effective at highlighting the critical feature matrix.

2) Using the GDAM Block to Enhance Global Discrimination Power: Given the inadequacy of MPAM, our goal is to learn group-dependent discriminant matrices $\boldsymbol{\Psi}_j$ that map multiprojection orthonormal matrices into a more differentiated space. Specifically, the GDAM block effectively contains globally important intergroup difference information. To achieve this objective, embedding the Fisher criteria to simultaneously maximize the distance of intergroup samples and minimize the distance of intragroup samples is attempted; the relevant details are shown in Fig. 2(c)–(d). Specifically, the objective function of GDAM is constructed by trace difference [24] as follows:

$$\min_{\boldsymbol{\Psi}_j} \sum_{j=1}^{2} \sum_{\boldsymbol{X}_k \in C_j} tr[S_w - \lambda S_b] \quad (10)$$

where λ is a scalar used to balance intragroup and intergroup terms. S_b and S_w define the intergroup dispersion and intragroup compactness, formulated as follows:

$$S_b = q - tr(\boldsymbol{\Psi}_1^T \boldsymbol{\Psi}_2 \boldsymbol{\Psi}_2{}^T \boldsymbol{\Psi}_1) \quad (11)$$

$$S_w = q - \sum_{j=1}^{2} \sum_{\boldsymbol{X}_k \in C_j} tr(\boldsymbol{X}_k^T \boldsymbol{\Psi}_j \boldsymbol{\Psi}_j{}^T \boldsymbol{X}_k) \quad (12)$$

In our Grassmann deep learning network, a Fisher criterion for Grassmann metric learning is employed; it aims to learn a discriminative space, thereby reducing and clarifying the example density of class boundaries, as shown in Fig. 2(e). Importantly, the group-dependent discriminant matrix $\boldsymbol{\Psi}_j$ obtained by GDAM acts as a projection matrix layer that allows the input to be mapped to a more discriminative manifold space, defined as $\boldsymbol{X}_{k,j} = \boldsymbol{\Psi}_j^T \boldsymbol{X}_{k-1,j}, j \in C_j$.

Unlike the traditional deep attention mechanism, our proposed attention mechanism not only considers the higher-order geometric relationships between individual brain networks but also avoids disrupting the spatially nonlinear structure of brain networks. Additionally, GDAM can emphasize the group-dependent feature representation and mitigate the interference of noise, which guides the whole model to learn more discriminative features. Thus, GDAM can potentially enhance the discrimination performance of the model on highly heterogeneous brain networks.

D. Optimization Scheme

The Grassmann deep learning framework is optimized by an alternative training strategy for two main reasons: (1) the attention mechanism consisting of MPAM and GDAM is updated based on an internal manifold iterative optimization strategy. In other words, this part requires only forward propagation to update the group-dependent discriminant matrix $\boldsymbol{\Psi}_j$ to guide the classification of the model. (2) Because the DA-DMHD model is represented by a complex matrix decomposition, model backpropagation cannot be achieved simply via elementwise operations in matrix form. As a result, the DA-DMHD model adopts the matrix backpropagation optimization strategy of [16] and [25]. The optimization scheme for matrix backpropagation is minutely described in the study [16]. Now, the updated strategy for the attention mechanism is further discussed.

Recalling Section III-C, the difficulty in solving (9) and (10) in our model arises from the fact that the closed solutions of the MPAM result $\boldsymbol{X}_k$ and the group-dependent discriminant matrix $\boldsymbol{\Psi}_j$ are unknown. Hence, this paper employs the Weiszfeld algorithm to overcome this challenge. Without loss of generality, the optimization process of MPAM and GDAM is summarized as follows.

1) Optimization of MPAM: The MPAM block finds the latent geometric mean X_k on the Grassmann manifolds that has the shortest geodesic distance from the multiprojection orthonormal matrix $\{\boldsymbol{X}_{k-1}^i \mid 1 \le i \le m\}$ located on the Grassmann manifolds. Therefore, the solution to the problem is obtained by using the Grassmann manifold gradient descent optimization method, which consists of four iterative steps:

1) Initialize $\boldsymbol{X}_k^{(t)}$ (t^{th} iteration) as the center points on the current Grassmann manifolds.
2) Calculate the gradient $\nabla_{\boldsymbol{X}_k}$ of the energy function in (9) with respect to each multi-projection orthonormal matrix $\boldsymbol{X}_{k-1}^i$ (i^{th} channel) on the current estimation of the manifold center $\boldsymbol{X}_k^{(t)}$ through (2). In this way, the mean tangent $\Delta \boldsymbol{X}_k^{(t+1)} \in \mathcal{T}_{\boldsymbol{X}_k^{(t)}}$ can be obtained by the following equations:

$$\begin{aligned} \Delta \boldsymbol{X}_k^{(t+1)} &= -\sum_{i=1}^{m} \nabla_{\boldsymbol{X}} d^2 \left(\boldsymbol{X}_k^{(t)}, \boldsymbol{X}_{k-1}^i\right) \\ &= 2 \sum_{i=1}^{m} \left(\boldsymbol{I}_q - \boldsymbol{X}_k^{(t)} \left(\boldsymbol{X}_k^{(t)}\right)^T\right) \\ &\quad \times \boldsymbol{X}_{k-1}^i \left(\boldsymbol{X}_{k-1}^i\right)^T \boldsymbol{X}_k^{(t)} \end{aligned} \quad (13)$$

3) Map the mean tangent $\Delta \boldsymbol{X}_k^{(t+1)}$ back to the Grassmann manifolds through $\boldsymbol{X}_k^{(t+1)} = \exp_{\boldsymbol{X}_k^{(t)}}(\Delta \boldsymbol{X}_k^{(t+1)})$ to update $\boldsymbol{X}_k^{(t+1)}$.
4) Iteratively perform steps (2)–(3) until convergence.

2) Optimization of GDAM: In our attention block, GDAM follows the idea of Fisher's criterion to learn a space such that

examples from the self-same group are closer and examples from different classes are more distant. Thus, GPAM extends the intergroup regularity term compared to MPAM, *i.e.*, the objective function is specified as follows:

$$\begin{aligned}&\min_{\boldsymbol{\Psi}_j}\sum_{j=1}^{2}\sum_{\boldsymbol{X}_k\in C_j} d^2(\boldsymbol{X}_k,\boldsymbol{\Psi}_j)-\lambda d^2(\boldsymbol{\Psi}_1,\boldsymbol{\Psi}_2)\\&=\min_{\boldsymbol{\Psi}_j}\sum_{j=1}^{2}\sum_{\boldsymbol{X}_k\in C_j}\left(q(1-\lambda)-tr(\boldsymbol{X}_k^T\boldsymbol{\Psi}_j{\boldsymbol{\Psi}_j}^T\boldsymbol{X}_k)\right)\\&\quad+\lambda tr(\boldsymbol{\Psi}_1^T\boldsymbol{\Psi}_2{\boldsymbol{\Psi}_2}^T\boldsymbol{\Psi}_1)\end{aligned}\tag{14}$$

Conventionally, the following two steps are alternated until convergence.

