Information Fusion 118 (2025) 102971

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full length article

Explainable multi-frequency and multi-region fusion model for affective brain-computer interfaces

Tao Wang [a,c], Rui Mao [d,*], Shuang Liu [a,b,c,**], Erik Cambria [d], Dong Ming [a,b,c]

[a] Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, 300072, China
[b] State Key Laboratory of Advanced Medical Materials and Devices, Tianjin, 300072, China
[c] Haihe Laboratory of Brain-Computer Interaction and Human–Machine Integration, Tianjin, 300384, China
[d] College of Computing and Data Science, Nanyang Technological University, Singapore, 639798, Singapore

ARTICLE INFO

Keywords:
Explainable AI
Emotion recognition
Brain-computer interfaces
Functional connectivity
EEG

ABSTRACT

An affective brain-computer interface (aBCI) has demonstrated great potential in the field of emotion recognition. However, existing aBCI models encounter significant challenges in explainability and the effective fusion of multi-frequency and multi-region features, which greatly limits their practical applicability. To address these issues, this paper proposes an explainable multi-frequency and multi-region fusion network (MFMR-FN), consisting of multi-frequency encoding and multi-region decoding networks. Specifically, in the encoding network, we leverage spectral graph theory and symmetric positive definite (SPD) matrix learning to adaptively encode EEG data into functional connectivity (FC) matrices with neurobiological information. Furthermore, a multi-frequency fusion algorithm, based on Riemannian geometry is designed to guide the network in achieving cross-frequency feature fusion. In the decoding network, we introduce a multi-region selection mechanism and a multi-scale Riemannian network, fusing brain network features from the whole brain, hemispheres, and local regions for coarse-to-fine emotion decoding. We conducted extensive experiments on the emotion recognition dataset (SEED) and the depression detection dataset (MODMA). The results show that MFMR-FN outperforms existing methods in multiple metrics and provides explainable FC features that reveal brain network patterns under different emotions and abnormal connectivity in depression. The proposed MFMR-FN is expected to improve the practicality and reliability of aBCIs in real-world and clinical applications.

1. Introduction

An affective brain-computer interface (aBCI) is a novel interface technology for decoding human emotional states, and it represents a major breakthrough from traditional motor BCI to the next generation of cognitive BCI [1]. Compared to emotion recognition based on facial expressions [2], text [3], posture [4,5], and multi-perceptual modality-based approaches [6], aBCIs can directly decode cognitive activity from the brain, enabling more reliable in detecting subconscious emotional states and less susceptible to conscious control or masking of emotions [7]. Furthermore, aBCIs not only recognize emotional states, but also regulate emotions through closed-loop systems, thus providing a new way for the diagnosis and treatment of mental disorders [8]. Therefore, the development of intelligent, efficient, and low-cost aBCIs has important scientific significance and application prospects. In aBCI research, electroencephalogram (EEG) signals have been widely used due to their high temporal resolution and non-invasiveness [9].

Traditional EEG features are typically extracted manually and focus on the time, frequency, and time-frequency domains [10–12]. Functional Connectivity (FC) is also an important feature, which represents the interactions and coordinated activities between different brain regions [13–15]. By extracting FC features, researchers can capture the information exchange between brain regions under different emotional states, making it widely used in aBCI tasks [16,17]. However, since manually crafted features highly depend on the researchers' prior knowledge, they often fail to capture features with strong discrimination and generalization. In recent years, convolutional neural networks (CNNs) [18] and long short-term memory networks (LSTMs) [19] have been widely applied to aBCIs and achieving significant progress. Furthermore, deep learning models based on FC, such as DGCNN [20], DeepfMRI [21], BrainGNN [22], and PPT-CNN [23], have also shown excellent emotion recognition performance. Compared to hand-crafted features, neural networks adaptively extract features from raw EEG data, enhancing recognition accuracy through end-to-end learning.

* Corresponding author.
** Corresponding author at: Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, 300072, China.
E-mail addresses: taowang2021@tju.edu.cn (T. Wang), rui.mao@ntu.edu.sg (R. Mao), shuangliu@tju.edu.cn (S. Liu), cambria@ntu.edu.sg (E. Cambria), richard-ming@tju.edu.cn (D. Ming).

https://doi.org/10.1016/j.inffus.2025.102971
Received 1 November 2024; Received in revised form 23 December 2024; Accepted 17 January 2025
Available online 24 January 2025

Although deep learning has made significant progress in the field of aBCIs, there are still several challenges that need to be addressed. (1) Deep learning models are often seen as "black box", with the extracted features lacking explainability [24]. In practical applications, the explainability of the model directly impacts its user trust [25], especially in scenarios such as aBCIs that require clear explanations of causality and decision-making. (2) Existing aBCI research has found that EEG signals from different frequency bands exhibit significant differences in emotional representations. For instance, low-frequency signals are primarily associated with relaxed and low-arousal emotions [26,27], while high-frequency signals are more active in high-arousal emotions [28,29]. However, current deep learning models often fail to effectively fuse cross-frequency features, which limits the recognition performance of aBCI-based models. (3) Existing research shows that asymmetric brain activity between the hemispheres and local brain network patterns are related to cognitive processes in emotion [30,31]. Nevertheless, most current research focuses on capturing coarse-grained features at the whole-brain scale, neglecting the deeper exploration of fine-grained features in hemispheric and local brain regions. Therefore, effective fusion of coarse-to-fine EEG features from different brain regions remains a critical challenge.

These limitations significantly hinder the generalizability and practicality of deep learning models for aBCI applications. To address the aforementioned issues, we propose an explainable multi-frequency and multi-region fusion network (MFMR-FN), consisting of a multi-frequency encoding network and a multi-region decoding network. It aims to improve the recognition accuracy of an aBCI model while also enhancing its explainability. Specifically, to address Issues (1) and (2), we constructed a multi-frequency encoding network. The encoding network consists of two modules, namely the functional connectivity-graph convolutional module (FG-GCN) and the bilinear and rectification matrix mapping modules (BiRe). These modules encode complex EEG signals into FC matrices, which reflect neural activity between different brain regions. Furthermore, by introducing a multi-frequency fusion algorithm based on Riemannian geometry, the encoding network is guided to extract cross-frequency fused FC. The output of this encoding network is the FC matrix with neurobiological information, enabling researchers to perform multi-dimensional brain network analysis. By analyzing these FC features, we can gain deep insights into the decision-making mechanisms of the encoding network. To tackle Issue (3), we propose a multi-region decoding network that integrates a multi-region selection mechanism with a multi-scale Riemannian network. This approach enables the network to capture both coarse-grained FC features at the whole-brain level and fine-grained features within hemispheric and localized brain regions. By fusing brain network patterns from coarse to fine, this method effectively enhances the decoding performance of aBCIs.

We conducted extensive experiments on two publicly available datasets, including SEED [32] for emotion recognition and MODMA [33] for depression recognition, to validate the generalization and explainability of the proposed MFMR-FN in various aBCI tasks. The experimental results demonstrate that our MFMR-FN outperforms existing methods across multiple metrics, achieving the best recognition performance. Furthermore, the model's output features demonstrate informative explainability, offering insights into brain network patterns associated with various emotional states and identifying abnormal connectivity in patients with depression. It significantly enhances the practical applicability of deep learning models in the field of aBCIs.

The contribution of this paper can be summarized as follows:

- The multi-frequency encoding network is proposed based on spectral graph theory and SPD matrix learning, which adaptively encodes EEG data into functional connections with neurobiological information. Meanwhile, a multi-frequency fusion algorithm based on Riemannian geometry is designed to guide the network in achieving cross-frequency feature fusion.
- The constructed decoding network captures coarse-grained features at the whole-brain scale as well as fine-grained features, such as hemispheric asymmetry and local brain network patterns, significantly improving the decoding performance of the aBCI model.
- The proposed network not only excels in performing the aBCI recognition task but also generates explainable brain functional connectivity matrices. This capability enables researchers to conduct comprehensive brain network analyses, gaining deeper insights besides analyzing emotional states.

The rest of this paper is organized as follows. Section 2 provides a brief review of related work. Section 3 describes the proposed MFMR-FN model in detail. The experimental setups and results are presented in Sections 4 and 5, respectively. Finally, conclusions and future work are discussed in Section 6.

2. Related work

With the rapid advancement of deep learning technology, particularly breakthroughs in the design and optimization of neural network architectures, deep learning models have shown great potential in aBCIs [34]. A core challenge in aBCI processing is extracting emotion-related features from complex EEG signals. Unlike traditional hand-crafted feature extraction methods, deep learning models can adaptively learn high-level features from raw EEG data, significantly improving the accuracy and robustness of emotion recognition [35]. Therefore, more researchers studied aBCIs via deep learning. For example, Gao et al. [36] proposed a GCN-based aBCI model, termed CU-GCN, which utilizes GCN to adaptively compute the connection probabilities between different EEG nodes and generate adaptive adjacency matrices, further improving the ability to capture spatial and temporal features in EEG signals. Tang et al. [37] proposed SiameseSPD-MR, which extracts spatial features from EEG signals through GCN, effectively capturing the connectivity information between brain regions. Gao et al. [38] introduced a framework, termed EEG-GCN, designed for single-view and multi-view aBCI tasks. This method employs GCN combined with spatio-temporal attention mechanisms to dynamically capture the temporal and spatial information in EEG signals. In this work, we improve the GCN to encode complex EEG signals into FC matrices that reflect neural activities between different brain regions, thereby constructing an explainable encoding network for aBCIs.

In aBCI research, EEG data analysis often focuses on exploring interactions between brain regions through FC matrices, which can be represented as the Symmetric Positive Definite (SPD) matrices [39]. These are defined within a Riemannian manifold, where geodesic distances serve as a key metric [40]. Deep learning methods often overlook the SPD geometric structure of FC matrices, resulting in the loss of critical brain network information [41]. Unlike traditional neural networks, the Riemannian network [42] operates on the Riemannian manifold where SPD matrices reside and achieve great success in FC matrix processing [43]. For instance, Zhang et al. [44] proposed a spatio-temporal joint optimization network based on Riemannian networks to capture the spatial relationships between different EEG brain regions. By applying bilinear matrix mapping, their model generates more discriminative SPD matrices, while preserving the geometric structure of the EEG covariance matrices.

Wu et al. [45] captured interactions between different brain regions through FC matrices and used Riemannian networks to extract high-level SPD representations embedded in FC matrices, further improving the accuracy and stability of emotion recognition. Kim et al. [46] proposed a DRBN model based on Riemannian networks, which learns covariance matrices of EEG signals on the Riemannian manifold for emotion recognition. This method extracts more discriminative features by learning the Riemannian barycenter of each class, improving the classification performance while preserving the geometric properties of SPD matrices.

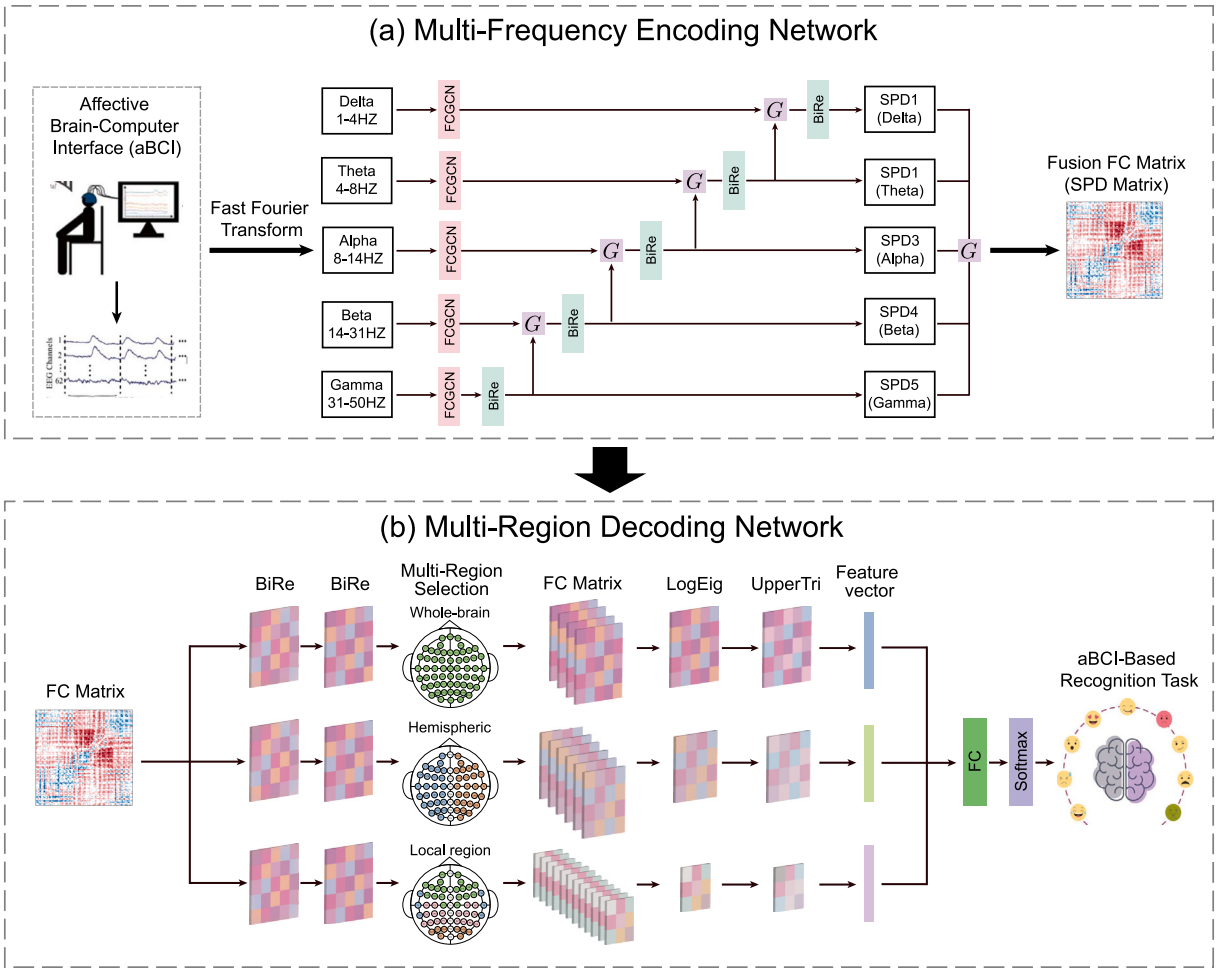

Fig. 1. The overall framework of the proposed MFMR-FN. (a) Multi-Frequency Encoding Network: EEG data is first decomposed into five frequency bands (Delta, Theta, Alpha, Beta, Gamma). Then the EEG data is encoded into FC matrices using FC-GCN and BiRe modules, and combined with the fusion algorithm G to extract cross-frequency FC features. (b) Multi-Region Decoding Network: The FC matrix extracted from the encoding network is fed into a three-branch Riemannian network. Each branch comprising the BiRe, LogEig, and UpperTri layers is combined with the multi-region selection mechanism to extract FC features at whole-brain, hemispheric, and local regions. Finally, these features are fused through a fully connected layer and fed into the softmax layer for final classification.

In this paper, we modify the Riemannian network's structure to capture high-level FC features during encoding, while preserving the explainability of the FC matrix. We also construct a multi-scale Riemannian network in the decoding network to enhance the decoding ability of the aBCI by fusing coarse-to-fine features across different brain regions.

3. Methodology

3.1. Method overview

The overview of the proposed MFMR-FN is shown in Fig. 1. The framework has two main components: the multi-frequency encoding network (Section 3.2) and the multi-region decoding network (Section 3.3). As shown in Fig. 1(a), in the multi-frequency encoding network, the EEG data is filtered into five frequency bands using fast fourier transform. The EEG signals in each frequency band are encoded through the FC-GCN (Section 3.2.1) and BiRe module (Section 3.2.2) to capture shallow and deep functional connectivity. We also propose a multi-frequency FC fusion algorithm (Section 3.2.3) for cross-frequency information fusion (As shown by G in Fig. 1(a)). This algorithm propagates fixed-band FC from each branch to adjacent branches for information fusion, followed by multi-band fusion encoding at the end of the network. The output of the encoding network is the FC matrix, which not only allows explainability analysis of brain networks, but also serves as input for the multi-region decoding network for classification tasks. As shown in Fig. 1(b), in the Multi-Region Decoding Network, we introduce a multi-region selection mechanism (Section 3.3.1) to extract features across different scales of brain regions. Based on this mechanism, we then construct a multi-scale Riemannian network (Section 3.3.2) for feature extraction and fusion across whole-brain, hemispheric, and local regions. This network captures coarse-to-fine FC patterns, improving the performance of the aBCI in various classification tasks.

3.2. Multi-frequency encoding network

In this section, we construct a multi-frequency encoding network, which is shown in Fig. 1(a). The encoding network extracts brain FC matrices from EEG data based on spectral graph theory and SPD matrix learning, while using cross-frequency fusion to further enhance the discriminability of aBCI emotion features. First, we construct an FG-GCN module to transform complex EEG signals into FC matrices with neurobiological information.

Then, we feed the FC matrices into the BiRe module for SPD matrix learning in the Riemannian manifold. This process extracts discriminable features while preserving the neural topological structures, ensuring the explainability of the extracted features. Moreover, given that the FC matrices output by each branch reside in Riemannian manifold space, traditional fusion strategies in Euclidean space are

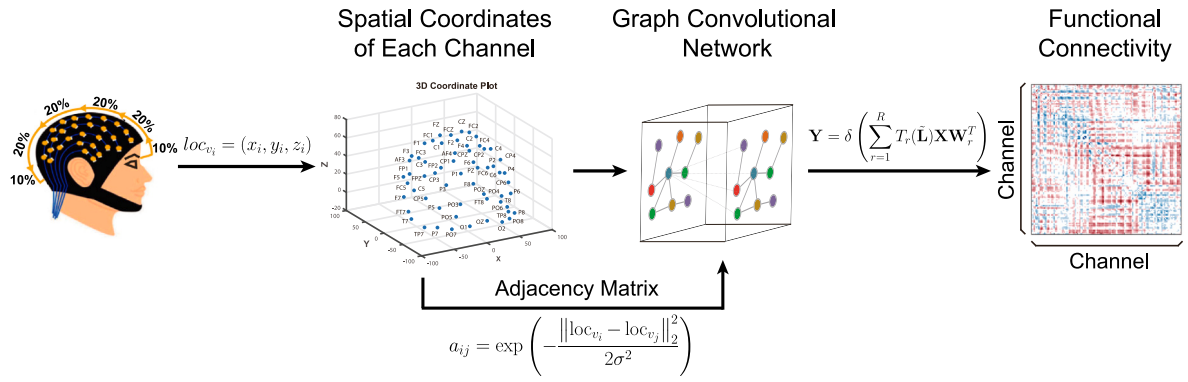

Fig. 2. The architecture of the proposed FC-GCN. EEG signals are processed by constructing an adjacency matrix based on the 3D scalp model using a Gaussian kernel. The GCN leverages the node feature matrix and adjacency matrix to learn spatial relationships within brain networks, subsequently generating the FC matrix.

not applicable. Therefore, a multi-frequency fusion algorithm based on Riemannian geometry is designed to guide the network in achieving cross-frequency feature fusion. The following sections provide a detailed explanation of the modules and the multi-frequency fusion algorithm.

3.2.1. Functional Connectivity-Graph Convolutional module (FC-GCN)

The Graph Convolutional Networks (GCNs), utilizing their unique spectral graph theory [47], can effectively learn spatial relationships within brain networks and have been widely applied in the analysis of brain FC [48]. Therefore, we construct the FC-GCN based on GCNs, which adaptively extracts FC patterns from EEG data. Meanwhile, to facilitate the subsequent operation, we arrange the EEG channels in the order of $[C_{left}, C_{right}]$, where C_{left} represents the channels located in the left hemisphere and C_{right} represents those in the right hemisphere. Within each hemisphere, the channels are organized by brain region: frontal, parietal, occipital, and temporal.

The input to GCNs includes a node feature matrix X and an adjacency matrix A. Inspired by the work of [37], we construct the GCN input by considering the anatomical structure of the 3D scalp model. Specifically, each EEG channel is treated as a node, and the adjacency matrix is computed using the Gaussian kernel function [49] based on the spatial distances between electrodes. The overall architecture of the FC-GCN module is shown in Fig. 2.

We define the node feature matrix as $X \in \mathbb{R}^{n\times c}$, where n is the number of EEG electrodes, and c is the number of features for each channel. The adjacency matrix $A \in \mathbb{R}^{n\times n}$ represents the connectivity between nodes, where the element a_{ij} indicates the connection weight between nodes i and j. The a_{ij} is defined as follows:

$$a_{ij} = \exp\left(-\frac{\left\|\mathrm{loc}_{v_i} - \mathrm{loc}_{v_j}\right\|_2^2}{2\sigma^2}\right), \tag{1}$$

where loc_{v_i} and loc_{v_j} represent the positions of nodes v_i and v_j in 3D scalp model, respectively. The σ denotes the width of the Gaussian kernel, and the diagonal elements of A are set to 1.

After constructing the graph, we utilize the GCN to learn high-level feature representations of the nodes. The operation is as follows:

$$Y = \delta\left(\sum_{r=1}^{R} T_r(\tilde{L})XW_r^T\right), \tag{2}$$

where δ represents the Rectified Linear Unit. The Chebyshev polynomial $T_r(\tilde{L})$ is used to approximate the rth order polynomial of the normalized Laplacian $\tilde{L}$. X is the input feature matrix, and W_r is the trainable weight matrix for rth order spectral filtering. We used GCN to obtain a high-level feature representation of the nodes. Subsequently, we compute FC from these nodes using Gaussian kernel functions. Specifically, for the obtained node $Y = [y_1, y_2, \ldots, y_n]^T \in \mathbb{R}^{n\times c}$, the FC matrix $F = [f_{ij}]_{n\times n}$ is computed as follows:

$$f_{ij} = \exp\left(-\gamma^2\|\mathbf{y}_i - \mathbf{y}_j\|_2^2\right), \tag{3}$$

where γ is a trainable scale parameter used to adaptively adjust the FC matrix F, and the F is an SPD matrix as indicated in Theorem 1. The detailed proof of Theorem 1 is given in Appendix A.

Theorem 1. *Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T \in \mathbb{R}^{n\times c}$ represent a set of data points, and the mapping $f : (X \times X) \to \mathbb{R}^{n\times n}$ is defined by: $f_{ij} = \exp\left(-\gamma^2\|\mathbf{x}_i - \mathbf{x}_j\|_2^2\right)$ with $1 \le i, j \le n$. Then $\mathbf{F} = [f_{ij}]_{n\times n}$ is a SPD matrix.*

3.2.2. Bilinear and Rectification matrix mapping module (BiRe)

The FC matrix F extracted by the FC-GCN module belongs to the set of SPD matrices, which form a connected Riemannian manifold S_{++}^n [50]. Therefore, to extract more discriminative FC representations, we construct the BiRe module to perform SPD matrix learning for the FC extracted by the FG-GCN module. Furthermore, the BiRe preserves the neural topological structures of the FC through matrix mapping, ensuring the explainability of the extracted FC matrices.

The BiRe module is composed of a bilinear mapping (BiMap) layer and an eigenvalue rectification (ReEig) layer [42], as shown in Fig. 3. The BiMap layer transforms the input SPD matrix into a more discriminative matrix using the bilinear mapping transformation matrix. During this process, the input matrix does not need to be vectorized, thus, the functional connectivity information between different brain regions contained in the SPD matrix structure is preserved. The ReEig layer follows BiMap layer to ensure the mapped matrix remains on the Riemannian manifold and to introduce non-linearity for mapping process. The function of ReEig layer is similar to the ReLU layer in classical convolutional neural networks.

Let $\boldsymbol{F}_{n-1} \in S_{++}^n$ be the input FC matrix to the nth BiRe layer, the BiRe layer is defined as follows:

$$\boldsymbol{F}_{b,n} = f_b\left(\boldsymbol{F}_{n-1}, \boldsymbol{W}_n\right) = \boldsymbol{W}_n\boldsymbol{F}_{n-1}\boldsymbol{W}_n^{\mathrm{T}} \tag{4}$$

$$\boldsymbol{F}_{r,n} = f_r\left(\boldsymbol{F}_{b,n}, \varepsilon\right) = \boldsymbol{U}_n\ \mathrm{Max}\left(\varepsilon\boldsymbol{I}, \Lambda_n\right)\boldsymbol{U}_n^{\mathrm{T}} \tag{5}$$

$$\mathrm{Max}\left(\varepsilon\boldsymbol{I}, \Lambda_n\right) = \boldsymbol{E}(i,i) = \begin{cases} \Lambda(i,i), & \text{if } \Lambda(i,i) > \varepsilon \\ \varepsilon, & \text{if } \Lambda(i,i) \le \varepsilon \end{cases} \tag{6}$$

$\boldsymbol{F}_{b,n}$ is the output of BiMap layer and $\boldsymbol{F}_{r,n}$ is the output of ReEig layer. $\boldsymbol{W}_n$ is the bilinear mapping transformation matrix, which is commonly required to be of full row-rank. $\boldsymbol{U}_n$ and Λ_n denote the eigenvectors and eigenvalues of input matrix $\boldsymbol{F}_{b,n}$, i.e., $\boldsymbol{F}_{b,n} = \boldsymbol{U}_n\Lambda_n\boldsymbol{U}_n^T$. $\boldsymbol{I}$ is the identity matrix. $\varepsilon > 0$ is a preset rectification threshold, which is used to replace null or small eigenvalues in Λ_n to obtain a new matrix $\boldsymbol{E}(i,i)$.

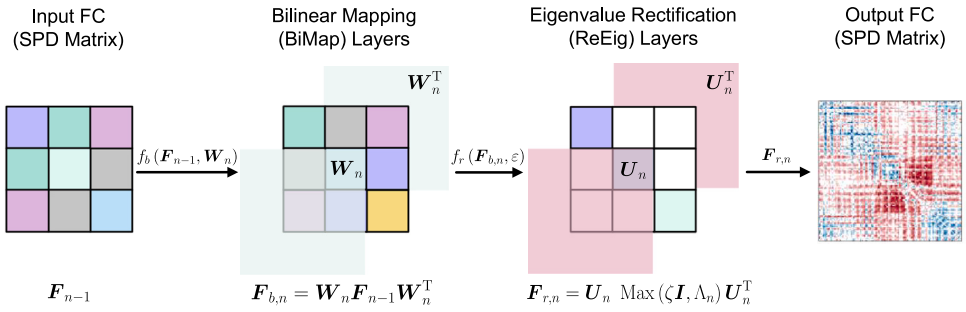

Fig. 3. The architecture of the BiRe module. $\boldsymbol{F}_{n-1}$ represents the input FC, which is an SPD matrix. The BiRe module consists of ReEig and BiMap layers. The ReEig layer replaces small eigenvalues with the preset threshold ε, ensuring the positive properties of mapped matrices. The BiMap layer transforms the input SPD matrix into a more discriminative matrix using the transformation matrix $\boldsymbol{W}$. Finally, the output matrix remains an SPD matrix with the same dimensions.

3.2.3. Multi-frequency FC fusion algorithm

In the forward propagation of different branches, inspired by the work of [51,52], we design a multi-frequency FC fusion algorithm. This algorithm propagates the SPD output from fixed-band branch to adjacent branches for information fusion, and finally achieves cross-frequency FC fusion at the end of the network. The above fusion process is defined as follows.