1) Given $\boldsymbol{\Psi}_j^{(t)}$ as the initial manifold evaluation center, we calculate the gradient $\nabla_{\boldsymbol{\Psi}}$ of (14) by (2). Then, the group-dependent discriminant matrix $\boldsymbol{\Psi}_j^{(t+1)}$ of the tangent $\Delta\boldsymbol{\Psi}_j^{(t+1)}\in\mathcal{T}_{\boldsymbol{\Psi}_j^{(t)}}$ can be obtained by

$$\begin{aligned}\Delta\boldsymbol{\Psi}_1^{(t+1)}=&-2\sum_{\boldsymbol{X}_k\in C_1}\left(\boldsymbol{I}_q-\boldsymbol{\Psi}_1^{(t)}(\boldsymbol{\Psi}_1^{(t)})^T\right)\boldsymbol{X}_k\boldsymbol{X}_k^T\boldsymbol{\Psi}_1^{(t)}\\&-\lambda\left(\boldsymbol{I}_q-\boldsymbol{\Psi}_1^{(t)}(\boldsymbol{\Psi}_1^{(t)})^T\right)\boldsymbol{\Psi}_2^{(t)}(\boldsymbol{\Psi}_2^{(t)})^T\boldsymbol{\Psi}_1^{(t)}\end{aligned}\tag{15}$$

$$\begin{aligned}\Delta\boldsymbol{\Psi}_2^{(t+1)}=&-2\sum_{\boldsymbol{X}_k\in C_2}\left(\boldsymbol{I}_q-\boldsymbol{\Psi}_2^{(t)}(\boldsymbol{\Psi}_2^{(t)})^T\right)\boldsymbol{X}_k\boldsymbol{X}_k^T\boldsymbol{\Psi}_2^{(t)}\\&-\lambda\left(\boldsymbol{I}_q-\boldsymbol{\Psi}_2^{(t)}(\boldsymbol{\Psi}_2^{(t)})^T\right)\boldsymbol{\Psi}_1^{(t)}(\boldsymbol{\Psi}_1^{(t)})^T\boldsymbol{\Psi}_2^{(t)}\end{aligned}\tag{16}$$

2) Map the mean tangent $\Delta\boldsymbol{\Psi}_j^{(t+1)}$ back to the Grassmann manifolds through $\boldsymbol{\Psi}_j^{(t+1)}=\exp_{\boldsymbol{\Psi}_j^{(t)}}\left(\Delta\boldsymbol{\Psi}_j^{(t+1)}\right)$ to update $\boldsymbol{\Psi}_j^{(t+1)}$.

Finally, the above alternating optimization strategy leads to the robust group-dependent discriminant matrix $\boldsymbol{\Psi}_j$, which provides clearer decision boundaries for subsequent classification.

E. Classification

In the testing phase, the given test data are first encoded using harmonic waves to make each brain network act as a point on the Grassmann manifold. The attention block is inactive during this phase, operating only during training by guiding the model to learn the discriminative Riemannian manifold representation. Subsequently, an output layer consisting of fully connected and softmax layers is used to perform brain network classification.

IV. Experiments

A. Data Preparation

The performance of the proposed DA-DMHD method is tested on two independent real datasets. One is the Alzheimer's Disease Neuroimaging Initiative dataset (ADNI), which includes structural brain connectivity from diffusion tensor images (DTIs) and functional connectivity from resting-state functional magnetic resonance imaging (rs-fMRI). Another is ADHD-200 on the functional connectivity from rs-fMRI, including the automatic anatomical labeling (AAL) atlas, Craddock's clustering 200 (CC200) atlas and Harvard-Oxford (HO) atlas. Here, the ADNI and ADHD-200 data are first constructed as brain networks (i.e., the preprocessing stage, as shown in Fig. 1).

ADNI about structural connectivity: A total of 506 subjects (including 168 cognitively normal (CN) subjects (77 males, 91 females; 73.7±5.5 years (yrs)), 167 subjects with mild cognitive impairment (MCI) (110 males, 57 females; 72±7.5 yrs), and 171 subjects with Alzheimer's disease (AD) (96 males, 75 females; 74.1±6.9 yrs)) from the raw ADNI database were selected. Each subject was scanned using T1-weighted MRI and diffusion-weighted MRI (DTI). To construct the adjacency matrix of the brain network, each DTI is first processed using an internal tractography pipeline to extract the structural brain network via the Destrieux atlas [26] with 148 regions of interest (ROIs). Then, each brain network generated a 148×148 connectivity matrix by applying surface seed-based probabilistic fiber tractography on the DTI data. Because detecting subtle differences between MCI and other stages is helpful in the diagnosis of brain diseases, the ADNI data were divided into three groups, CN vs. MCI, AD vs. MCI, and CN vs. AD, in follow-up experiments.

ADNI about functional connectivity: rs-fMRI images of 465 subjects (including 282 CN subjects (110 males, 172 females; 73.1±7.6 yrs) and 183 subjects with MCI (104 males, 79 females; 74.8±7.7 yrs)) from the raw ADNI data were collected. First, the standard preprocessing procedure used in [27] and [28] was employed. Then, spatial smoothing, temporal time correction, time prewhitening, global drift removal and bandpass filtering (0.01-0.1 Hz) were applied to the rs-fMRI images via the FEAT command in the FMRIB Software Library (FSL) [29]. Finally, the Destrieux Atlas was adopted to construct 148 ROIs and calculated different ROIs by Pearson correlation coefficients (PCCs) to construct functional brain networks.

ADHD-200: By collecting the sample without considering artifacts and removing subjects with missing values and/or bad spatial normalization, 756 subjects were obtained from the raw ADHD-200 database. Each subject was registered separately by the AAL, CC200 and HO atlases. Then, different ROIs were calculated via PCCs to construct the functional brain connectivity for each subject. The ADHD-200 dataset has three labels, namely, typically developing children (TDC), ADHD combined (ADHD-C) type or ADHD inattentive (ADHD-I) type [30], but binary diagnosis tasks are the focus of this work. Here, ADHD-C and ADHD-I were merged into the ADHD group to implement the binary diagnostic task of ADHD vs. TDC.

Then, the structural or functional brain networks are encoded as harmonic waves by eigen-decomposition and used as input to the model.

B. Comparative Methods and Settings

In this paper, the proposed algorithm is methodically compared with some representative brain network classification methods and other state-of-the-art classification methods, which are broadly classified into the following five categories.

- **Traditional machine learning-based methods:** Linear discriminant analysis (LDA), random forest (RF), and AdaBoost (ADB) [7].
- **Multiple statistical feature-based methods:** Multiple manifold metric learning (MMML) [31] and multikernel manifold metric learning (MKMML) [32].

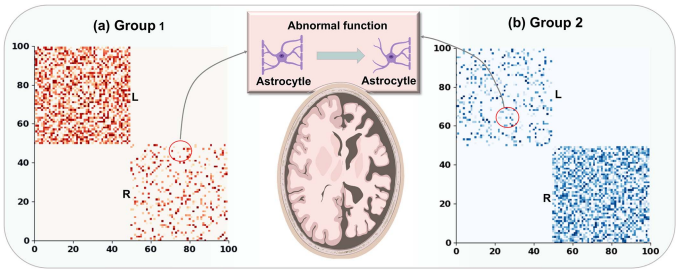

Fig. 3. Schematic diagram of the hemispheric asymmetric synthesis data. L and R denote the left and right hemispheres of the brain, respectively.

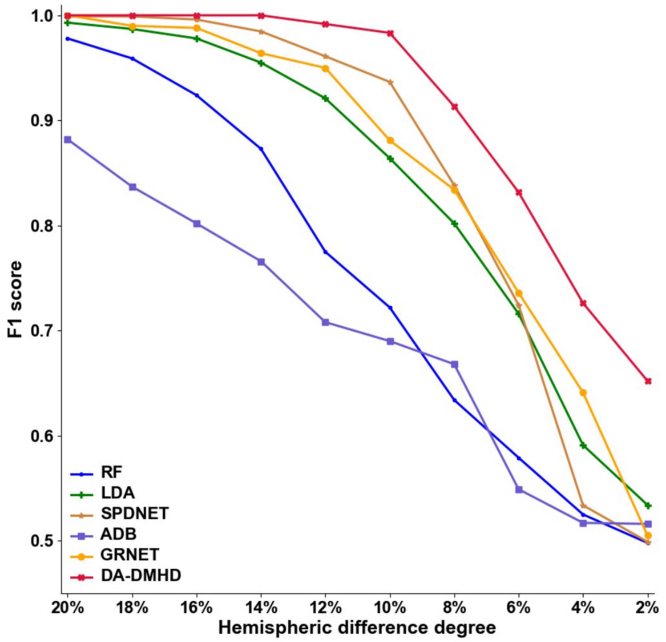

Fig. 4. F1-score of different methods on hemispheric asymmetric data with different sparsity. The horizontal coordinate indicates the degree of difference between the two groups of synthetic data.