For the SPD matrices from multiple branches $\{F_i\}_{i=1}^N \in \mathcal{S}_{++}^n$, the fusion operation is defined as:

$$F_i^* = \begin{cases} G\left(F_i\right) & i = 1 \\ G\left(F_i + F_{i-1}^*\right) & i > 1 \end{cases} \tag{7}$$

F_i^* represents the output of the ith branch. $G(\cdot)$ is the fusion operation, which is equivalent to computing the central position of the SPD matrix set $\{F_i\}_{i=1}^N \in \mathcal{S}_{++}^n$. Thus, we need to find a point $\mathfrak{G}$ in the Riemannian manifold $\mathcal{S}_{++}^n$ that minimizes the weighted sum of squared Riemannian distances to all points in the set $\{F_i\}$. We refer to $\mathfrak{G}$ as the weighted Riemannian centroid, which is defined as:

$$\begin{aligned} \mathfrak{G} &= \mathrm{Rie}_{\boldsymbol{w}}\left(\left\{F_i\right\}_{i=1}^N\right) \\ &= \underset{\boldsymbol{F}_\mu \in \mathcal{S}_{++}^n}{\operatorname{argmin}} \sum_{i=1}^N w_i \delta_{\mathfrak{G}}^2\left(\boldsymbol{F}_i, \boldsymbol{F}_\mu\right); \text{ s.t. } w_i \geq 0, \sum_{i=1}^N w_i = 1, \end{aligned} \tag{8}$$

where $\boldsymbol{F}_\mu$ represents a point in the Riemannian manifold $\mathcal{S}_{++}^n$. The set of weights $\{w_i\}_{i=1}^N$ is used to compute the Riemannian centroid and satisfies convexity constraints, meaning all w are non-negative and sum to 1. $\delta_{\mathfrak{G}}^2\left(\boldsymbol{F}_i, \boldsymbol{F}_\mu\right)$ denotes the squared Riemannian distance from point $\boldsymbol{F}_i$ to point $\boldsymbol{F}_\mu$. The Riemannian distance between two points is defined as follows:

$$\delta_{\mathfrak{G}}\left(\boldsymbol{F}_i, \boldsymbol{F}_\mu\right) = \frac{1}{2}\left\|\log\left(\boldsymbol{F}_i^{-\frac{1}{2}} \boldsymbol{F}_u \boldsymbol{F}_i^{-\frac{1}{2}}\right)\right\|_F, \tag{9}$$

where $\|\cdot\|_F$ represents the Frobenius norm, and log denotes the matrix logarithm operation applied to the SPD matrix.

For the set of SPD matrices $\{\boldsymbol{F}_i\}_{i=1}^N$, when $N = 2$, the closed-form solution exists for Eq. (8), which corresponds to the geodesic distance between the points F_1 and F_2:

$$\begin{aligned} \mathfrak{G} &= \mathrm{Bar}_{(w,1-w)}\left(F_1, F_2\right) \\ &= F_2^{\frac{1}{2}}\left(F_2^{-\frac{1}{2}} F_1 F_2^{-\frac{1}{2}}\right)^w F_2^{\frac{1}{2}}; \text{ s.t. } w \in [0,1] \end{aligned} \tag{10}$$

We set $w = 0.5$. Therefore, for the fusion of two FC matrices, their Riemannian barycenter $\mathfrak{G}$ can be directly obtained using Eq. (10).

However, when $N > 2$ (in this paper, $N = 5$), the closed-form optimal solution for Eq. (8) has not yet been theoretically resolved. Therefore, we designed the multi-frequency FC fusion algorithm to approximate the Riemannian barycenter $\mathfrak{G}$. Specifically, in each iteration, we first select an initial point $\mathfrak{G}$ in the Riemannian manifold $\mathcal{S}_{++}^n$ as the starting estimate of the Riemannian barycenter. Then, using the logarithm operation $\mathrm{Log}_{\mathfrak{G}}(\cdot)$, we map all points $\left\{F_i\right\}_{i=1}^5$ (the five FC matrices to be fused from the 5 branches) to the tangent space $\mathcal{T}\mathfrak{G}$. These mapped results are denoted as $\{\boldsymbol{H}_i\}_{i=1}^5$. In the tangent space, we compute the arithmetic mean $C_{\mathcal{T}_{\mathfrak{G}}}$ of all the mapped points. Finally, we apply the exponential operation $\mathrm{Exp}_{\mathfrak{G}}(\cdot)$ to map the obtained arithmetic mean back to the manifold space $\mathcal{S}_{++}^n$, generating a new barycenter estimate $\mathfrak{G}^*$. The $\mathfrak{G}^*$ serves as the base point for the next iteration, and the process is repeated until convergence to a stable barycenter. This procedure is shown in Fig. 4, where the five SPD matrices represent the brain FC in five frequency bands.

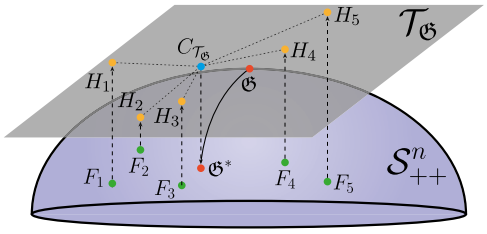

Fig. 4. The illustration of the multi-frequency FC fusion process (when $N > 2$).

We present the above multi-frequency FC fusion algorithm in Algorithm 1. For the square root $(\cdot)^{\frac{1}{2}}$ and inverse square root $(\cdot)^{-\frac{1}{2}}$ operations for SPD matrices, we define their backpropagation in the Riemannian manifold $\mathcal{S}_{++}^n$ using the Daleckĭi-Kreĭn formula [53]. Considering the nonlinear function $f(F)$ (such as $F^{\frac{1}{2}}$ or $F^{-\frac{1}{2}}$), the gradient of the lth layer $\frac{\partial \mathcal{L}^{(l)}}{\partial F}$ is given by:

$$\frac{\partial \mathcal{L}^{(l)}}{\partial C} = U\left(L \odot \left(U^T \frac{\partial \mathcal{L}^{(l+1)}}{\partial f(F)} U\right)\right) U^T, \tag{11}$$

where U is the eigenvector matrix of the input SPD matrix F. $\mathcal{L}$ denotes the loss function. $\odot$ represents the Hadamard product. L is the Loewner matrix, which is defined as:

$$L_{ij} = \begin{cases} \frac{f(\sigma_i) - f(\sigma_j)}{\sigma_i - \sigma_j} & \text{if } \sigma_i \neq \sigma_j \\ f'(\sigma_i) & \text{if } \sigma_i = \sigma_j \end{cases} \tag{12}$$

σ_i represents the eigenvalues of F. $f(\sigma_i)$ corresponds to the function applied to the eigenvalues. By using L and U, the gradient can be backpropagated to the manifold space $\mathcal{S}_{++}^n$, ensuring accurate gradient updates.

3.3. Multi-region decoding network

Previous studies have shown that asymmetric interhemispheric activity and local brain network patterns are linked to emotional cognition and mental disorders [31,54]. Therefore, for the construction of the decoding network, it is important to capture not only coarse-grained FC features at the whole-brain scale, but also to extract fine-grained features from the hemispheres and local brain regions.

Algorithm 1 Multi-branch Riemannian network fusion mechanism

Input: The set of SPD matrices from different network branches $\{\boldsymbol{F}_i\}_{i=1}^{N}$, s.t. $N \geq 2$

Output: The fused SPD matrix $\mathfrak{G}$

initialization: $\mathfrak{G} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{F}_i$, criterion $\Omega = 10^8$, tolerance $\varepsilon = 10^{-8}$, iteration $t = 1$, maxiter $t_{\max} = 50$
if $N = 2$ **then**
  # Compute the geodesic distance of two matrices
  $\mathfrak{G} = \boldsymbol{F}_2^{\frac{1}{2}}\left(\boldsymbol{F}_2^{-\frac{1}{2}}\boldsymbol{F}_1\boldsymbol{F}_2^{-\frac{1}{2}}\right)^{\frac{1}{2}}\boldsymbol{F}_2^{\frac{1}{2}}$
end if
# Iterative method for more than two matrices
if $N > 2$ **then**
  while $\Omega > \varepsilon$ and $t < t_{\max}$ **do**
    # Map SPD matrices to tangent space
    $H_i = \mathrm{Log}_{\mathfrak{G}}(\mathbf{F}_i) = \mathfrak{G}^{\frac{1}{2}}\log\left(\mathfrak{G}^{-\frac{1}{2}}\mathbf{F}_i\mathfrak{G}^{-\frac{1}{2}}\right)\mathfrak{G}^{\frac{1}{2}}$
    # Calculate arithmetic mean in tangent space
    $C_{\mathcal{T}_{\mathfrak{G}}} = \frac{1}{N}\sum_{i=1}^{N} H_i$
    # Map mean back to Riemannian manifold
    $\mathfrak{G}^* = \mathrm{Exp}_{\mathfrak{G}}\left(C_{\mathcal{T}_{\mathfrak{G}}}\right) = \mathfrak{G}^{\frac{1}{2}}\exp\left(\mathfrak{G}^{-\frac{1}{2}}C_{\mathcal{T}_F}\mathfrak{G}^{-\frac{1}{2}}\right)\mathfrak{G}^{\frac{1}{2}}$
    # Check convergence using distance
    $\Omega = \delta_{\mathfrak{G}}(\mathfrak{G}, \mathfrak{G}^*) = \frac{1}{2}\left\|\log\left(\mathfrak{G}^{-\frac{1}{2}}\mathfrak{G}^*\mathfrak{G}^{-\frac{1}{2}}\right)\right\|_F$
    # Update barycenter estimate
    $\mathfrak{G} = \mathfrak{G}^*$
    $t = t + 1$
  end while
end if

For this purpose, we proposed a multi-region decoding network, which is shown in Fig. 1(b). The network has three branches that extract multi-scale brain network patterns from the whole-brain, hemispheric, and local region scales. The following sections detail the multi-region selection mechanism and the multi-scale Riemannian network employed in the decoding network.

3.3.1. Multi-region selection mechanism

To extract brain network patterns at three different scales, we have already reorganized the input EEG channels in Section 3.2.1. Therefore, the channel order in the input FC should be $[C_{left}, C_{right}]$, where C_{left} refers to the left hemisphere channels and C_{right} to the right. Within each hemisphere, channels are arranged by region: frontal, parietal, occipital, and temporal. This arrangement allows for better multi-region selection.

For the whole-brain scale, we use the full-channel FC without applying any selection mechanism to the input FC. For the hemispheric scale, we aim for the network to focus on extracting asymmetric interhemispheric activity, which is relevant to the cognitive processes in emotion. Therefore, we need to extract the input FC matrix by splitting it into left and right hemispheres. However, directly splitting the matrix may result in non-SPD matrices, which would destroy the neurobiological information encoded in the SPD matrix and also cannot be processed in the Riemannian network. To resolve this, we propose a multi-region selection mechanism. Fig. 5 illustrates the implementation of the multi-region selection mechanism in hemispheric scale. In Fig. 5(a), we define the FC matrix output from the BiRe module as $F \in \mathbb{R}^{B\times N\times N}$. B is the number of samples in each mini-batch, and N is the number of EEG channels. In probability theory, for a given random matrix F of size $N \times N$, its distribution can be described by the covariance matrix C of size $N^2 \times N^2$.

The covariance matrix C is defined as follows:

$$C = \mathrm{Var}(\mathrm{vec}(F)) = E\left[(\mathrm{vec}(F) - \mu)(\mathrm{vec}(F) - \mu)^T\right], \tag{13}$$

where $\mathrm{vec}(F)$ represents the vectorization of matrix F, μ is the mean of $\mathrm{vec}(F)$, and E denotes the expectation operator. The vectorization $\mathrm{vec}(\cdot)$ converts an $N \times N$ matrix into an $N^2 \times 1$ vector. This process is illustrated in Fig. 5(b)–(c).

Furthermore, for the FC matrix F in Fig. 5(a), we apply a sliding window of size $W \times W$ with stride W to select W_n regions of interest. W_n is defined as:

$$W_n = \left(\frac{N - W}{W} + 1\right)^2 \tag{14}$$

In the hemispheric-scale selection mechanism, W is set to $N/2$, which is half the number of EEG channels. We use four colors to represent the selected $W_n = 4$ regions of interest: intra-left hemisphere, left-to-right hemisphere, right-to-left hemisphere, and intra-right hemisphere. Each $W \times W$ region in Fig. 5(a) corresponds to a $W^2 \times W^2$ principal submatrices in Fig. 5(c). According to Theorem 2, the selected principal submatrices are still SPD matrices. Therefore, we concatenate these submatrices to form W_n SPD matrices in Fig. 5(d), which captures hemispheric brain network activity and is used in our multi-scale Riemannian network. The detailed proof of Theorem 2 is given in Appendix B.

For the local region scale, we further extract FC features within and between local regions using the proposed multi-region selection mechanism. Specifically, in the local scale, the sliding window parameter W is set to match the number of channels in the smallest brain region. For example, for 64-channel EEG data, W is defined as the number of channels in the left temporal lobe. This ensures the sliding window size is consistent with the spatial resolution of the selected brain region, thus capturing local brain network patterns more accurately.

Theorem 2. *For a covariance matrix $C \in \mathbb{R}^{n\times n}$, which is a SPD matrix, its principal submatrices $(C)_i \in \mathbb{R}^{w\times w}$ are also SPD matrices and form a connected Riemannian manifold $\mathcal{S}_{++}^{w}$.*

3.3.2. Multi-scale Riemannian network

Inspired by the work of [55], we constructed the multi-scale Riemannian network to learn multi-scale brain network patterns from obtained FC matrices, enabling more precise aBCI recognition tasks. The framework of the proposed multi-scale Riemannian network is illustrated in Fig. 1(b). The network is composed of three branches, each designed to capture brain network patterns at different levels: whole-brain, hemispheric, and local regions. In each branch, two BiRe modules (described in Section 3.2.2) are first used to extract low-dimensional FC features. Then, using the multi-region selection mechanism (detailed in Section 3.3.1), we extract FC features at three different scales. Subsequently, the LogEig and UpperTri layers are applied for high-dimensional feature extraction. Finally, the feature vectors from all branches are fused through a fully connected layer and fed to the softmax layer for final classification.

In particular, the LogEig and UpperTri layers are defined as follows. The LogEig layer imparts a Lie group structure to elements on Riemannian manifolds, allowing the matrix to be simplified into a flat space where traditional Euclidean computations can be applied [56]. The LogEig layer is defined as follows:

$$\mathbf{F}_{l,n} = f_l\left(\mathbf{F}_{n-1}\right) = \mathbf{Q}_{n-1}\log\left(\mathbf{A}_{n-1}\right)\mathbf{Q}_{n-1}^{\mathrm{T}}, \tag{15}$$

where $\mathbf{Q}_{n-1}$ and $\mathbf{A}_{n-1}$ denote the eigenvectors and eigenvalues of the input matrix respectively, and $\log(\cdot)$ is the matrix logarithm operation. The UpperTri layer extracts the upper triangular part of the input FC matrix (excluding the diagonal) to reduce computational complexity, then flattens these elements into a vector. The LogEig layer is defined as $\mathbf{V} = \mathrm{vec}(\mathbf{T} \odot \mathbf{F})$. Where, $\mathbf{V}$ represents the input matrix, and $\mathbf{V}$ is the output vector. $\mathbf{T}$ is the upper triangular mask matrix, where $T_{ij} = 1$

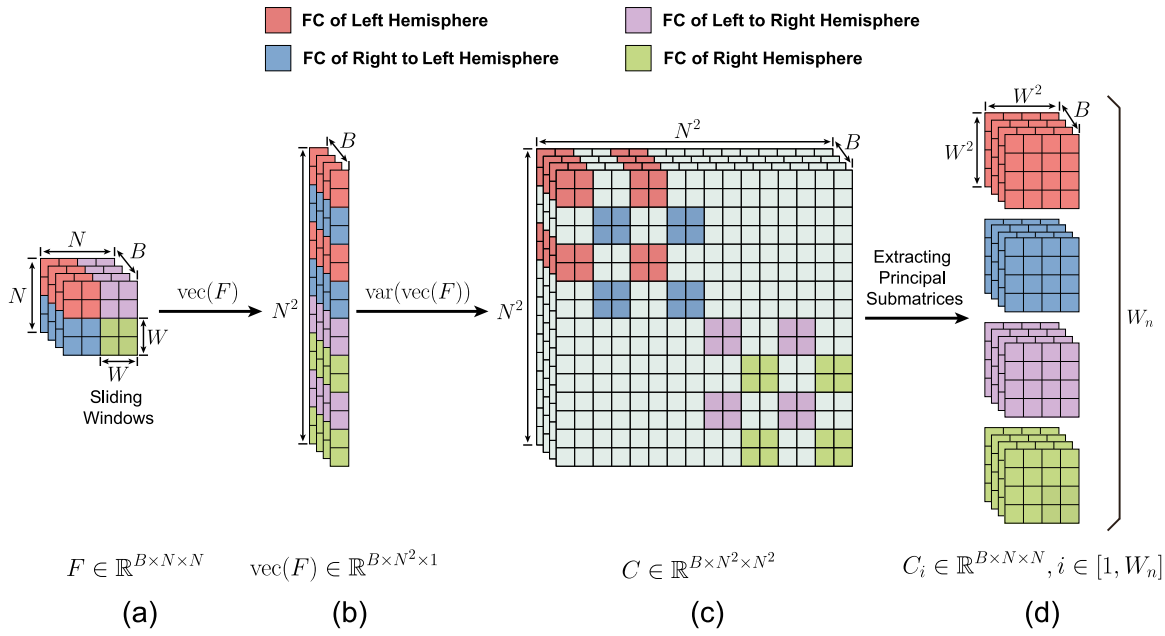

Fig. 5. The illustration of the multi-region selection mechanism (in hemispheric scale). (a) The FC matrix $F \in \mathbb{R}^{B\times N\times N}$ is output from the BiRe module. We apply a sliding window of size $W \times W$ with stride W ($W = N/2$), dividing F into four regions: intra-left hemisphere, left-to-right, right-to-left, and intra-right hemisphere. (b) The matrix F is vectorized into $\mathrm{vec}(F) \in \mathbb{R}^{B\times N^2\times 1}$. (c) Using the vector from (b), we compute the covariance matrix C corresponding to the divided random matrices from (a). (d) We select principal submatrices corresponding to the regions segmented in (a), and concatenate them to obtain W_n SPD matrices of size $W^2 \times W^2$, reflecting brain network activity across different regions.

if $i < j$, otherwise $T_{ij} = 0$. $\odot$ denotes the Hadamard product, and $\mathrm{vec}(\cdot)$ refers to the vectorization operation, excluding zero elements. The fusion and classification operation performed at the fully connected layer with the softmax layer can be represented as:

$$Prob = \mathrm{Softmax}\left(W \cdot \mathrm{concat}\left(V_w, V_h, V_l\right) + b\right), Prob \in \mathbb{R}^Z, \tag{16}$$

where $Prob$ represents the final classification probability, W is the weight matrix of the fully connected layer, and b is the bias vector. (V_w, V_h, V_l) are the brain network features extracted from the whole-brain, hemispheric, and local region scales, respectively. $\mathrm{concat}(\cdot)$ denotes the concatenation of outputs from all scales along the feature dimension, followed by batch normalization. Z is the total number of classification classes.

4. Experiment

4.1. Datasets

To validate the applicability and explainability of MFMR-FN in various aBCI tasks, we used two publicly available datasets for evaluation and testing, including SEED for emotion recognition and MODMA for depression recognition. These datasets were collected using EEG caps with 62 and 128 electrode channels, respectively, further allowing us to verify the robustness of our method across different aBCI devices. The two datasets are described in Table 1.

4.1.1. SEED

The SEED database [32] consists of data from 15 participants (7 male and 8 female). During the experiment, each participant was asked to watch 15 movie clips with different emotional labels (five neutral, positive, and negative), with each clip lasting around four minutes. All participants were required to perform three experimental sessions, with a gap of at least one week between sessions. The EEG signals were recorded using the ESI NeuroScan System2 at a sampling rate of 1000 Hz from a 62-channel electrode cap. For EEG signal processing, the raw EEG data were downsampled to 200 Hz and bandpass-filtered between 0.5 Hz and 75 Hz. We employed an ICA-based strategy to remove artifacts such as electrooculography (EOG) interference. Furthermore, we remove both channels CB1 and CB2. To minimize non-experimental interference, we selected relatively stable mid-session data for each trial, resulting in three-minute EEG segments. Subsequently, a non-overlapping sliding window of 6 s was applied, dividing each trial into 30 segments.

Table 1
The description of SEED and MODMA datasets.

	SEED	MODMA
Classification Categories	Emotion 3 emotions (neutral, positive, negative)	Depression 2 groups (MDD and NC)
Participants	15 (each has 3 emotions)	53 (24 MDD, 29 NC)
Device	ESI NeuroScan 62-channel	Net Station 128-channel
Sampling rate	1000 Hz	250 Hz
Preprocessing	Bandpass (0.5–75 Hz), CB1/CB2 removal, Sliding window (6 s)	Bandpass (0.5–75 Hz), Notch (50 Hz) Sliding window (4 s)

4.1.2. MODMA

The MODMA dataset is a public dataset for EEG-based depression research [33]. The dataset consists of 24 patients with major depressive disorder (MDD) and 29 normal control (NC). The EEG signals were collected using a 128-channel electrode cap at a sampling rate of 250 Hz through the Net Station system. To eliminate baseline drift and electrical interference, we applied a 0.5 Hz to 75 Hz bandpass filter and a 50 Hz notch filter. Subsequently, the EEG signals were re-referenced using the Common Average Reference (CAR) technique, and an ICA-based approach was employed to remove artifacts such as EOG. Similar to the SEED dataset, we selected relatively stable mid-session data for each trial, resulting in EEG segments of two minutes. A non-overlapping sliding window of 4 s was then used, dividing each trial into 30 segments.

Table 2
Details of training hyperparameter settings for different datasets.

Hyperparameters	SEED	MODMA
Adjacency matrix parameter σ	1.5	1.5
Chebyshev polynomial order r	3	2
Transformation matrix W_n size	60 × 60	128 × 128
Rectification threshold ϵ	1e−8	1e−5
Learning rate	1e−4	1e−3
Batch size	32	32
Number of epochs	500	300
Hidden nodes in FC layer	64	128
Loss function	Cross-entropy	Cross-entropy
Optimizer	Adam	Adam

Table 3
Performance comparison of the proposed method with state-of-the-art methods on SEED dataset.

Research	Method	Accuracy
Li et al. [57]	DAN	83.81/8.56
Zhong et al. [58]	RGNN	85.30/6.72
Li et al. [59]	GMSS	86.52/6.22
Guo et al. [60]	HVF_2N-DBR	89.33/10.13
Xu et al. [61]	GRU-Conv	87.04/13.35
Li et al. [62]	GUSA	91.77/5.91
Guo et al. [63]	FC-FAN	92.12/3.32
Chen et al. [64]	GDDN	92.54/03.65
Our method	MFMR-FN	**94.22/3.54**

4.2. Setups

Subject-independent experiments are better suited for real-world applications as they more accurately evaluate the model's predictive performance on new participants. Therefore, this study focuses on subject-independent experiments for both emotion recognition and depression detection. We employed the Leave-One-Subject-Out (LOSO) cross-validation strategy for the experiments. Specifically, the data from each subject was used as the test set, while the data from the other participants were used as the training set. This process was repeated until each subject's data had been tested once. The final evaluation metrics were computed by averaging the results of all subjects, with both the mean and standard deviation reported (i.e., "mean/std"). For the SEED dataset, we selected accuracy, recall, F1-score, and AUC as evaluation metrics. For the MODMA dataset, we evaluated performance using accuracy, sensitivity, specificity, and F1 score, which are commonly used metrics for binary disease detection. The experiments were conducted on two NVIDIA GeForce RTX 4090 GPUs, utilizing CUDA 12.1 through the PyTorch API. Table 2 provides the hyperparameter configurations for training on the two datasets.

4.3. Baselines

Our method is compared with several baselines.

DAN [57]: Deep adaptation network (DAN) has a five-layer structure with the first three as fully connected layers for feature extraction, and the last two employing multi-kernel maximum mean discrepancies, allowing for effective cross-subject emotion recognition based on EEG.

RGNN [58]: Regularized graph neural network (RGNN) utilizes graph neural networks to handle EEG-based emotion recognition.

GMSS [59]: Graph-based multi-task self-supervised learning (GMSS) model integrates spatial and frequency jigsaw puzzle tasks with contrastive learning to improve EEG emotion recognition.

HVF_2N-DBR [60]: Horizontal and vertical feature fusion network based on different brain regions (HVF2N-DBR) model partitions EEG signals into four brain regions and uses a HVF2M module to capture features in multiple orientations.

GRU-Conv [61]: Hybrid gated recurrent unit and convolutional neural network (GRU-Conv) combines GRU for capturing temporal features with CNN for extracting spatial information.

GUSA [62]: Graph-based unsupervised subdomain adaptation (GUSA) model employs a multi-level graph convolution module to dynamically model EEG channels, alongside node-wise, class-level, and emotion-wise domain constraints to extract more fine-grained EEG representations.

FC-FAN [63]: Functional connectivity-enhanced feature-grouped attention network (FC-FAN) utilizes TFGAM and F2GAM attention modules to enhance key temporal and spatial features, optimizing EEG-based emotion recognition.

GDDN [64]: Graph domain disentanglement network (GDDN) includes a graph domain disentanglement module, a classifier aggregation module, and an auxiliary supervision module to preserve feature integrity and improve model robustness.

EDRF [65]: Explainable depression recognition framework (EDRF) leverages a hybrid architecture combining 1DCNN, LSTM, and GCN for comprehensive spatiotemporal feature extraction from EEG data.

mKTAChSel [66]: Modified kernel-target alignment channel selection (mKTAChSel) algorithm refines EEG channel selection method for depression detection. The mKTAChSel aligns channel features with target labels using an optimized KTA objective function.

SGP-SL [67]: Self-attention graph pooling with soft label (SGP-SL) model refines local and global EEG channel connections with self-attention pooling and employs soft labels to enhance discrimination, which significantly improves the accuracy of depression detection.

TPTLP [68]: Twin pascal's triangles lattice pattern (TPTLP) model uses a handcrafted lattice structure inspired by Pascal's triangle to extract unique textural features from EEG signals. TPTLP provides a computationally efficient, lightweight alternative to depression classification.

MS^2-GNN [69]: Modal-shared modal-specific graph neural network (MS^2-GNN) integrates shared and specific GNNs for EEG-based depression detection. It applies multimodal attention for compact feature representation and uses a reconstruction network to preserve data integrity.

CNN-GRU [70]: Spatiotemporal convolutional neural network-gated recurrent unit (CNN-GRU) model transforms EEG signals into brain maps that integrate spatial, frequency, and temporal features. The streamlined structure of CNN-GRU improves depression prediction accuracy while remaining computationally efficient.

SSPA-GCN [71]: Secondary subject partitioning and attention mechanism based on graph convolution network (SSPA-GCN) uses a channel-feature attention module to prioritize key EEG features and a sub-domain clustering method to improve depression detection accuracy.

5. Results

In this section, we provide a detailed description of our experiments, covering five aspects: the comparison with existing methods (Section 5.1), explainability analysis of the model (Section 5.2), comparison with traditional FC extraction methods (Section 5.3), ablation experiments of the encoding and decoding networks (Section 5.4), and performance evaluation based on confusion matrices (Section 5.5).

5.1. Comparison with different methods

In this section, we compared the proposed MFMR-FN with state-of-the-art methods on the SEED and MODMA datasets. Since most of the existing methods published only accuracy in the SEED dataset, we performed the comparison mainly for accuracy. For the MODMA dataset, since existing methods typically report accuracy, sensitivity, specificity, and F1 score, we also performed a full comparative analysis on these metrics. These results are shown in Tables 3 and 4. As shown in Table 3, compared to existing methods, our method achieved the highest accuracy of 94.22%, with improvements ranging from 1.68% to 10.41%.