- **Graph-based learning method:** Dynamic graph convolutional neural network (DGCNN) [9].
- **Deep learning-based methods:** Deep neural network (DNN) [8], DNN joint center loss (DNNCL) [33], convolutional neural network (CNN) [10], simple symmetric positive definite (SPD) manifold deep learning network (SYMNET, including SYMNET1 and SYMNET2) [17], Grassmann manifold network (GRNET) [16], and SPD network (SPDNET) [15], and Multilevel functional connectivity fusion classification framework (MFC) [34].
- **High-order tensor-based method:** Higher-order singular value decomposition (SVD) with sparse logistic regression (HOSVD) [35].

In the experiments, the proposed method is tested on two publicly available neurodegenerative disease datasets (i.e., ADNI and ADHD-200). In dataset division, 80% of the original data was divided into a training set and 20% was divided into a test set. Then, $5-$fold cross-validation was performed on the training set to evaluate our method and comparative methods. Then, the trained model was applied on the tested dataset. The experiment was repeated independently ten times. The averaged results were stored to evaluate the statistical power of our method over comparative ones. The average of five evaluation metrics was adopted to measure the performance of the algorithm: accuracy (ACC), recall (REC), precision (PRE), area under the ROC curve (AUC), and F1-score. To ensure the fairness of the experiment, the parameters of DA-DMHD and the other comparison methods were empirically adjusted. Notably, the MFC method was tailored for functional brain networks, so MFC was tested only on the dataset including the functional brain network. Additionally, a paired-sample T-test was performed on our proposed method and the suboptimal comparison methods. $\star$ was used to indicate that our results are significantly different ($P < 0.05$) compared to those of the suboptimal algorithm. In the supplementary material, a selection of parameters for the DA-DMHD and comparison methods is provided.

C. Experiment on Synthetic Data

Clinical neuroscience broadly recognizes lateralization and asymmetry of the brain in many psychiatric and neurological disorders [36]. Considering this, synthetic data with similar hemispheric asymmetry were designed without losing the neuroscientific context, as shown in Fig. 3. Here, the NetworkX toolkit [37] was applied to construct two groups of hemispheric lesion difference samples (i.e., the left brain or the right brain differs in sparsity), shown in (a) and (b) of Fig. 3. With the hypothesis of neurological disorders, two significant purposes exist for the synthetic data constructed: to examine the classification performance of DA-DMHD and to assess the stability of DA-DMHD for unbalanced samples. Correspondingly, traditional machine learning (TML) methods, SPDNET and GRNET, are used as the main comparison methods for this experiment. Notably, both GRNET and DA-DMHD inputs are harmonic waves [18], while the SPDNET input is the symmetric positive definite form of the brain network and TML methods use vectors as input.

1) Classification Performance Evaluation of DA-DMHD: In this part, multiple brain lateralization data were constructed to validate the performance of the proposed DA-DMHD algorithm in classifying high-dimensional network data. Here, it was assumed that the synthetic data are sample-balanced, i.e., each group contains 1000 samples. Then, the left hemisphere sparsity of group 1 is set to a fixed value, while the right hemisphere sparsity is set to an adjustable value. Therein, the sparsity of the right hemispheric is adjusted to make the degree of difference from the left hemispheric in the range of [20%,2%] with a step of 2%. For group 2, the sparsity of the left and right hemispherics is set in the opposite way to that of group 1. As the sparsity of the two groups of networks changes, the heterogeneity and noise between the groups also change, which in turn affects the final classification effect. Finally, the classification power of the DA-DMHD method and the comparison method was discussed by visualizing the F1-score of each algorithm.

From the overall view of Fig. 4, the classification accuracy of our proposed method is better than that of the selected comparison method. Geometrically, traditional machine learning methods shrink each object into feature vectors for subsequent classification tasks without considering the relationship

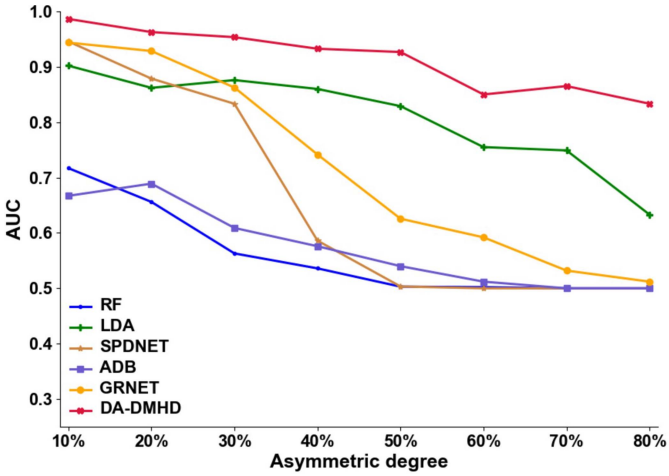

Fig. 5. AUC scores of different methods in unbalanced samples.

between different edges, resulting in poor classification results. In contrast, this relational property of the brain was indirectly exploit through manifolds, so the classification performances of SPDNET, GRNET, and DA-DMHD were better. The methods using harmonic waves as input (including GRNET and DA-DMHD) outperformed other forms of input, which further illustrates the effectiveness of harmonic waves as brain network embedding. Moreover, a geometric attention mechanism was integrated into the model to optimize the spacing between different manifold data, which enhanced the discriminative power of the model. More important, DA-DMHD achieved an F1-score of 0.6 or more, despite only a 2% degree of difference between the two groups of synthetic brain hemisphere data. As a result, this shows that our proposed model can effectively capture the microstructure between individual brain networks and reduce intergroup heterogeneity or/and noise, thereby reducing the rate of misdiagnosis.

2) Performance Testing of the Sample Imbalance Algorithm: Based on the above observations, the combination of Grassmann manifold geometry and the proposed attention block can effectively improve the classification ability of the model. However, this conclusion is currently limited to the ideal sample balance stage. Therefore, to further verify the reliability and robustness of DA-DMHD, several unbalanced sample experiments were constructed. (1) In this experiment, the degree of difference between the left and right hemispheric of group 1 and group 2 was fixed at 10%. (2) Group 1 was fixed to be 1000 samples, while setting the sample size of group 2 to range from 90% to 20% of group 1 with a 10% decreasing step. (3) To demonstrate the reliability of the DA-DMHD method more intuitively, AUC diagrams were drawn for the DA-DMHD method and the comparison method on this synthetic dataset, as shown in Fig. 5.

As shown in Fig. 5, the proposed DA-DMHD method achieves much better reliability on the unbalanced samples than other methods in all cases, while maintaining an AUC score of 0.8 or higher. Notably, SPDNET and GRNET perform poorly in addressing extremely imbalanced samples. Nevertheless, the harmonic-based GRNET method is still much better than SPDNET, suggesting that the model's performance can be improved by building the network from harmonic waves. In contrast, the classification performance of LDA with discriminative information tends to be stable. The above analysis implies that SPDNET and GRNET have difficultly achieving high reliability with imbalanced samples via low-dimensional mapping alone. Thus, deep Grassmann models need to be not only separable but also discriminative. Therefore, the MPAM block was introduced to integrate low-dimensional features to highlight the more essential separability features. In addition, a group-dependent low-dimensional discriminant projection matrix is obtained by introducing the GDAM block to improve the discriminability of the model. In summary, the synthetic experiments show that our proposed two blocks can effectively improve the reliability and robustness of the deep Grassmann model.

D. Results of Brain Disease Classification

To understand the effectiveness of our method on real high-dimensional brain networks, various experiments on two neurological disease datasets that contain ADNI and ADHD-200 were conducted, as described before. The bolded and underlined values in the statistical table are the optimal and suboptimal values, respectively. The dark gray background shows the deep learning methods in the statistics table.