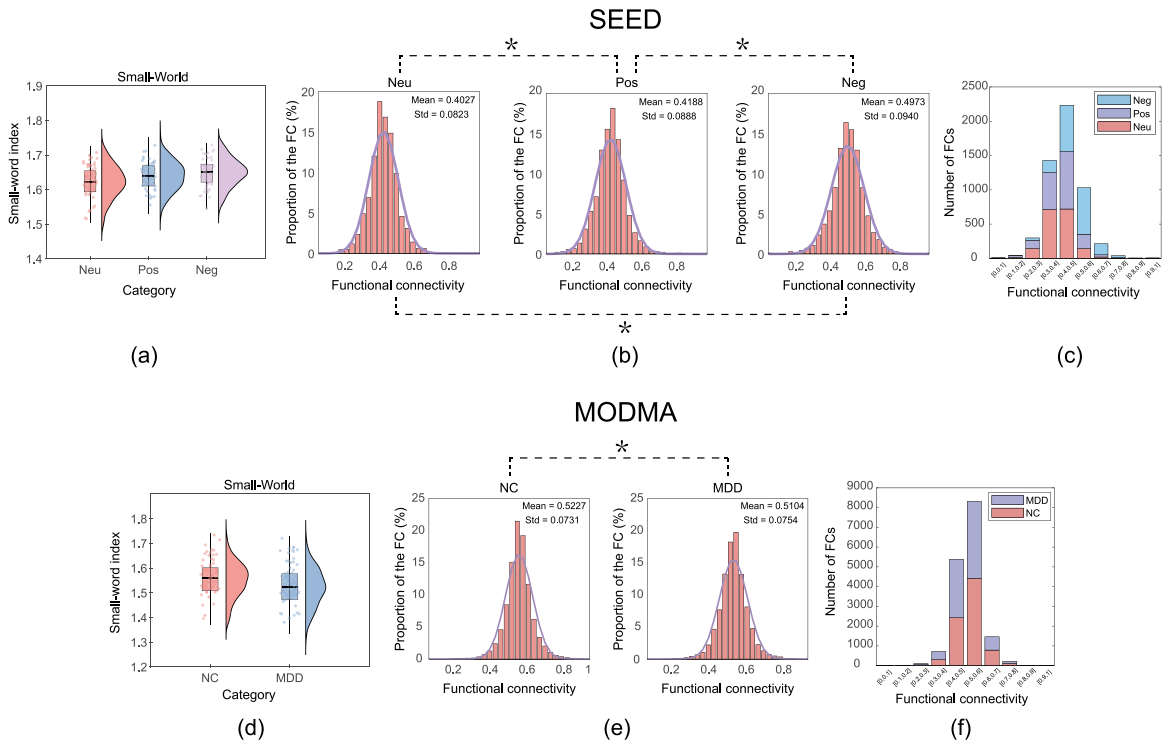

Fig. 6. Small-world and connectivity strength analysis in SEED and MODMA datasets. SEED: (a) small-world properties under different emotional states, (b) probability density function of functional connectivity, and (c) histogram of connectivity strength. MODMA: (d)–(f) represent the corresponding results for healthy individuals and MDD patients.

Table 4
Performance comparison of the proposed method with state-of-the-art methods on MODMA dataset.

Research	Method	Accuracy	Sensitivity	Specificity	F1-score
Shen et al. [65]	EDRF	76.50	75.70	77.30	75.90
Shen et al. [66]	mKTAChSel	81.60	–	–	89.97
Chen et al. [67]	SGP-SL	84.91	87.50	–	84.00
Tasci et al. [68]	TPTLP	83.96	76.14	90.41	81.10
Chen et al. [69]	MS^2-GNN	86.49	87.50	–	84.85
Liu et al. [70]	CNN-GRU	89.63	90.24	89.63	90.19
Zhang et al. [71]	SSPA-GCN	92.87	92.00	**93.59**	92.12
Our method	MFMR-FN	**93.96**	**94.95**	92.78	**93.97**

In Table 4, we compared our MFMR-FN with various machine learning and deep learning methods. Our model also achieved the best results in terms of accuracy (93.96%), sensitivity (94.95%), and F1-score (93.97%), outperforming existing methods. This demonstrates that the proposed MFMR-FN shows excellent classification performance and generalization ability in aBCI tasks, compared to state-of-the-art baselines.

5.2. Explainability analysis

The proposed encoding network captures multi-frequency fused FC and its output FC is highly explainable. This is the main improvement of our network compared to traditional neural networks. Therefore, in this section, we perform extensive experiments to analyze the explainability of the FC learned by the encoding network.

5.2.1. Small-world and connectivity analysis

We first evaluated whether the obtained fusion FC reflects real brain activity by introducing "small-world properties" [72] as an evaluation metric. Existing neuroscience research suggests that small-world properties are an important indicator of brain network efficiency and modularity, and healthy individuals typically exhibit stable small-world properties [73].

The small-world property is defined in Appendix C. The results of the small-world analysis are shown in Fig. 6(a) and (d). We can observe that the FCs extracted from both datasets have significant small-world properties ($\sigma > 1$) [74], consistent with known brain activity patterns. Furthermore, Fig. 6(a) shows that small-world properties are significantly higher in positive and negative compared to neutral emotions. This may reflect more efficient information transmission in the brain during high arousal emotions [75]. Fig. 6(d) shows a reduction in small-world properties in depressed patients, indicating the reduced information transfer efficiency and disrupted modular structure of the patient's brain [76]. We further analyzed the connection strength of the obtained fusion FC. Fig. 6(b) and (c) show the probability density function and histogram of connectivity strength for three emotions in the SEED dataset. Fig. 6(e) and (f) display the distribution in MDD patients and healthy individuals from the MODMA dataset. As shown in Fig. 6(b) and (e), the extracted FC in both datasets follows a normal distribution similar to actual brain networks. This finding deserves our attention because although brain FC is widely known to follow a normal distribution [77], traditional methods struggle to capture it, whereas our model accurately captures this physiological feature. Additionally, as shown in Fig. 6(b) and (c), negative emotions exhibit the strongest connectivity, which may be due to the widespread neural activity associated with negative emotional regulation [78]. Fig. 6(e) and (f) show that brain connectivity in healthy individuals is significantly higher than in MDD patients. which is also consistent with existing research findings [79]. In summary, the small-world and connectivity strength analyses confirm that the FC extracted by encoding network reflects physiologically meaningful brain activity patterns, further validating the explainability of our model.

5.2.2. Brain network topological properties analysis

In this section, we further explore the topological properties of the FC extracted by the encoding network, focusing on functional integration and segregation. These properties help understand complex brain information processing under different cognitive states, verifying the explainability of the obtained fusion FC.

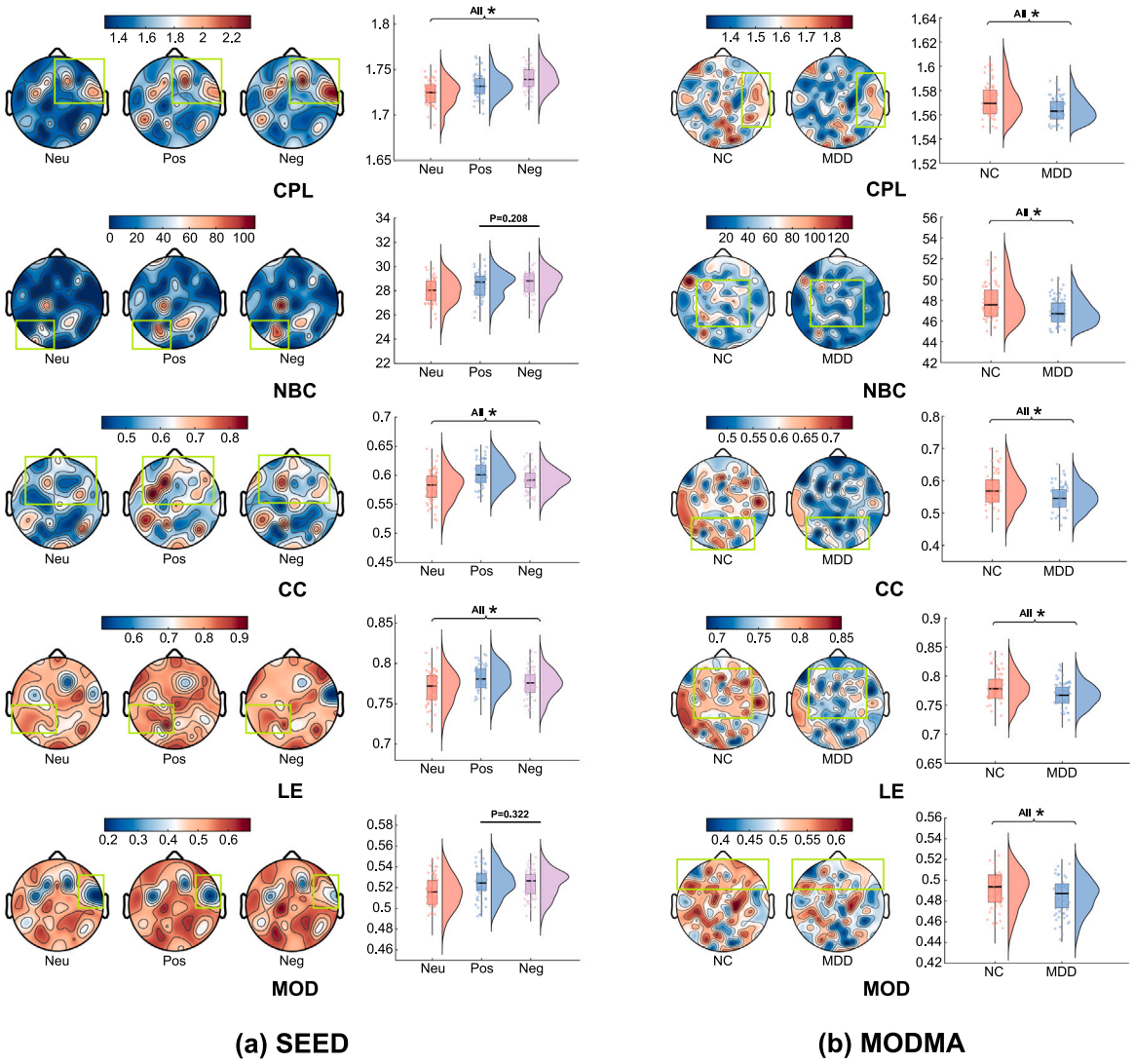

Fig. 7. The analysis of brain network topological properties in (a) SEED and (b) MODMA dataset. In each figure, the left brain maps show feature values for each EEG channel, with green boxes highlighting regions of significant differences. The right violin graphs illustrate the value distributions, where "All ∗" represents significant differences ($p < 0.05$) in both between-group comparisons and pairwise post-hoc comparisons. The horizontal lines with p-values indicate non-significant differences between the two groups.

We first applied the Cluster Span Threshold [80] to select a threshold and binarize the FC. Subsequently, we calculated functional integration metrics, including Characteristic Path Length (CPL) and Network Betweenness Centrality (NBC), as well as functional segregation metrics such as Clustering Coefficient (CC), Local Efficiency (LE), and Modularity (MOD). These metrics were calculated using the Brain Connectivity Toolbox [81] in the MATLAB environment. The formulas and descriptions of these metrics are provided in Appendix C.

We calculated the metrics for functional integration and separation on both datasets, and the results are presented in Fig. 7. The brain maps in Fig. 7 demonstrate the feature values of each EEG channel. For the SEED dataset, one-way ANOVA was used for inter-group comparisons, while independent sample t-tests with Bonferroni correction were applied to the MODMA dataset. Fig. 7(a) shows that both positive and negative emotions exhibit more significant functional integration and segregation compared to neutral emotions. Specifically, negative emotions exhibit higher integration features (CPL, NBC), while positive emotions show higher segregation features (CC, LE, MOD). These results suggest that negative emotions enhance the brain's integrative properties, while positive emotions strengthen local connectivity. Fig. 7(b) shows that MDD has reduced values across all metrics, indicating weakened global and local information processing.

In summary, these results confirm that the FC extracted by the encoding network effectively captures brain network integration and segregation patterns, demonstrating strong inter-group differences in emotion recognition and depression analysis. The model shows high biological explainability, and effectively relieves the 'black box' problem in deep learning.

5.3. Comparison with other FC methods

In this section, a comparison is made between the FC extracted by encoding network and other FC extraction methods, such as Coherence (COH), Pearson Correlation Coefficient (COR), and Phase Locking Value (PLV) [82]. The results are shown in Fig. 8, where the sparsity level is set to 0.05 for the SEED dataset and 0.01 for the MODMA dataset. As shown in Fig. 8, compared to traditional FC extraction method, the obtained fusion FC from our model captures more significant inter-group differences across both datasets. Furthermore, we conducted a performance evaluation of the obtained fusion FC against traditional methods using multiple classifiers, and the results are listed in Table 5. We vectorized these FC matrices and input them into classical classifiers.

Table 5
Performance comparison of the obtained fusion FC and traditional methods using multiple classifiers in SEED and MODMA datasets.

Method	Classifier	SEED				MODMA			
		Accuracy	Recall	F1-score	AUC	Accuracy	Sensitivity	Specificity	F1-score
COH	KNN	51.26/8.37	51.49/7.65	49.65/9.15	69.10/5.25	59.72/9.98	65.65/12.12	54.53/13.76	59.99/9.00
	LDA	60.30/6.11	60.26/5.78	59.57/6.16	75.13/3.43	60.25/11.12	74.13/9.25	52.12/10.46	61.16/10.26
	SVM	63.11/5.38	63.23/5.81	62.75/5.72	72.37/4.15	66.15/9.92	76.94/10.88	53.22/12.43	65.56/10.15
	RF	67.41/6.23	67.97/6.42	67.17/6.27	76.20/3.18	70.27/11.93	80.87/10.06	64.19/8.68	70.76/10.67
	MLP	69.07/5.50	67.60/3.89	63.88/7.13	80.22/5.88	74.04/11.69	81.42/12.42	70.14/7.89	74.73/13.34
COR	KNN	55.55/6.62	55.97/5.53	53.69/5.89	71.56/3.89	57.54/10.11	65.70/8.45	50.28/13.01	58.58/8.01
	LDA	64.27/6.95	64.89/7.18	64.09/6.82	76.94/4.68	56.02/11.39	65.46/8.03	51.59/11.13	59.39/11.51
	SVM	70.79/6.86	70.93/6.78	70.34/6.73	80.65/3.75	59.30/12.89	53.09/11.62	67.80/12.90	57.41/11.36
	RF	72.30/6.25	72.41/6.54	71.78/6.90	87.99/3.85	63.82/9.83	77.73/10.70	51.93/11.09	67.63/10.11
	MLP	74.25/5.82	74.38/3.59	72.80/4.72	92.16/3.86	67.58/10.46	77.62/12.17	59.13/10.39	73.82/8.94
PLV	KNN	54.95/5.80	55.57/6.52	52.84/5.79	70.28/5.39	57.56/15.33	54.13/14.92	69.02/16.81	54.03/10.07
	LDA	64.72/5.48	64.93/4.69	64.33/4.77	74.76/3.93	58.04/11.73	57.79/12.93	63.48/10.55	57.33/8.94
	SVM	71.27/5.06	70.76/5.89	70.88/5.98	82.06/3.36	64.66/8.86	77.53/11.94	51.92/10.23	68.32/9.32
	RF	72.15/4.62	71.55/4.52	71.36/4.58	79.06/3.10	66.15/9.92	76.94/10.88	53.22/11.43	65.56/8.15
	MLP	73.92/4.35	73.88/3.69	73.12/5.12	87.07/1.61	72.76/8.52	87.81/7.00	59.29/10.96	78.47/7.31
Our FC	KNN	69.00/6.18	67.01/5.80	64.69/6.19	77.28/5.02	63.15/11.17	76.67/14.48	50.09/14.84	66.05/12.06
	LDA	71.83/6.06	71.78/6.35	69.54/5.96	80.17/3.12	69.95/11.96	65.47/11.37	73.69/12.77	79.77/7.77
	SVM	77.02/4.88	77.51/4.77	76.13/4.87	82.45/3.51	74.82/9.84	89.22/8.10	58.70/10.83	79.35/7.30
	RF	81.14/3.06	81.35/2.74	81.16/3.17	87.93/3.43	78.41/8.60	86.93/10.61	**76.46/9.27**	80.34/8.65
	MLP	**85.64/3.69**	**85.76/4.03**	**85.37/4.08**	**94.42/1.90**	**82.69/7.25**	**90.43/8.93**	75.95/8.91	**84.78/6.90**

Table 6
Performance comparison of FC for different frequency bands in SEED and MODMA datasets.