1) Classification Performance on ADNI: During this subset of experiments, DA-DMHD on ADNI data are tested (including structural and functional connectivity). Moreover, classification for the ADNI dataset is particularly challenging since the physiological distinction of MCI lies at the decision boundary between clinical CN and AD, which is incredibly subtle in terms of alterations in brain networks. The experimental results of all methods are summarized in Table II.

Among the classification results reported in Table II, the accuracy obtained by the proposed DA-DMHD method for all the classification tasks is consistently the highest among all methods. These results suggest that our proposed DA-DMHD model combined with nonlinear and Fisher discriminative attention mechanisms can be used to improve the diagnosis of brain diseases within the brain network-based study.

Furthermore, the classification results produced by SYMNET1/SYMNET2 in the ADNI classification task are both inferior to those of traditional machine learning methods. The primary reason may be that SYMNET1/SYMNET2 was initially designed for image and video recognition tasks, making it difficult to extract the underlying structural information in the brain network to improve the classification performance. Furthermore, the classification performances of SPDNET, MMML, and MR-MML on the ADNI dataset are lower than the classification performance of GRNET. Meanwhile, GRNET achieves better classification accuracy than the DGCNN and CNN algorithms on the AD vs. MCI and CN vs. AD classification tasks for the structural network. For the functional connectivity of CN vs. MCI classification task, the GRNET and DA-DMHD methods based on harmonic waves construction outperformed most vector and symmetric matrix-based methods. Interestingly, this is consistent with our conclusions in the synthesis experiments. Thus, harmonic waves are more effective than using brain network data directly.

Another interesting observation is that the vector-form DNN brain network classification method outperforms most matrix-form feature-based learning methods. This result is mainly because the DNN input is preprocessed to contain highly discriminative feature vectors, which improves the model's ability to discriminate between individual brain networks. Paradoxically,

TABLE II
CLASSIFICATION PERFORMANCES ON ADNI

Task/Modal	Method	ACC	REC	PRE	AUC	F1-score
CN vs. MCI Structural Connectivity	RF	0.794±0.016	0.774±0.064	0.822±0.073	0.796±0.021	0.784±0.019
	ADB	0.709±0.058	0.708±0.065	0.705±0.054	0.706±0.055	0.704±0.061
	LDA	0.678±0.035	0.800±0.040	0.686±0.101	0.678±0.042	0.663±0.071
	HOSVD	0.607±0.117	0.628±0.181	0.730±0.056	0.617±0.115	0.635±0.106
	MMML	0.688±0.094	0.706±0.117	0.651±0.134	0.651±0.133	0.673±0.114
	MRMML	0.547±0.146	0.554±0.162	0.549±0.151	0.571±0.128	0.551±0.155
	DGCNN	0.787±0.083	0.725±0.127	0.828±0.101	0.787±0.084	0.770±0.102
	DNNCL	0.821±0.044	0.806±0.048	0.831±0.063	0.822±0.045	0.828±0.053
	DNN	0.824±0.025	0.797±0.139	**0.856±0.046**	0.828±0.039	0.819±0.039
	CNN	0.769±0.015	**0.844±0.015**	0.730±0.056	0.765±0.014	0.782±0.036
	SYMNET1	0.553±0.058	0.565±0.073	0.572±0.089	0.575±0.053	0.565±0.073
	SYMNET2	0.532±0.069	0.544±0.075	0.569±0.095	0.553±0.086	0.552±0.070
	SPDNET	0.736±0.020	0.735±0.020	0.769±0.076	0.701±0.122	0.722±0.005
	GRNET	0.764±0.015	0.760±0.036	0.784±0.033	0.779±0.009	0.771±0.009
	DA-DMHD	**0.827±0.029**	0.821±0.023	0.833±0.059	**0.847±0.034***	**0.858±0.030***
AD vs. MCI Structural Connectivity	RF	0.721±0.043	0.668±0.126	0.763±0.081	0.728±0.079	0.690±0.041
	ADB	0.734±0.056	0.734±0.112	0.724±0.041	0.737±0.057	0.718±0.042
	LDA	0.670±0.077	0.653±0.056	0.661±0.087	0.671±0.071	0.653±0.070
	HOSVD	0.615±0.066	0.722±0.037	0.629±0.101	0.635±0.053	0.622±0.080
	MMML	0.779±0.072	0.850±0.064	0.703±0.084	0.752±0.048	0.769±0.076
	MRMML	0.727±0.079	0.796±0.070	0.642±0.120	0.706±0.087	0.707±0.096
	DGCNN	0.779±0.059	0.706±0.185	0.821±0.050	0.774±0.068	0.745±0.123
	DNNCL	0.817±0.076	0.847±0.114	0.804±0.064	0.814±0.064	0.822±0.080
	DNN	0.831±0.052	**0.879±0.066**	0.816±0.102	0.840±0.046	0.839±0.048
	CNN	0.743±0.015	0.727±0.058	0.764±0.084	0.749±0.017	0.739±0.013
	SYMNET1	0.668±0.031	0.686±0.073	0.635±0.055	0.722±0.032	0.655±0.034
	SYMNET2	0.562±0.064	0.576±0.079	0.537±0.139	0.605±0.074	0.540±0.090
	SPDNET	0.739±0.031	0.739±0.031	0.733±0.054	0.764±0.049	0.746±0.036
	GRNET	0.821±0.023	0.863±0.026	0.771±0.042	0.860±0.012	0.813±0.027
	DA-DMHD	**0.856±0.032***	0.872±0.043	0.856±0.023	**0.873±0.037**	**0.873±0.026***
CN vs. AD Structural Connectivity	RF	0.753±0.039	0.721±0.115	0.721±0.164	0.757±0.041	0.735±0.031
	ADB	0.702±0.081	0.750±0.116	0.699±0.075	0.709±0.082	0.699±0.075
	LDA	0.728±0.043	0.750±0.039	0.720±0.041	0.726±0.043	0.711±0.064
	HOSVD	0.622±0.040	0.749±0.115	0.659±0.167	0.625±0.080	0.654±0.079
	MMML	0.738±0.070	0.770±0.076	0.719±0.069	0.714±0.077	0.743±0.069
	MRMML	0.712±0.077	0.753±0.065	0.693±0.096	0.750±0.095	0.720±0.080
	DGCNN	0.828±0.063	0.815±0.134	0.860±0.108	0.820±0.074	0.822±0.062
	DNNCL	0.881±0.019	0.762±0.043	0.899±0.050	0.881±0.020	0.878±0.022
	DNN	0.893±0.066	0.865±0.095	0.914±0.052	0.895±0.068	0.888±0.070
	CNN	0.810±0.030	0.721±0.131	0.878±0.049	0.801±0.043	0.781±0.071
	SYMNET1	0.669±0.059	0.772±0.089	0.614±0.075	0.750±0.061	0.676±0.037
	SYMNET2	0.621±0.051	0.665±0.083	0.592±0.069	0.687±0.089	0.622±0.049
	SPDNET	0.742±0.030	0.738±0.040	0.766±0.049	0.679±0.114	0.716±0.075
	GRNET	0.849±0.038	0.870±0.038	0.824±0.075	0.860±0.047	0.844±0.044
	DA-DMHD	**0.903±0.013**	**0.885±0.013***	**0.934±0.029***	**0.918±0.011***	**0.915±0.014***
CN vs. MCI Functional Connectivity	RF	0.613±0.033	0.964±0.034	0.626±0.027	0.539±0.025	0.746±0.028
	ADB	0.586±0.032	0.938±0.063	0.630±0.032	0.538±0.038	0.732±0.030
	LDA	0.562±0.025	0.685±0.038	**0.827±0.025**	0.530±0.014	0.652±0.038
	HOSVD	0.619±0.036	**0.998±0.002**	0.643±0.034	0.563±0.046	0.754±0.008
	MMML	0.589±0.047	0.754±0.027	0.719±0.069	0.564±0.048	0.707±0.047
	MRMML	0.572±0.055	0.686±0.032	0.696±0.097	0.519±0.024	0.690±0.061
	DGCNN	0.641±0.023	0.953±0.042	0.635±0.027	0.553±0.027	0.762±0.026
	DNNCL	0.606±0.060	0.987±0.006	0.606±0.069	0.503±0.002	0.753±0.020
	DNN	0.624±0.049	0.967±0.061	0.625±0.050	0.529±0.050	0.757±0.020
	CNN	0.626±0.050	0.996±0.004	0.620±0.046	0.527±0.024	0.763±0.034
	MFC	0.663±0.017	0.901±0.125	0.683±0.071	**0.600±0.059**	0.753±0.050
	SYMNET1	0.621±0.041	0.838±0.067	0.674±0.080	0.568±0.022	0.743±0.035
	SYMNET2	0.631±0.048	0.796±0.073	0.719±0.072	0.551±0.083	0.753±0.051
	SPDNET	0.652±0.019	0.901±0.062	0.664±0.014	0.567±0.021	0.763±0.019
	GRNET	0.659±0.061	0.956±0.056	0.668±0.065	0.540±0.041	**0.784±0.049**
	DA-DMHD	**0.665±0.012**	0.926±0.077	0.662±0.006	0.589±0.028	0.771±0.024

TABLE III
CLASSIFICATION PERFORMANCES ON ADHD

Task/Atlas	Method	ACC	REC	PRE	AUC	F1-score
TDC vs. ADHD AAL	RF	0.646±0.056	**0.997±0.002**	0.660±0.072	0.528±0.044	0.782±0.040
	ADB	0.650±0.058	**0.997±0.002**	0.689±0.082	0.579±0.073	0.781±0.042
	LDA	0.622±0.042	0.810±0.057	0.685±0.093	0.555±0.043	0.730±0.035
	HOSVD	0.661±0.041	0.963±0.024	0.695±0.030	0.584±0.026	0.784±0.016
	MMML	0.517±0.087	0.594±0.087	0.700±0.186	0.568±0.078	0.636±0.114
	MRMML	0.475±0.087	0.571±0.048	0.615±0.169	0.504±0.132	0.587±0.104
	DGCNN	0.661±0.035	0.925±0.061	0.670±0.035	0.549±0.023	0.777±0.038
	DNNCL	0.624±0.023	0.761±0.030	0.691±0.053	0.569±0.053	0.722±0.020
	DNN	0.667±0.023	0.781±0.037	0.722±0.022	0.618±0.017	0.750±0.037
	CNN	0.684±0.023	0.915±0.047	0.696±0.036	0.593±0.035	0.788±0.013
	MFC	0.698±0.025	0.834±0.073	0.726±0.027	0.642±0.020	0.775±0.036
	SYMNET1	0.530±0.036	0.618±0.024	0.668±0.066	0.578±0.041	0.641±0.037
	SYMNET2	0.562±0.032	0.635±0.019	0.724±0.049	0.585±0.050	0.676±0.028
	SPDNET	0.668±0.028	0.970±0.029	0.668±0.039	0.670±0.029	0.790±0.020
	GRNET	0.644±0.011	0.644±0.009	**0.979±0.013**	0.612±0.021	0.777±0.004
	DA-DMHD	**0.709±0.036**	0.693±0.016	0.895±0.070	**0.671±0.024**	**0.796±0.031**
TDC vs. ADHD HO	RF	0.636±0.015	0.783±0.075	0.661±0.029	0.543±0.022	0.760±0.049
	ADB	0.631±0.032	**0.998±0.002**	0.656±0.014	0.536±0.028	0.773±0.021
	LDA	0.611±0.033	0.794±0.049	0.623±0.021	0.539±0.022	0.747±0.034
	HOSVD	0.595±0.048	0.613±0.052	**0.708±0.044**	0.589±0.051	0.656±0.054
	MMML	0.591±0.021	0.899±0.044	0.621±0.010	0.527±0.011	0.735±0.016
	MRMML	0.545±0.073	0.664±0.069	0.597±0.106	0.527±0.048	0.627±0.088
	DGCNN	0.657±0.035	0.918±0.042	0.666±0.026	0.565±0.023	0.772±0.018
	DNNCL	0.628±0.031	0.984±0.015	0.665±0.014	0.552±0.038	0.769±0.025
	DNN	0.669±0.015	0.856±0.092	0.702±0.024	0.609±0.034	0.768±0.032
	CNN	0.650±0.035	0.897±0.069	0.662±0.029	0.564±0.021	0.776±0.010
	MFC	0.674±0.032	0.877±0.083	0.689±0.036	0.596±0.044	0.770±0.038
	SYMNET1	0.593±0.033	0.874±0.091	0.628±0.019	0.491±0.033	0.729±0.033
	SYMNET2	0.625±0.022	0.918±0.014	0.645±0.017	0.519±0.024	0.753±0.045
	SPDNET	0.657±0.027	0.941±0.037	0.660±0.020	0.557±0.036	0.775±0.017
	GRNET	0.653±0.020	0.979±0.023	0.655±0.011	0.543±0.022	**0.785±0.013**
	DA-DMHD	**0.679±0.034**	0.910±0.090	0.685±0.045	**0.613±0.075**	0.777±0.022
TDC vs. ADHD CC200	RF	0.643±0.033	**0.997±0.003**	0.645±0.031	0.526±0.026	0.776±0.026
	ADB	0.634±0.026	0.995±0.005	0.668±0.029	0.557±0.037	0.756±0.035
	LDA	0.626±0.032	0.923±0.066	0.641±0.025	0.515±0.023	0.756±0.035
	HOSVD	0.612±0.033	0.715±0.032	0.651±0.055	0.600±0.033	0.678±0.035
	MMML	0.603±0.035	0.937±0.039	0.624±0.017	0.572±0.048	0.749±0.024
	MRMML	0.515±0.074	0.597±0.071	0.580±0.067	0.503±0.036	0.599±0.067
	DGCNN	0.651±0.028	0.977±0.033	0.647±0.029	0.530±0.029	0.778±0.029
	DNNCL	0.593±0.086	0.886±0.228	0.615±0.054	0.484±0.033	0.717±0.121
	DNN	0.682±0.032	0.869±0.113	0.697±0.027	0.595±0.064	0.770±0.052
	CNN	0.660±0.013	0.934±0.049	0.665±0.018	0.564±0.021	0.760±0.049
	MFC	**0.695±0.022**	0.907±0.049	0.698±0.021	0.611±0.047	0.787±0.018
	SYMNET1	0.536±0.057	0.726±0.075	0.612±0.033	0.466±0.051	0.664±0.051
	SYMNET2	0.629±0.047	0.981±0.037	0.633±0.053	0.503±0.051	0.769±0.085
	SPDNET	0.649±0.018	0.989±0.009	0.646±0.018	0.522±0.018	0.781±0.012
	GRNET	0.669±0.030	0.998±0.003	0.658±0.021	0.549±0.042	**0.793±0.015**
	DA-DMHD	0.692±0.030	0.897±0.078	**0.707±0.032**	**0.625±0.041**	0.784±0.045

DNNCL with discriminative loss has lower classification performance than the DNN, demonstrating that the existing discriminative block in Euclidean space is not suitable for complex brain network data. By contrast, for the proposed method, the state-of-the-art classification performance confirms its effectiveness. Interestingly, our proposed attention module not only effectively improves the interpretability of the model but also enables the model to achieve an impressive classification result, with even fewer layers than traditional deep learning methods.