Frequency	SEED				MODMA			
	Accuracy	Recall	F1-score	AUC	Accuracy	Sensitivity	Specificity	F1-score
Delta	88.29/6.17	88.69/6.28	88.19/6.27	97.28/2.45	75.77/10.39	86.07/11.75	67.90/10.78	78.83/9.32
Theta	82.63/8.67	82.56/9.11	82.22/9.53	93.77/5.87	81.24/9.91	89.42/10.28	75.76/9.39	83.55/7.25
Alpha	76.88/10.50	77.54/9.85	76.55/10.22	90.04/7.32	88.73/6.99	91.40/5.76	86.18/6.67	88.47/3.73
Beta	86.97/4.62	86.84/4.45	86.58/4.74	97.52/1.32	85.40/9.69	87.53/8.59	85.22/6.87	85.83/5.26
Gamma	90.56/5.57	90.44/5.33	90.15/6.20	97.00/1.58	86.53/7.29	89.99/6.14	88.36/6.75	88.34/3.98
Fusion	**94.22/3.54**	**94.26/3.53**	**93.97/3.70**	**99.46/0.71**	**93.96/4.64**	**94.95/3.43**	**92.78/3.29**	**93.97/2.71**

As shown in Table 5, the obtained fusion FC outperforms traditional methods across all classifiers in both the SEED and MODMA datasets. This improvement highlights that, compared to traditional FC extraction methods, our model is able to extract FC features with greater discriminability and robustness.

5.4. Ablation studies

5.4.1. Ablation studies on the encoding networks

In Section 3.2, our encoding network fuses FC across five frequency bands to produce the fused FC. To verify the effectiveness of multi-frequency FC fusion, we conducted ablation experiments comparing the performance of individual frequency bands against our fused FC. The results are shown in Fig. 9, where the sparsity level is set to 0.05 for the SEED dataset and 0.01 for the MODMA dataset. The results demonstrate that while individual bands show advantages in capturing specific inter-group differences, the fused FC comprehensively integrates unique features from all bands, significantly improving discriminability.

Table 6 shows the classification results of FCs in different frequency bands on SEED and MODMA datasets. The experimental results show that the fused FC achieved the best recognition performance in both datasets. In the SEED dataset, the fusion FC achieves an accuracy of 94.22%, with an improvement of 3.66–17.34% compared to the single-band performance. In the MODMA dataset, the obtained fusion FC increased accuracy to 93.96%, outperforming individual bands by 5.23–18.19%. These findings validate the effectiveness and necessity of the proposed multi-frequency fusion algorithm in aBCI tasks.

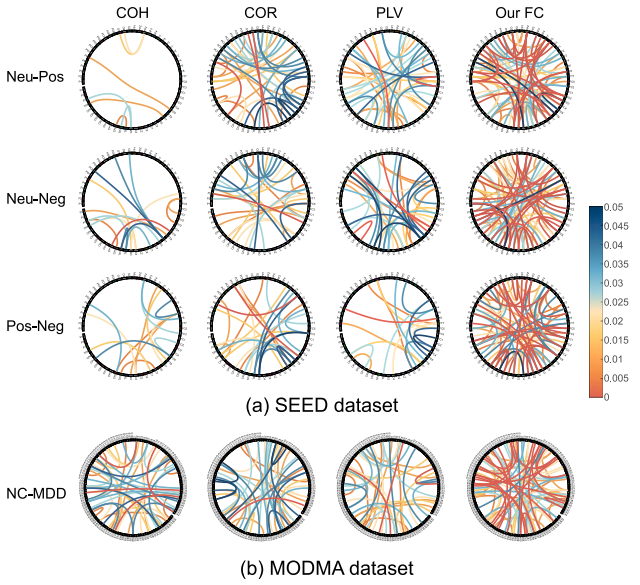

Fig. 8. Comparison of different FC extraction methods in (a) SEED and (b) MODMA datasets. The circular plots illustrate the FC patterns, where the nodes around the circle represent EEG channels and lines indicate significant connectivity differences between groups. The color of the lines indicates the p-value.

5.4.2. Ablation studies on the decoding networks

In Section 3.3, the decoding network further improves recognition performance by fusing FC across whole-brain (V_w), hemispheric(V_h),

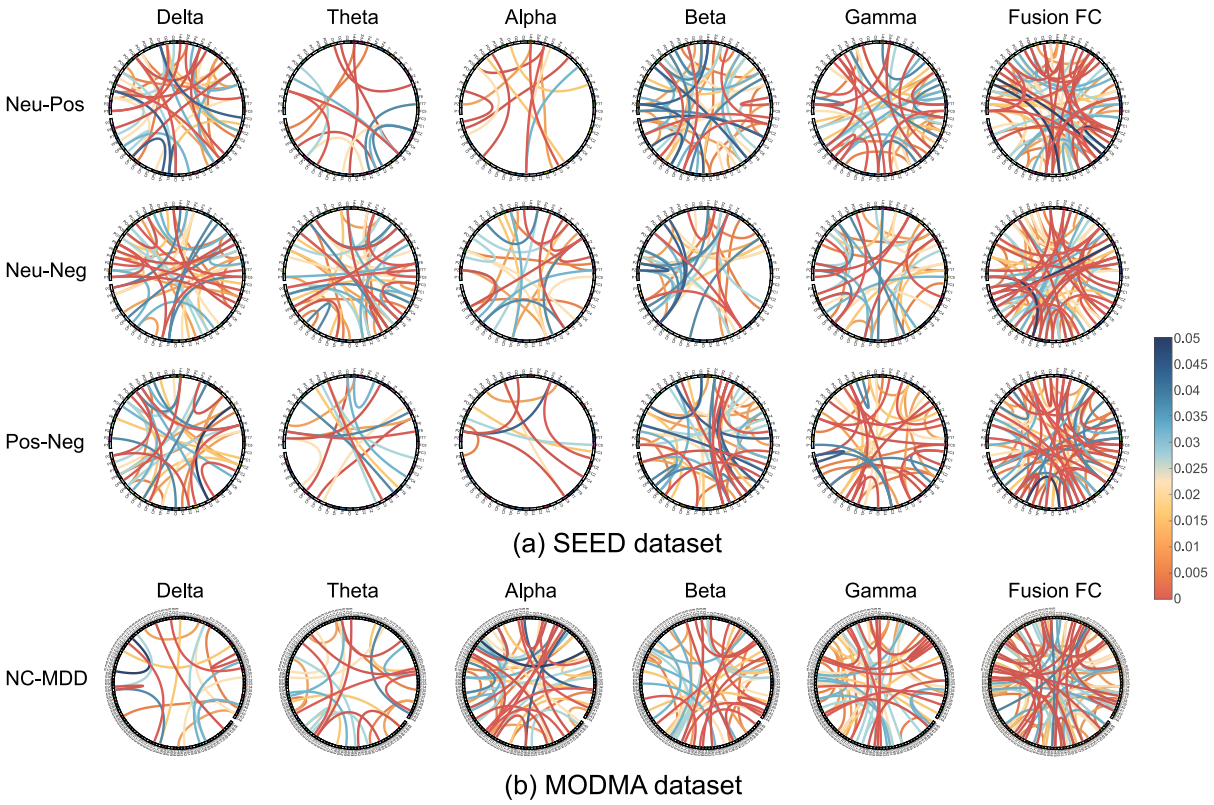

Fig. 9. Comparison of individual frequency FC and fused FC in (a) SEED and (b) MODMA datasets. The circular plots illustrate the FC patterns, where the nodes around the circle represent EEG channels and lines indicate significant connectivity differences between groups. The color of the lines indicates the *p*-value.

Table 7
Performance comparison of different brain region fusion methods in SEED dataset.

Region	Accuracy	Recall	F1-score	AUC
V_w	84.91/7.46	84.40/8.13	84.18/8.22	94.47/4.84
V_h	85.64/7.40	85.75/6.84	85.35/7.09	97.89/3.55
V_l	87.41/5.06	87.85/4.65	87.29/5.07	97.03/2.00
V_w+V_h	90.65/5.46	90.73/5.53	90.49/5.57	98.18/1.34
V_w+V_l	92.15/5.03	92.07/4.64	91.84/5.17	98.64/1.03
V_h+V_l	91.39/4.61	91.50/4.94	91.21/4.75	98.85/1.63
V_w+V_h+V_l	**94.22/4.54**	**94.26/4.53**	**93.97/4.70**	**99.46/0.71**

Table 8
Performance comparison of different brain region fusion methods in MODMA dataset.

Region	Accuracy	Sensitivity	Specificity	F1-score
V_w	83.27/8.06	87.29/7.16	83.71/7.24	82.54/8.92
V_h	87.73/6.82	91.74/7.41	84.89/6.63	87.98/7.38
V_l	86.02/6.43	90.65/6.45	84.72/9.87	87.45/3.20
V_w+V_h	90.27/5.45	93.01/5.94	87.00/4.22	90.18/4.44
V_w+V_l	89.51/3.06	91.80/4.68	86.94/6.46	90.14/3.63
V_h+V_l	91.67/3.77	92.92/4.11	89.41/5.58	92.40/3.27
V_w+V_h+V_l	**93.96/4.64**	**94.95/3.43**	**92.78/3.29**	**93.97/3.71**

and local region(V_l) scales. To verify the effectiveness of the proposed multi-region selection mechanism, we conducted ablation experiments comparing performance using each scale independently, as well as their various combinations. The results are shown in Tables 7 and 8. As shown in Table 7, in the SEED dataset, the model fusing whole-brain, hemispheric, and local region achieved best performance across all evaluation metrics. As shown in Table 8, in the MODMA dataset, the proposed strategy also performed well with an accuracy of 93.96%. These results demonstrate that the proposed multi-region selection strategy fuses brain network features from coarse to fine, significantly improving the performance of the decoding network.

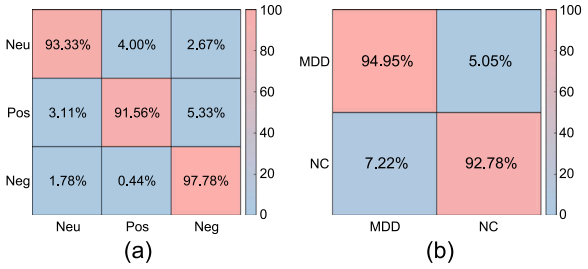

Fig. 10. The confusion matrices of the proposed method in the (a) SEED and (b) MODMA datasets.

5.5. Confusion matrix analysis

Fig. 10 presents the confusion matrices of the proposed MFMR-FN for the SEED and MODMA datasets. As shown in Fig. 10(a), in the SEED dataset, the model performed well in recognizing negative and neutral emotions in the SEED dataset, with accuracies of 97.78% and 93.33%, respectively. However, some confusion existed between positive and negative emotions, particularly with positive emotion being misclassified as negative. This phenomenon may be attributed to the high physiological arousal levels present in both emotional states, leading to similar activation patterns in certain brain regions or similar brain network features. As shown in Fig. 10(b), the model can accurately distinguish between the two groups, with accuracy maintained at a high level. Nevertheless, some NC samples were misclassified as MDD, which may be due to these NC individuals exhibiting some mild depressive traits during testing.

In summary, MFMR-FN shows high accuracy on both the SEED and MODMA datasets, indicating its robustness in handling different aBCI tasks.

6. Conclusion

This paper proposes an explainable multi-frequency and multi-region fusion model (MFMR-FN) for an aBCI, significantly improving recognition accuracy and explainability in various aBCI tasks. The MFMR-FN consists of a multi-frequency encoding network and a multi-region decoding network. The encoding network is proposed based on FC-GCN and BiRe modules, which adaptively encodes EEG data into FC matrices with neurobiological information. Furthermore, a novel multi-frequency fusion algorithm based on Riemannian geometry is designed to guide the network in achieving cross-frequency feature fusion. The encoding network outputs FC matrices with explainability, allowing for multi-dimensional brain network analysis. The decoding network employs a multi-region selection mechanism and a multi-scale Riemannian network, integrating brain network features from the whole brain, hemispheres, and local regions for coarse-to-fine emotional decoding. We validated the model's performance and explainability on public datasets for emotion recognition and depression detection. Experimental results show that MFMR-FN outperforms existing methods on multiple metrics. Furthermore, the model provides explainable FC features, revealing brain network patterns under different emotions and abnormal brain connectivity in depression.

The proposed MFMR-FN is expected to improve the practicality and reliability of aBCIs in real-world and clinical applications. Given the explainability of the proposed method, we will analyze the relationship between concept mapping-represented cognitive states [83–85] and brain signals in future work to bridge the gap between cognitive theory and neural representation.

Although the proposed model has the above advantages, there are still limitations and constraints. First, the proposed encoding and decoding networks have not fully considered the relative contributions of different frequency bands and brain regions in emotion recognition. Future research could incorporate advanced attentional mechanisms such as self-attention or multi-head attention, specifically designed for multi-frequency bands and multiple brain regions. These mechanisms have the potential to enhance the effectiveness of information fusion, thereby improving the performance of the aBCI. Second, the proposed model primarily relies on static functional connectivity, without extracting dynamic information transfer or causal relationships between brain regions. Future studies may introduce effective connectivity analysis, such as Granger Causality and Dynamic Causal Modeling (DCM), to capture the complex dynamic causal relationships between brain regions, thus further enhancing the explainability of the model.

CRediT authorship contribution statement

Tao Wang: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Rui Mao:** Writing – review & editing, Methodology, Conceptualization. **Shuang Liu:** Writing – review & editing, Supervision, Investigation, Conceptualization. **Erik Cambria:** Writing – review & editing, Supervision, Methodology. **Dong Ming:** Writing – review & editing, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the STI2030-Major Projects (No. 2021ZD0202000), and the National Natural Science Foundation of China (No. 81925020).

Appendix A. Proof of Theorem 1

First, we prove symmetry of the matrix $F = \left[f_{ij}\right] \in \mathbb{R}^{n\times n}$. Due to the symmetry of the Euclidean distance, i.e., $\|\mathbf{x}_i - \mathbf{x}_j\|_2 = \|\mathbf{x}_j - \mathbf{x}_i\|_2$, for all i, j, we obtain the following equation:

$$f_{ij} = \exp\left(-\beta\|\mathbf{x}_i - \mathbf{x}_j\|_2^2\right) = \exp\left(-\beta\|\mathbf{x}_j - \mathbf{x}_i\|_2^2\right) = f_{ji} \tag{17}$$

where $\beta = \gamma^2$, therefore F is a symmetric matrix.

Second, we prove the positive definiteness of the matrix F. Let $\mathbf{w} = [w_1, w_2, \ldots, w_n]^T \in \mathbb{R}^n$ denotes an arbitrary nonzero vector, and the quadratic form $\mathcal{P}$ is defined as follows:

$$\mathcal{P} = \mathbf{w}^T F \mathbf{w} = \sum_{i,j=1}^{n} w_i w_j f_{ij}. \tag{18}$$

Using the definition of f_{ij}, we can represent $\mathcal{P}$ as follows:

$$\mathcal{P} = \sum_{i,j=1}^{n} w_i w_j \exp\left(-\beta\|\mathbf{x}_i - \mathbf{x}_j\|_2^2\right), \tag{19}$$

Using the expansion of the norm squared, we rewrite $\mathcal{P}$ as follows:

$$\mathcal{P} = \sum_{i,j=1}^{n} w_i w_j \exp\left(-\beta\left(x_i - x_j\right)^T\left(x_i - x_j\right)\right) \tag{20}$$

Applying the characteristic function of a multivariate normal distribution, let $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, 2\beta\mathbf{I})$, then the expectation expression can be rewritten as follows:

$$\mathcal{P} = \mathbb{E}_{\mathbf{Z}}\left[\sum_{i,j=1}^{n} w_i w_j \exp\left(i(\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{Z}\right)\right] \tag{21}$$

By the linearity property of the exponential function, this simplifies further to:

$$\mathcal{P} = \mathbb{E}_{\mathbf{Z}}\left[\left(\sum_{i=1}^{n} w_i \exp\left(i\mathbf{x}_i^T\mathbf{Z}\right)\right)\left(\sum_{j=1}^{n} w_j \exp\left(-i\mathbf{x}_j^T\mathbf{Z}\right)\right)\right] \tag{22}$$

Let $u_i = \exp(i\mathbf{x}_i^T\mathbf{Z})$, then:

$$\mathcal{P} = \mathbb{E}_{\mathbf{Z}}\left[\left(\sum_{i=1}^{n} w_i u_i\right)^2\right] > 0 \tag{23}$$

For all non-zero vector $\mathbf{w} \in \mathbb{R}^n$, $\mathcal{P} = \mathbf{w}^T F\mathbf{w} > 0$, so we conclude that F is a symmetric positive definite (SPD) matrix. Therefore, Theorem 1 is established.

Appendix B. Proof of Theorem 2

First, we prove that the covariance matrix $C \in \mathbb{R}^{n\times n}$ is a symmetric positive definite (SPD) matrix. The covariance matrix C is defined as follows:

$$C = \mathrm{Var}(\mathrm{vec}(F)) = E\left[(\mathrm{vec}(F) - \mu)(\mathrm{vec}(F) - \mu)^T\right] \tag{24}$$

where $\mathrm{vec}(F)$ denotes the vectorization of matrix F, and μ is the mean of $\mathrm{vec}(F)$. E represents the expectation operator. Since C is the expected outer product of the centered vectorized matrix F, C is symmetric. Next, we will prove the positive definiteness of C.

Let $\mathbf{v} = [v_1, v_2, \ldots, v_n]^T \in \mathbb{R}^n$ be an arbitrary non-zero vector. Define the quadratic form $\mathcal{P}$ as follows:

$$\mathcal{P} = \mathbf{v}^T C\mathbf{v} = \mathbf{v}^T E\left[(\mathrm{vec}(F) - \mu)(\mathrm{vec}(F) - \mu)^T\right]\mathbf{v} \tag{25}$$

Due to the linearity of expectation, we can expand the expression as:

$$\mathcal{P} = E\left[\mathbf{v}^T(\mathrm{vec}(F) - \mu)(\mathrm{vec}(F) - \mu)^T\mathbf{v}\right] \tag{26}$$

Assuming F consists of n samples F_i, then:

$$\mathcal{P} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{v}^T(\mathrm{vec}(F_i) - \mu)(\mathrm{vec}(F_i) - \mu)^T\mathbf{v} \tag{27}$$

Table C.1
Brain network metrics for integration and segregation.

Metric	Mathematical formulation	Explanation
Small-World Property (σ)	$\sigma = \frac{C/C_{\text{rand}}}{L/L_{\text{rand}}}$ where C and C_{rand} represent the clustering coefficients of the actual and random networks, respectively, L and L_{rand} represent their characteristic path lengths.	σ measures the balance between high local clustering and short global path length, indicating an efficient small-world network structure.
Characteristic Path Length (CPL)	$CPL = \frac{1}{N}\sum_{i=1}^{N} L_i = \frac{1}{N}\sum_{i=1}^{N} \frac{\sum_{j\in N, j\neq i} d_{ij}}{N-1}$ where L_i is the average shortest path length from node i to all other nodes, d_{ij} is the shortest path length between nodes i and j, and N is the total number of nodes in the network.	CPL reflects the global integration efficiency of the network by indicating the average shortest path length between all pairs of nodes.
Network Betweenness Centrality (NBC)	$NBC = \frac{1}{N}\sum_{i=1}^{N} NBC_i = \frac{1}{N}\sum_{i=1}^{N}\sum_{h\neq i\neq j} \frac{\rho_{hj}(i)}{\rho_{hj}}$ where NBC_i is the betweenness centrality of node i, ρ_{hj} represents the total number of shortest paths between nodes h and j, and $\rho_{hj}(i)$ is the number of those paths that pass through node i.	NBC reflects the integration efficiency of the network by averaging the betweenness centrality of all nodes, highlighting key nodes in network communication.
Clustering Coefficient (CC)	$CC = \frac{1}{N}\sum_{i=1}^{N} CC_i = \frac{1}{N}\sum_{i=1}^{N} \frac{2t_i}{k_i(k_i-1)}$ where CC_i is the clustering coefficient of node i, t_i is the number of triangles involving node i, and k_i is the number of edges connected to node i.	CC measures the likelihood that a node's neighbors are also connected, reflecting the local clustering or segregation in the network.
Local Efficiency (LE)	$LE = \frac{1}{N}\sum_{i=1}^{N} LE_i = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{k_i(k_i-1)} \sum_{j,h\in G_i} \frac{1}{d_{jh}}$ where LE_i is the local efficiency of node i, G_i represents the subgraph formed by the neighbors of node i, and d_{jh} is the shortest path length between nodes j and h within G_i.	LE measures the average efficiency of information transfer within the local neighborhood of a node when that node is removed, reflecting local connectivity and resilience.
Modularity (MOD)	$MOD = \sum_{u\in M}\left[e_{uu} - (\sum_{v\in M} e_{uv})^2\right]$ where M represents the distinct modules within the network, e_{uu} is the ratio of connections within module u, and $\sum_{v\in M} e_{uv}$ is the fraction of connections between nodes in module u and nodes in any module v.	MOD quantifies the extent to which a network can be divided into distinct modules, with high modularity indicating strong functional segregation.

Let $\mathbf{y}_i = \text{vec}(F_i) - \mu$, then:

$$\mathcal{P} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}^T\mathbf{y}_i)^2 \tag{28}$$

Since$(\mathbf{v}^T\mathbf{y}_i)^2 \geq 0$, and for $\mathbf{v} \neq 0$, there exists at least one $\mathbf{y}_i$ such that $\mathbf{v}^T\mathbf{y}_i \neq 0$, then:

$$\mathcal{P} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}^T\mathbf{y}_i)^2 > 0 \tag{29}$$

Thus, for any non-zero vector $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v}^T C\mathbf{v} > 0$, proving that C is positive definite. Therefore, the covariance matrix C is symmetric and positive definite (SPD) matrix. Next, we prove that the principal submatrices of the covariance matrix C are also SPD matrices. Since C is symmetric, any principal submatrix $(C)_i$ must also be symmetric, as removing rows and columns does not affect symmetry. To prove the positive definiteness of the principal submatrix $(C)_i$, let $\mathbf{y} \in \mathbb{R}^w$ be an arbitrary non-zero vector (where w is the dimension of the submatrix). Extend $\mathbf{y}$ to a non-zero vector $\mathbf{x} \in \mathbb{R}^n$ (where n is the dimension of the original matrix), such that $\mathbf{x}$ is zero in the removed dimensions and corresponds to $\mathbf{y}$ in the retained dimensions. Thus, we have:

$$\mathbf{y}^T(C)_i\mathbf{y} = \mathbf{x}^T C\mathbf{x} \tag{30}$$

Since C is positive definite, for any non-zero $\mathbf{x}$, $\mathbf{x}^T C\mathbf{x} > 0$. Therefore, for any non-zero $\mathbf{y}$, $\mathbf{y}^T(C)_i\mathbf{y} > 0$. Therefore, the principal submatrix $(C)_I$ is also SPD matrix, and these submatrices form a connected Riemannian manifold $\mathcal{S}_{++}^w$.

In conclusion, we have proven that the covariance matrix C is SPD matrix, and its set of principal submatrices $(C)_i$ is also SPD matrix, forming a connected Riemannian manifold $\mathcal{S}_{++}^w$. Therefore, Theorem 2 is established.

Appendix C. Brain network metrics

For detailed calculations regarding the small-world property and various metrics of brain network integration and segregation, please refer to Table C.1.

Data availability

Data will be made available on request.

References

[1] D. Wu, B.-L. Lu, B. Hu, Z. Zeng, Affective brain–computer interfaces (abcis): A tutorial, Proc. IEEE 111 (10) (2023) 1314–1332.
[2] K. Chen, X. Yang, C. Fan, W. Zhang, Y. Ding, Semantic-rich facial emotional expression recognition, IEEE Trans. Affect. Comput. 13 (4) (2022) 1906–1916.
[3] R. Mao, Q. Liu, K. He, W. Li, E. Cambria, The biases of pre-trained language models: An empirical study on prompt-based sentiment analysis and emotion detection, IEEE Trans. Affect. Comput. 14 (3) (2023) 1743–1753.
[4] T. Wang, J. Sun, J. Chao, S. Zheng, C. Zhao, C. Wu, H. Peng, A novel gait analysis method based on the pseudo-velocity model for depression detection, in: 2020 IEEE International Conference on E-Health Networking, Application & Services, HEALTHCOM, IEEE, 2021, pp. 1–6.
[5] T. Wang, S. Liu, F. He, W. Dai, M. Du, Y. Ke, D. Ming, Emotion recognition from full-body motion using multiscale spatio-temporal network, IEEE Trans. Affect. Comput. 15 (3) (2024) 898–912.
[6] C. Fan, J. Lin, R. Mao, E. Cambria, Fusing pairwise modalities for emotion recognition in conversations, Inf. Fusion 106 (2024) 102306.
[7] S. Qiu, Y. Chen, Y. Yang, P. Wang, Z. Wang, H. Zhao, Y. Kang, R. Nie, A review on semi-supervised learning for EEG-based emotion recognition, Inf. Fusion (2023) 102190.
[8] F.S. de Aguiar Neto, J.L.G. Rosa, Depression biomarkers using non-invasive EEG: A review, Neurosci. Biobehav. Rev. 105 (2019) 83–93.

[9] E.H. Houssein, A. Hammad, A.A. Ali, Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review, Neural Comput. Appl. 34 (15) (2022) 12527–12557.

[10] G. Li, D. Ouyang, Y. Yuan, W. Li, Z. Guo, X. Qu, P. Green, An EEG data processing approach for emotion recognition, IEEE Sens. J. 22 (11) (2022) 10751–10763.

[11] S.A. Khoshnevis, R. Sankar, Applications of higher order statistics in electroencephalography signal processing: A comprehensive survey, IEEE Rev. Biomed. Eng. 13 (2019) 169–183.

[12] C.E. Valderrama, A comparison between the hilbert-huang and discrete wavelet transforms to recognize emotions from electroencephalographic signals, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2021, pp. 496–499.

[13] T. Xu, L. Xu, H. Zhang, Z. Ji, J. Li, A. Bezerianos, H. Wang, Effects of rest-break on mental fatigue recovery based on EEG dynamic functional connectivity, Biomed. Signal Process. Control 77 (2022) 103806.

[14] H. Wang, X. Liu, H. Hu, F. Wan, T. Li, L. Gao, A. Bezerianos, Y. Sun, T.-P. Jung, Dynamic reorganization of functional connectivity unmasks fatigue related performance declines in simulated driving, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (8) (2020) 1790–1799.

[15] H. Wang, X. Liu, J. Li, T. Xu, A. Bezerianos, Y. Sun, F. Wan, Driving fatigue recognition with functional connectivity based on phase synchronization, IEEE Trans. Cogn. Dev. Syst. 13 (3) (2020) 668–678.

[16] C. Chen, Z. Li, F. Wan, L. Xu, A. Bezerianos, H. Wang, Fusing frequency-domain features and brain connectivity features for cross-subject emotion recognition, IEEE Trans. Instrum. Meas. 71 (2022) 1–15.

[17] X. Liu, T. Li, C. Tang, T. Xu, P. Chen, A. Bezerianos, H. Wang, Emotion recognition and dynamic functional connectivity analysis based on EEG, IEEE Access 7 (2019) 143293–143302.

[18] D. Huang, S. Chen, C. Liu, L. Zheng, Z. Tian, D. Jiang, Differences first in asymmetric brain: A bi-hemisphere discrepancy convolutional neural network for EEG emotion recognition, Neurocomputing 448 (2021) 140–151.