2) Classification Performance on ADHD-200: For further evaluation, the proposed algorithm was applied to the more challenging ADHD-200 dataset. This dataset contains adolescents with actively developing brains, resulting in high variability between groups. Moreover, it is difficult to draw a line between normal levels of ADHD symptoms and clinically significant levels that require intervention [38]. Thus, identifying differences between these groups is an inherent challenge. Furthermore, compared to previous experiments on Alzheimer's disease, the ADHD-200 classification task suffers from sample imbalance, which increases the classification difficulty for the highly heterogeneous brain network. The final average classification results are shown in Table III.

Referring to the classification results in Table III, our observations were summarized in terms of the following aspects. First, in our experiments, part of the algorithm suffers from undesirable behavior, favoring the prediction of all test instances as TDC to obtain the maximum recall while sacrificing precision. For example, the traditional machine learning methods, the DGCNN and CNN, achieve better recall with our careful tuning, but the AUC value is still below 0.6. Simultaneously, the lower classification accuracies obtained by almost all classification methods are further evidence that the ADHD classification task is challenging. Notably, the MFC method performed slightly better than DA-DMHD on the CC200 atlas of ADHD-200, but its AUC was still lower than that of DA-DMHD. Our proposed DA-DMHD outperforms SPDNET and GRNET in terms of classification, which indicates that the low-dimensional Grassmann manifold feature matrix learned by projection mapping alone is insufficient for highly heterogeneous ADHD data. This result further demonstrates that the proposed DA-DMHD method is useful in building reliable deep Grassmann geometric models while focusing on learning discriminative subspace feature representations.

E. Ablation Studies

According to the experimental results, the proposed DA-DMHD method is superior to some representative methods for brain network classification. However, the above experiments only verified the effectiveness of the algorithm in general. To further investigate the utility of the rectifying and attention layers in our method, the ADHD-200 dataset of the AAL atlas was taken as an example to perform ablation experiments with the proposed algorithm. To do this, DA-DMHD was partitioned into

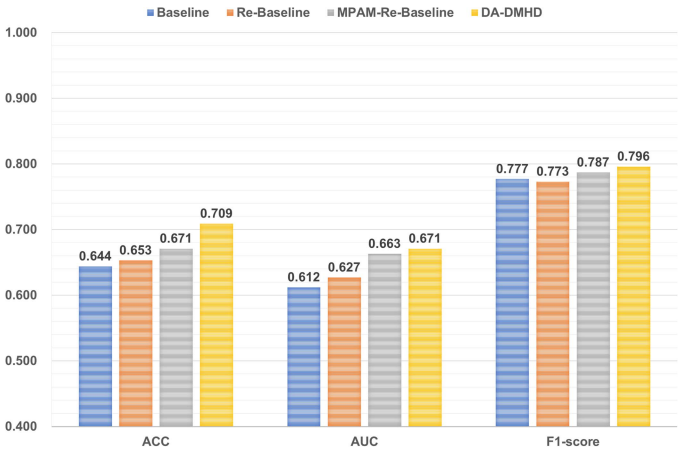

Fig. 6. Classification performance of the DA-DMHD model for the ablation experiment on the ADHD-200 dataset.

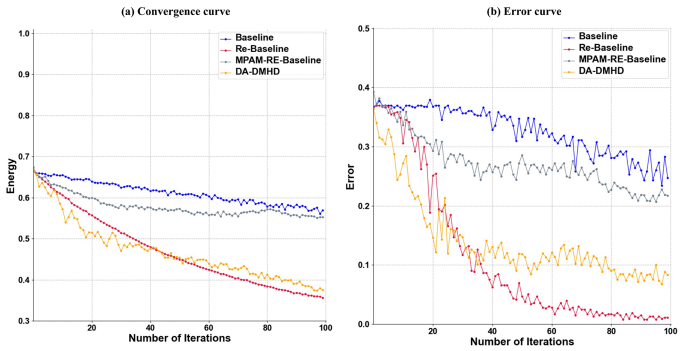

Fig. 7. Convergence performance and error trends of the DA-DMHD model for ablation experiments on the ADHD-200 dataset.

four parts: (1) GRNET (Baseline), (2) the rectifying layer and the baseline (Re-Baseline), (3) MAPM, a rectifying layer and the baseline (MPAM-Re-Baseline), and (4) DA-DMHD. Finally, the classification performance of different parts of DA-DMHD in Fig. 6 is reported via three metrics, ACC, AUC, and F1-score.

As shown in Fig. 6, the proposed DA-DMHD achieved much better ADHD-200 classification performance than the other combinations in all cases. Thus, focusing on only a single separability block (such as the FRMap layer) may not adequately represent the classified features associated with brain diseases, which often occur in multiple brain regions. Notably, the discriminative attention block and the nonlinear block of DA-DMHD can be seamlessly integrated into the Baseline. The combination of these blocks alleviates the problems of intergroup heterogeneity and intragroup ambiguity in complex brain network classification tasks, thereby highlighting the characteristic representation of lesioned brain regions. As indicated by the ACC and AUC values in Fig. 6, embedding the proposed block into the Baseline method can increase the classification performance, further illustrating the effectiveness of our proposed block. The above observations suggest that the algorithm in this paper has the potential to highlight meaningful brain disease feature representations and mitigate the negative effects (especially for sample imbalance data) to learn a more robust classification subspace.

Moreover, the convergence and error trend of the four-part algorithm for the ablation experiment are experimentally demonstrated. From Fig. 7(a)(b), the proposed blocks all achieve good convergence and learning ability. Notably, Re-baseline obtained smooth convergence results and low learning errors, which illustrates that nonlinearity enables the model to learn the topology of the brain network better and improves the separability. Paradoxically, the results in Fig. 6 indicate that overfitting occurs when introducing only the rectifying layer. The main reason is that the brain network data are highly heterogeneous intergroup data and have intragroup similarity, which increases the difficulty of algorithmic classification. To address the above issues, our proposed attentional mechanism focuses on learning discriminative brain disease-specific manifold feature representations, allowing the model to gravitate towards a more efficient resolution of the tension between intergroup heterogeneity and intragroup similarity.

TABLE IV
LIST OF TOP-10 SIGNIFICANT BRAIN CONNECTIVITIES FROM FIG. 8

ADNI (Structural Connectivity)			ADHD (AAL)		
Index	Row(ROI label)	Col(ROI label)	Index	Row(ROI label)	Col(ROI label)
1	53 (S_front_middle)	144 (S_precentral-sup-part)	1	43 (Calcarine_L)	88 (Temporal_Pole_Mid_R)
2	30 (G_precuneus)	104 (G_precuneus)	2	51 (Occipital_Mid_L)	104 (Cerebelum_8_R)
3	31 (G_rectus)	134 (S_oc-temp_lat)	3	47 (Lingual_L)	72 (Paracentral_Lobule_L)
4	2 (G_and_S_occipital_inf)	71 (S_subparietal)	4	116 (Vermis_10)	47 (Lingual_L)
5	6 (G_and_S_cingul-Ant)	65 (S_parieto_occipital)	5	51 (Occipital_Mid_L)	90 (Temporal_Inf_R)
6	104 (G_precuneus)	145 (S_subparietal)	6	7 (Frontal_Mid_L)	88 (Temporal_Pole_Mid_R)
7	45 (S_central)	139 (S_parieto_occipital)	7	30 (Insula_R)	102 (Cerebelum_7b_R)
8	98 (G_orbital)	142 (S_precentral-inf-part)	8	54 (Occipital_Inf_R)	116 (Vermis_10)
9	109 (G_temp_sup-Plan_polar)	110 (G_temp_sup-Plan_tempo)	9	105 (Cerebelum_9_L)	116 (Vermis_10)
10	25 (G_pariet_inf-Angular)	114 (Lat_Fis-ant-Vertical)	10	69 (Paracentral_Lobule_L)	57 (Postcentral_L)

F. Discussion

1) Discovering Significant Connections and Brain Regions:

The ability of clinical translation is of critical importance for computer-aided diagnosis. To do this, the connections with potential biological significance was investigated based on the brain network modeled by our proposed DA-DMHD method. These connections have been identified as potential biomarkers for imaging in early and/or advanced neurological diseases. Here, the weighted activation of the first layer is visualized, inspired by [39], to show important brain network connections and brain regions. According to (4), the connection weight matrix $\boldsymbol{W}_k \in \mathbb{R}^{n_k \times n_{k-1}} (n_k \leq n_{k-1})$ maps the input matrix (harmonic waves) to a more low-dimensional Grassmann manifold space n_k. As a result, $\boldsymbol{W}_k^T \boldsymbol{W}_k$ yields a symmetric matrix of $n_{k-1} \times n_{k-1}$ as the initial brain network representation, which can be understood as the importance of the corresponding edge in the initial matrix for the final classification. The diagonal elements of $\boldsymbol{W}_k^T \boldsymbol{W}_k$ correspond to the importance of brain regions. $\boldsymbol{W}_k^T \boldsymbol{W}_k$ was visualized as a brain network in Fig. 8. Only the top 10 connections with the largest average weights are shown, which are the most discriminative connections for CN vs. AD and TDC vs. ADHD in Fig. 8(a) and (b), respectively. Significant brain connections and node information are listed in Table IV. First, as shown in Fig. 8(a), 10 connections that span across 19 nodes (i.e., brain regions) are listed (full node labels in [26]) to verify that the model can distinguish between important brain regions for AD. From this, several temporal regions (i.e., cingulum [40], precuneus [41], and subparietal regions [42]) and many others were observed, which are consistent with the current findings in AD diagnosis. Specifically, subparietal regions [42] play a crucial role in face recognition and memory, which have been reported to have more significant brain atrophy in early AD. Additionally, most of the AD-related nodes are located in the default mode network (DMN) [43]. For instance, the angular [44] has been confirmed to interact strongly

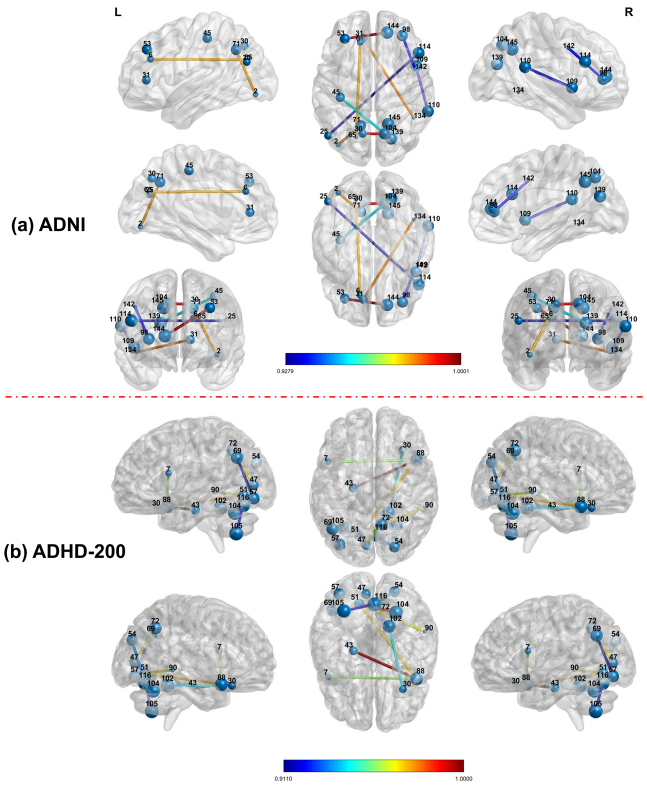

Fig. 8. Top 10 connectivities from ADNI (structural connectivity) (a) and ADHD-200(AAL) (b) analyses. The nodes represent the corresponding brain regions, whose size indicates the importance of the node, with larger regions being more important. The color of the connecting lines indicates the importance of each edge.

with other large-scale brain networks, suggesting that it may be involved in information integration over the whole brain.

Next, Fig. 8(b) gives 10 connections that span across 15 nodes (full node names are in [45]). From Fig. 8(b), several temporal regions (midtemporal pole [46], and cerebellum [47]) are shown; the results of these visualizations and other identified regions have been well documented in numerous ADHD reports. Therein, decreased volume of vermis [48] has been demonstrated to be a nonprogressive anatomical alteration in ADHD, which makes the cerebellar hemispheres a better target for clinical intervention. The paracentral region [49] has asymmetric alterations in the development of ADHD, which may be a prerequisite for neurodevelopmental disorders. Notably, Fig. 8(b) shows that the nodes of ADHD are clustered in the cerebellum and closely connected, which becomes an important sign for the algorithm to diagnose the disease. Indeed, it has been shown that some functional connections between the cerebellum and DMN regions are strongly correlated and influence the behavioral state of the brain [50].

In short, as an auxiliary diagnostic method, our proposed DA-DMHD method identifies subject-specific distinctive pathological regions, helping to simplify the diagnostic process for physicians. Thus, our proposed DA-DMHD method not only allows for accurate stratification of brain diseases but also identifies potential biomarkers associated with the disease, providing a new window for clinical diagnosis.

2) Limitations: In the current work, our proposed DA-DMHD method achieved good performance in diagnosing neurological disorders using single modality structural or functional brain networks. As a result, the DA-DMHD model does not consider fusing multimodal complementary information, which limits the model's performance in discovering correlations between structural and functional brain connections. Emerging evidence shows that functional-structural relationships exhibit a hierarchical structure in the brain. Stronger structural connections between ROIs imply higher functional interactions [51], [52]. In light of this, multimodal information was incorporated into our follow-up work. Another limitation of this work is that the optimization of the model needs to be divided into two stages. This indicates that our proposed approach is not a strictly end-to-end process, which may limit the model's ability to diagnose neurological disorders. In subsequent work, a manifold back-propagation optimization method that allows the DA-DMHD model to be optimized uniformly can be constructed.

V. Conclusion

In this paper, a dual-attention deep manifold harmonic discrimination (DA-DMHD) method for computer-aided degenerative disease diagnosis is presented. The primary distinctions between the work presented here and recent work on brain network diagnostic methods are the type of model input, the application space of the model, and the incorporation of discriminative attention. The utilization of harmonic waves (manifold-valued data) to encode the potential topology of the brain network and apply it to a deep Grassmann model with nonlinearity and discriminability is of interest. Theoretically, one could easily apply traditional deep learning methods to such data by ignoring the structure of the brain network and simply vectorizing it. However, such methods usually lead to inaccurate classification or high false positive rates due to the geometric structure of the brain network being ignored. Therefore, it is crucial to consider the geometry of the brain network data and implement the intrinsic operations allowed by the space in which the data are situated. Moreover, in contrast to existing deep Grassmann models that focus on improving only model separability, the heart of our approach is the development of a feature-driven discriminative attention block in deep Grassmann manifolds to achieve high robustness in degenerative disease diagnosis. The experimental results show that our method can not only identify discriminative brain network feature representations through harmonic waves but also has better diagnostic performance and robustness than other several state-of-the-art methods.

References

[1] S. Przedborski et al., "Series introduction: Neurodegeneration: What is it and where are we?," *J. Clin. Investigation*, vol. 111, no. 1, pp. 3–10, 2003.
[2] 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.
[3] C. J. Stam, "Modern network science of neurological disorders," *Nature Rev. Neurosci.*, vol. 15, no. 10, pp. 683–695, 2014.
[4] T. Xie and Y. He, "Mapping the Alzheimer's brain with connectomics," *Front. Psychiatry*, vol. 2, pp. 77–85, 2012.
[5] L. Wang et al., "Learning clique subgraphs in structural brain network classification with application to crystallized cognition," *NeuroImage*, vol. 225, pp. 1–16, 2021.
[6] M. A. Myszczynska et al., "Applications of machine learning to diagnosis and treatment of neurodegenerative diseases," *Nature Rev. Neurol.*, vol. 16, no. 8, pp. 440–456, 2020.
[7] I. Bilgen et al., "Machine learning methods for brain network classification: Application to autism diagnosis using cortical morphological networks," *J. Neurosci. Methods*, vol. 343, pp. 87–99, 2020.

[8] Y. Kong et al., "Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier," *Neurocomputing*, vol. 324, pp. 63–68, 2019.

[9] Y. Wang et al., "Dynamic graph CNN for learning on point clouds," *ACM Trans. Graph.*, vol. 38, no. 5, pp. 1–12, 2019.

[10] H. Chen et al., "A deep learning framework for identifying children with ADHD using an EEG-based brain network," *Neurocomputing*, vol. 356, pp. 83–96, 2019.

[11] A. Qiu et al., "Manifold learning on brain functional networks in aging," *Med. Image Anal.*, vol. 20, no. 1, pp. 52–60, 2015.

[12] L. Dodero et al., "Kernel-based analysis of functional brain connectivity on Grassmann manifold," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 604–611.

[13] 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.

[14] Y. Zhang, H. Zhang, L. Xiao, Y. Bai, V. D. Calhoun, and Y.-P. Wang, "Multi-modal imaging genetics data fusion via a hypergraph-based manifold regularization: Application to schizophrenia study," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2263–2272, Sep. 2022.

[15] Z. Huang et al., "Detecting brain state changes by geometric deep learning of functional dynamics on Riemannian manifold," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2021, pp. 543–552.

[16] Z. Huang et al., "Building deep networks on Grassmann manifolds," in *Proc. AAAI Conf. Artif. Intell.*, 2018, Art. no. 401.

[17] R. Wang, X.-J. Wu, and J. Kittler, "SymNet: A simple symmetric positive definite manifold deep learning method for image set classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 5, pp. 2208–2222, May 2022.

[18] S. Atasoy et al., "Human brain networks function in connectome-specific harmonic waves," *Nature Commun.*, vol. 7, no. 1, pp. 1–10, 2016.

[19] 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.

[20] S. Atasoy et al., "Connectome-harmonic decomposition of human brain activity reveals dynamical repertoire re-organization under LSD," *Sci. Rep.*, vol. 7, no. 1, pp. 1–18, 2017.

[21] A. Edelman et al., "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.

[22] Y. Chikuse, *Statistics on Special Manifolds*. Berlin, Germany: Springer, 2012.

[23] R. Subbarao and P. Meer, "Nonlinear mean shift for clustering over analytic manifolds," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 1168–1175.

[24] Z. Li, F. Nie, X. Chang, and Y. Yang, "Beyond trace ratio: Weighted harmonic mean of trace ratios for multiclass discriminant analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 10, pp. 2100–2110, Oct. 2017.

[25] C. Ionescu, O. Vantzos, and C. Sminchisescu, "Matrix backpropagation for deep networks with structured layers," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 2965–2973.

[26] C. Destrieux et al., "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature," *NeuroImage*, vol. 53, no. 1, pp. 1–15, 2010.

[27] D. Zhu et al., "Connectome-scale assessments of structural and functional connectivity in MCI," *Hum. Brain Mapping*, vol. 35, no. 7, pp. 2911–2923, 2014.

[28] L. Zhang et al., "Predicting brain structural network using functional connectivity," *Med. Image Anal.*, vol. 79–92, pp. 1–14, 2022.

[29] M. Jenkinson et al., "FSL," *NeuroImage*, vol. 62–74, no. 2, pp. 782–790, 2012.

[30] M. R. Brown et al., "ADHD-200 global competition: Diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements," *Front. Syst. Neurosci.*, vol. 6, pp. 69–81, 2012.

[31] R. Wang, X.-J. Wu, K.-X. Chen, and J. Kittler, "Multiple manifolds metric learning with application to image set classification," in *Proc. 24th Int. Conf. Pattern Recognit.*, 2018, pp. 627–632.

[32] R. Wang, X.-J. Wu, K.-X. Chen, and J. Kittler, "Multiple Riemannian manifold-valued descriptors based image set classification with multi-kernel metric learning," *IEEE Trans. Big Data*, vol. 8, no. 3, pp. 753–769, Jun. 2022.

[33] Y. Wen et al., "A discriminative feature learning approach for deep face recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 499–515.

[34] Y. Liang and G. Xu, "Multi-level functional connectivity fusion classification framework for brain disease diagnosis," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 6, pp. 2714–2725, Jun. 2022.

[35] L. Zhan et al., "Boosting brain connectome classification accuracy in Alzheimer's disease using higher-order singular value decomposition," *Front. Neurosci.*, vol. 9, pp. 257–269, 2015.

[36] N. Lubben et al., "The enigma and implications of brain hemispheric asymmetry in neurodegenerative diseases," *Brain Commun.*, vol. 3, no. 3, pp. 1–19, 2021.

[37] E. L. Platt, *Network Science With Python and NetworkX Quick Start Guide: Explore and Visualize Network Data Effectively*. Birmingham, U.K.: Packt Publishing Ltd, 2019.

[38] D. Dai et al., "Classification of ADHD children through multimodal magnetic resonance imaging," *Front. Syst. Neurosci.*, vol. 6, pp. 63–75, 2012.

[39] J. Ma et al., "Multimodality Alzheimer's disease analysis in deep Riemannian manifold," *Inf. Process. Manage.*, vol. 59, no. 4, pp. 1–8, 2022.

[40] M. Bozzali et al., "Damage to the cingulum contributes to Alzheimer's disease pathophysiology by deafferentation mechanism," *Hum. Brain Mapping*, vol. 33, no. 6, pp. 1295–1308, 2012.

[41] G. Karas et al., "Precuneus atrophy in early-onset Alzheimer's disease: A morphometric structural MRI study," *NeuroRadiology*, vol. 49, no. 12, pp. 967–976, 2007.

[42] O. Woolnough et al., "Category selectivity for face and scene recognition in human medial parietal cortex," *Curr. Biol.*, vol. 30, no. 14, pp. 2707–2715, 2020.

[43] M. D. Greicius et al., "Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 13, pp. 4637–4642, 2004.

[44] D. Vatansever et al., "Angular default mode network connectivity across working memory load," *Hum. Brain Mapping*, vol. 38, no. 1, pp. 41–52, 2017.

[45] N. Tzourio-Mazoyer et al., "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, 2002.

[46] K. Rubia et al., "Temporal lobe dysfunction in medication-naive boys with attention-deficit/hyperactivity disorder during attention allocation and its relation to response variability," *Biol. Psychiatry*, vol. 62, no. 9, pp. 999–1006, 2007.

[47] I. Ivanov et al., "Cerebellar morphology and the effects of stimulant medications in youths with attention deficit-hyperactivity disorder," *NeuroPsychoparmacology*, vol. 39, no. 3, pp. 718–726, 2014.

[48] S. Mackie et al., "Cerebellar development and clinical outcome in attention deficit hyperactivity disorder," *Amer. J. Psychiatry*, vol. 164, no. 4, pp. 647–655, 2007.

[49] S. Chen et al., "Asymmetry in cortical and subcortical structures of the brain in children and adolescents with attention-deficit/hyperactivity disorder," *Neuropsychiatric Dis. Treat.*, vol. 17, pp. 493–503, 2021.

[50] Y. Sun et al., "Differentiating boys with ADHD from those with typical development based on whole-brain functional connections using a machine learning approach," *Neuropsychiatric Dis. Treat.*, vol. 16, pp. 691–704, 2020.

[51] L. Xiao et al., "Multi-hypergraph learning-based brain functional connectivity analysis in fMRI data," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1746–1758, May 2020.

[52] W. Hu et al., "Interpretable multimodal fusion networks reveal mechanisms of brain cognition," *IEEE Trans. Med. Imag.*, vol. 40, no. 5, pp. 1474–1483, May 2021.