[19] L. Wang, W. Huang, Z. Yang, C. Zhang, Temporal-spatial-frequency depth extraction of brain-computer interface based on mental tasks, Biomed. Signal Process. Control 58 (2020) 101845.

[20] T. Song, W. Zheng, P. Song, Z. Cui, EEG emotion recognition using dynamical graph convolutional neural networks, IEEE Trans. Affect. Comput. 11 (3) (2018) 532–541.

[21] A. Riaz, M. Asad, E. Alonso, G. Slabaugh, DeepFMRI: End-to-end deep learning for functional connectivity and classification of ADHD using fMRI, J. Neurosci. Methods 335 (2020) 108506.

[22] U. Mahmood, Z. Fu, V.D. Calhoun, S. Plis, A deep learning model for data-driven discovery of functional connectivity, Algorithms 14 (3) (2021) 75.

[23] S.-E. Moon, C.-J. Chen, C.-J. Hsieh, J.-L. Wang, J.-S. Lee, Emotional EEG classification using connectivity features and convolutional neural networks, Neural Netw. 132 (2020) 96–107.

[24] E. Cambria, R. Mao, M. Chen, Z. Wang, S.-B. Ho, Seven pillars for the future of artificial intelligence, IEEE Intell. Syst. 38 (6) (2023) 62–69.

[25] S. Han, R. Mao, E. Cambria, Hierarchical attention network for explainable depression detection on Twitter aided by metaphor concept mappings, in: Proceedings of the 29th International Conference on Computational Linguistics, COLING, 2022, pp. 94–104.

[26] M.R. Mowla, R.I. Cano, K.J. Dhuyvetter, D.E. Thompson, Affective brain-computer interfaces: Choosing a meaningful performance measuring metric, Comput. Biol. Med. 126 (2020) 104001.

[27] Z. He, K. Yang, N. Zhuang, Y. Zeng, Processing of affective pictures: A study based on functional connectivity network in the cerebral cortex, Comput. Intell. Neurosci. 2021 (1) (2021) 5582666.

[28] Z.-M. Wang, R. Zhou, Y. He, X.-M. Guo, Functional integration and separation of brain network based on phase locking value during emotion processing, IEEE Trans. Cogn. Dev. Syst. 15 (2) (2020) 444–453.

[29] M.A. Maria, M. Akhand, A.A. Hossain, M.A.S. Kamal, K. Yamada, A comparative study on prominent connectivity features for emotion recognition from EEG, IEEE Access 11 (2023) 37809–37831.

[30] Z.-M. Wang, Z.-Y. Chen, J. Zhang, EEG emotion recognition based on PLV-rich-club dynamic brain function network, Appl. Intell. 53 (14) (2023) 17327–17345.

[31] M. Aslan, M. Baykara, T.B. Alakuş, Analysis of brain areas in emotion recognition from EEG signals with deep learning methods, Multimedia Tools Appl. 83 (11) (2024) 32423–32452.

[32] W.-L. Zheng, J.-Y. Zhu, B.-L. Lu, Identifying stable patterns over time for emotion recognition from EEG, IEEE Trans. Affect. Comput. 10 (3) (2017) 417–429.

[33] H. Cai, Z. Yuan, Y. Gao, S. Sun, N. Li, F. Tian, H. Xiao, J. Li, Z. Yang, X. Li, et al., A multi-modal open dataset for mental-disorder analysis, Sci. Data 9 (1) (2022) 178.

[34] Y. Zhang, H. Liu, D. Zhang, X. Chen, T. Qin, Q. Zheng, EEG-based emotion recognition with emotion localization via hierarchical self-attention, IEEE Trans. Affect. Comput. 14 (3) (2022) 2458–2469.

[35] S. Liu, Z. Wang, Y. An, J. Zhao, Y. Zhao, Y.-D. Zhang, EEG emotion recognition based on the attention mechanism and pre-trained convolution capsule network, Knowl.-Based Syst. 265 (2023) 110372.

[36] H. Gao, X. Wang, Z. Chen, M. Wu, Z. Cai, L. Zhao, J. Li, C. Liu, Graph convolutional network with connectivity uncertainty for EEG-based emotion recognition, IEEE J. Biomed. Health Inf. (2024).

[37] Y. Tang, D. Chen, J. Wu, W. Tu, J.J. Monaghan, P. Sowman, D. Mcalpine, Functional connectivity learning via siamese-based SPD matrix representation of brain imaging data, Neural Netw. 163 (2023) 272–285.

[38] Y. Gao, X. Fu, T. Ouyang, Y. Wang, EEG-GCN: Spatio-temporal and self-adaptive graph convolutional networks for single and multi-view EEG-based emotion recognition, IEEE Signal Process. Lett. 29 (2022) 1574–1578.

[39] K. You, H.-J. Park, Re-visiting Riemannian geometry of symmetric positive definite matrices for the analysis of functional connectivity, NeuroImage 225 (2021) 117464.

[40] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Multiclass brain–computer interface classification by Riemannian geometry, IEEE Trans. Biomed. Eng. 59 (4) (2011) 920–928.

[41] Y.-J. Suh, B.H. Kim, Riemannian embedding banks for common spatial patterns with EEG-based SPD neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2021, pp. 854–862.

[42] Z. Huang, L. Van Gool, A riemannian network for spd matrix learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2017, pp. 2036–2042.

[43] Y. Wang, S. Qiu, X. Ma, H. He, A prototype-based SPD matrix network for domain adaptation EEG emotion recognition, Pattern Recognit. 110 (2021) 107626.

[44] X. Zhang, D. Lu, J. Shen, J. Gao, X. Huang, M. Wu, Spatial-temporal joint optimization network on covariance manifolds of electroencephalography for fatigue detection, in: 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, IEEE, 2020, pp. 893–900.

[45] M. Wu, R. Ouyang, C. Zhou, Z. Sun, F. Li, P. Li, A study on the combination of functional connection features and Riemannian manifold in EEG emotion recognition, Front. Neurosci. 17 (2024) 1345770.

[46] B.H. Kim, J.W. Choi, H. Lee, S. Jo, A discriminative SPD feature learning approach on Riemannian manifolds for EEG classification, Pattern Recognit. 143 (2023) 109751.

[47] D. Klepl, M. Wu, F. He, Graph neural network-based eeg classification: A survey, IEEE Trans. Neural Syst. Rehabil. Eng. 32 (2024) 493–503.

[48] G. Qu, W. Hu, L. Xiao, J. Wang, Y. Bai, B. Patel, K. Zhang, Y.-P. Wang, Brain functional connectivity analysis via graphical deep learning, IEEE Trans. Biomed. Eng. 69 (5) (2021) 1696–1706.

[49] Y. Yin, X. Zheng, B. Hu, Y. Zhang, X. Cui, EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM, Appl. Soft Comput. 100 (2021) 106954.

[50] O. Tuzel, F. Porikli, P. Meer, Human detection via classification on riemannian manifolds, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007, pp. 1297–1304.

[51] L. Yang, M. Arnaudon, F. Barbaresco, Riemannian median, geometry of covariance matrices and radar target detection, in: The 7th European Radar Conference, IEEE, 2010, pp. 415–418.

[52] T. Wang, C. Li, C. Wu, C. Zhao, J. Sun, H. Peng, X. Hu, B. Hu, A gait assessment framework for depression detection using kinect sensors, IEEE Sens. J. 21 (3) (2020) 3260–3270.

[53] F. Nielsen, R. Bhatia, Matrix Information Geometry, Springer, 2013.

[54] N. Horato, L.A. Quagliato, A.E. Nardi, The relationship between emotional regulation and hemispheric lateralization in depression: a systematic review and a meta-analysis, Transl. Psychiatry 12 (1) (2022) 162.

[55] Z. Chen, T. Xu, X.-J. Wu, R. Wang, Z. Huang, J. Kittler, Riemannian local mechanism for spd neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 37, 2023, pp. 7104–7112.

[56] T. Wang, S. Liu, F. He, M. Du, W. Dai, Y. Ke, D. Ming, Affective body expression recognition framework based on temporal and spatial fusion features, Knowl.-Based Syst. 308 (2025) 112744.

[57] H. Li, Y.-M. Jin, W.-L. Zheng, B.-L. Lu, Cross-subject emotion recognition using deep adaptation networks, in: Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part V 25, Springer, 2018, pp. 403–413.

[58] P. Zhong, D. Wang, C. Miao, EEG-based emotion recognition using regularized graph neural networks, IEEE Trans. Affect. Comput. 13 (3) (2020) 1290–1301.

[59] Y. Li, J. Chen, F. Li, B. Fu, H. Wu, Y. Ji, Y. Zhou, Y. Niu, G. Shi, W. Zheng, GMSS: Graph-based multi-task self-supervised learning for EEG emotion recognition, IEEE Trans. Affect. Comput. 14 (3) (2022) 2512–2525.

[60] W. Guo, G. Xu, Y. Wang, Horizontal and vertical features fusion network based on different brain regions for emotion recognition, Knowl.-Based Syst. 247 (2022) 108819.

[61] G. Xu, W. Guo, Y. Wang, Subject-independent EEG emotion recognition with hybrid spatio-temporal GRU-Conv architecture, Med. Biol. Eng. Comput. 61 (1) (2023) 61–73.

[62] X. Li, C.P. Chen, B. Chen, T. Zhang, Gusa: Graph-based unsupervised subdomain adaptation for cross-subject EEG emotion recognition, IEEE Trans. Affect. Comput. (2024).

[63] W. Guo, Y. Li, M. Liu, R. Ma, Y. Wang, Functional connectivity-enhanced feature-grouped attention network for cross-subject EEG emotion recognition, Knowl.-Based Syst. 283 (2024) 111199.

[64] B. Chen, C.P. Chen, T. Zhang, GDDN: Graph domain disentanglement network for generalizable EEG emotion recognition, IEEE Trans. Affect. Comput. (2024).
[65] J. Shen, J. Chen, Y. Ma, Z. Cao, Y. Zhang, B. Hu, Explainable depression recognition from EEG signals via graph convolutional network, in: 2023 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, IEEE, 2023, pp. 1406–1412.
[66] J. Shen, X. Zhang, X. Huang, M. Wu, J. Gao, D. Lu, Z. Ding, B. Hu, An optimal channel selection for EEG-based depression detection via kernel-target alignment, IEEE J. Biomed. Health Inf. 25 (7) (2020) 2545–2556.
[67] T. Chen, Y. Guo, S. Hao, R. Hong, Exploring self-attention graph pooling with EEG-based topological structure and soft label for depression detection, IEEE Trans. Affect. Comput. 13 (4) (2022) 2106–2118.
[68] G. Tasci, H.W. Loh, P.D. Barua, M. Baygin, B. Tasci, S. Dogan, T. Tuncer, E.E. Palmer, R.-S. Tan, U.R. Acharya, Automated accurate detection of depression using twin pascal's triangles lattice pattern with EEG signals, Knowl.-Based Syst. 260 (2023) 110190.
[69] T. Chen, R. Hong, Y. Guo, S. Hao, B. Hu, MS^2-GNN: Exploring GNN-based multimodal fusion network for depression detection, IEEE Trans. Cybern. 53 (12) (2022) 7749–7759.
[70] W. Liu, K. Jia, Z. Wang, Z. Ma, A depression prediction algorithm based on spatiotemporal feature of EEG signal, Brain Sci. 12 (5) (2022) 630.
[71] Z. Zhang, Q. Meng, L. Jin, H. Wang, H. Hou, A novel EEG-based graph convolution network for depression detection: incorporating secondary subject partitioning and attention mechanism, Expert Syst. Appl. 239 (2024) 122356.
[72] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world'networks, Nature 393 (6684) (1998) 440–442.
[73] E. Bullmore, O. Sporns, Complex brain networks: graph theoretical analysis of structural and functional systems, Nature Rev. Neurosci. 10 (3) (2009) 186–198.
[74] D.S. Bassett, E.T. Bullmore, Small-world brain networks revisited, Neurosci. 23 (5) (2017) 499–516.
[75] M. Ma, Y. Li, Z. Xu, Y. Tang, J. Wang, Small-world network organization of functional connectivity of EEG gamma oscillation during emotion-related processing, in: 2012 5th International Conference on BioMedical Engineering and Informatics, IEEE, 2012, pp. 597–600.
[76] Y. Li, D. Cao, L. Wei, Y. Tang, J. Wang, Abnormal functional connectivity of EEG gamma band in patients with depression during emotional face processing, Clin. Neurophysiol. 126 (11) (2015) 2078–2089.
[77] M. Cirunay, G. Ódor, I. Papp, Scale-free behavior of weight distributions of connectomes, 2024, arXiv preprint arXiv:2407.17220.
[78] T.J. Amos, B. Guragai, Q. Rao, W. Li, Z. Jin, J. Zhang, L. Li, Task functional networks predict individual differences in the speed of emotional facial discrimination, NeuroImage 297 (2024) 120715.
[79] J. Hu, J. Luo, Z. Xu, B. Liao, S. Dong, B. Peng, G. Hou, Spatio-temporal learning and explaining for dynamic functional connectivity analysis: Application to depression, J. Affect. Disord. (2024).
[80] S. Sun, X. Li, J. Zhu, Y. Wang, R. La, X. Zhang, L. Wei, B. Hu, Graph theory analysis of functional connectivity in major depression disorder with high-density resting state EEG data, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (3) (2019) 429–439.
[81] M. Xia, J. Wang, Y. He, BrainNet viewer: a network visualization tool for human brain connectomics, PLoS One 8 (7) (2013) e68910.
[82] A.M. Bastos, J.-M. Schoffelen, A tutorial review of functional connectivity analysis methods and their interpretational pitfalls, Front. Syst. Neurosci. 9 (2016) 175.
[83] M. Ge, R. Mao, E. Cambria, Explainable metaphor identification inspired by conceptual metaphor theory, Proc. AAAI Conf. Artif. Intell. 36 (10) (2022) 10681–10689.
[84] R. Mao, Q. Lin, Q. Liu, G. Mengaldo, E. Cambria, Understanding public perception towards weather disasters through the lens of metaphor, in: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI, International Joint Conferences on Artificial Intelligence Organization, 2024, pp. 7394–7402.
[85] R. Mao, T. Zhang, Q. Liu, A. Hussain, E. Cambria, Unveiling diplomatic narratives: Analyzing United Nations Security Council debates through metaphorical cognition, in: Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci), Vol. 46, 2024, pp. 1709–1716.