

# Decoding Multi-Brain Motor Imagery From EEG Using Coupling Feature Extraction and Few-Shot Learning

Li Zhu<sup>1</sup>, Youyang Liu<sup>2</sup>, Riheng Liu<sup>3</sup>, Yong Peng<sup>4</sup>, *Member, IEEE*, Jianting Cao<sup>5</sup>, *Member, IEEE*, Junhua Li<sup>6</sup>, *Senior Member, IEEE*, and Wanzeng Kong<sup>7</sup>, *Senior Member, IEEE*

**Abstract**—Electroencephalography (EEG)-based motor imagery (MI) is one of brain computer interface (BCI) paradigms, which aims to build a direct communication pathway between human brain and external devices by decoding the brain activities. In a traditional way, MI BCI relies on a single brain, which suffers from the limitations, such as low accuracy and weak stability. To alleviate these limitations, multi-brain BCI has emerged based on the integration of multiple individuals' intelligence. Nevertheless, the existing decoding methods mainly use linear averaging or feature integration learning from multi-brain EEG data, and do not effectively utilize coupling relationship features, resulting in undesired decoding accuracy. To overcome these challenges, we proposed an EEG-based multi-brain MI decoding method, which utilizes coupling feature extraction and few-shot learning to capture coupling relationship features among multi-brains with only limited EEG data. We performed an experiment to collect EEG data from multiple persons who engaged in the same task simultaneously and compared the methods on the collected data. The comparison results showed that our proposed method improved the performance by 14.23% compared to the

single-brain mode in the 10-shot three-class decoding task. It demonstrated the effectiveness of the proposed method and usability of the method in the context of only small amount of EEG data available.

**Index Terms**—Few-shot learning, hypergraph, multi-brain brain computer interface, motor imagery, tensor decomposition.

## I. INTRODUCTION

MOTOR imagery (MI) is one of the classic brain computer interface (BCI) paradigms, which has been widely used in neural rehabilitation. It refers to the utilization of imagined motor movements, rather than actual physical actions [1]. This endogenous spontaneous EEG pattern differs from the evoked EEG activity, since it does not require any external stimulus, but relies solely on the imagined movements generated by the user. The EEG data collected from scalp of users can be used to accurately identify the intended motor tasks. Among various types of EEG signals, MI has received considerable attention and is regarded as a flexible measure of brain activities. As a result, MI-BCI has broad applications in assisting patients with object control and self-care, as well as serving as a tool for rehabilitation and physiotherapy to help patients recover their motor abilities to the fullest extent possible [2], [3]. The high temporal resolution of EEG data is a desirable merit that enables the investigation and diagnosis of various brain disorders or mechanisms [4], [5]. MI is a response to the cognitive task of imagining hand or leg movements and has been widely examined for BCI applications. As such, automated MI classification using machine learning [6], [7] and deep learning techniques [8], [9] has been extensively explored. However, the single-brain MI-BCI paradigm faces several technical challenges, such as low recognition accuracy and weak stability since the decision is made by a single person, and there may be a lack of cross-checking EEG data [10].

The emergence of EEG hyperscanning technology aims to overcome the limitations of traditional single-brain BCI systems. Hyperscanning refers to the technique that synchronously records the brain activity from two or more users who engage in a specific cognitive task. Therefore, the neural mechanisms of social interaction between brains are revealed. Multi-brain BCI can be implemented using

Manuscript received 24 March 2023; revised 11 July 2023 and 4 November 2023; accepted 20 November 2023. Date of publication 23 November 2023; date of current version 30 November 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFE0114900; in part by the National Science Foundation of China under Grant 62301196, Grant 61971173, and Grant U20B2074; in part by the Key Research and Development Project of Zhejiang Province under Grant 2020C04009, Grant 2023C03026, Grant 2021C03001, and Grant 2021C03003; and in part by the Zhejiang Provincial Department of Education Science and Research Project under Grant Y202249732. (*Corresponding author: Yong Peng.*)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee of Saitama Institute of Technology under Approval No. 2018-01.

Li Zhu, Yong Peng, and Wanzeng Kong are with the School of Computer Science and Technology, Hangzhou 310018, China, and also with the Zhejiang Key Laboratory of Brain-Machine Collaborative Intelligence, Hangzhou 310018, China (e-mail: yongpeng@hdu.edu.cn).

Youyang Liu is with the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China.

Riheng Liu is with the HDU-ITMO Joint Institute, Hangzhou Dianzi University, Hangzhou 310018, China.

Jianting Cao is with the Graduate School of Engineering, Saitama Institute of Technology, Fukaya 369-0293, Japan.

Junhua Li is with the School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ Colchester, U.K.

Digital Object Identifier 10.1109/TNSRE.2023.3336356

such techniques. The advent of hyperscanning technology has allowed the study of social interaction neural mechanisms to advance from a “stand-alone paradigm” to a “two-brain/multi-brain neuroscience” by involving multiple individuals and improved the ecological validity beyond observation of single individual [11]. Though different brains might have different underlying mechanisms in encoding the external world, the mental distance among these encoded stimuli is expected to remain constant to a great extent. Therefore, it is rational to presume that inter-brain coupling relationships exist in a representation space of a higher level [12]. This is the reason why multi-brain synergistic BCI can be utilized in multi-brain interaction experiments, such as MI. EEG-based hyperscanning is a straightforward and cost-effective method for recording brain activities, which is widely utilized in multi-brain BCI systems.

The hyperscanning technique is often utilized to study social interaction, including the joint decision and initiator-receiver interactions [13]. In joint decision hyperscanning research, it is indicated that an increasing number of subjects improves the classification performance of multi-brain joint operation hyperscanning systems through the fusion of multi-brain EEG signals [14]. The application of multi-brain BCI in group rehabilitation offers benefits such as peer social support, increased accessibility and opportunity for treatment [15].

To address the challenges mentioned above, we proposed a novel coupling feature-based few-shot learning decoding method for multi-brain MI-BCI. This approach achieved the improved classification accuracy with a limited amount of EEG data. In summary, the present study comprises the following contributions.

- A multi-brain MI experimental paradigm is designed to incorporate idle state detection, with the aim of considering the interactions and coupling features among multi-brain. Therefore, more comprehensive information such as the inter-brain connectivity and synchronization patterns is captured, which was not typically provided in traditional single-brain studies. This could potentially be taken as a new paradigm for rehabilitation.
- The proposed method consists of a feature extraction module for mining cross-brain coupling information, as well as a few-shot learning module to conduct feature classification. Specifically, this method leverages hypergraph learning to extract interpretable representations of brain-to-brain coupling relationships, which can describe higher-order relationships between multiple nodes. Furthermore, it exploits the property that tensor decomposition can extract discriminative features of high-dimensional data. We employ a few-shot learning MI-BCI module to deal with the common problem that only limited amount of data is usually available in MI-BCI field.
- Experiments evaluated the effectiveness of the coupling features as well as the recognizing MI task based on EEG only using a small amount of data, which could facilitate the development of stable MI-BCI with limited training data. In addition, A new evaluation metric, to quantify the impact of inter-brain coupling coordination relationships on the joint classification decision among multi-brain is proposed. The results show that there is a positive

correlation between the cross-brain coupling coordination degree and the classification accuracy.

We structure the remainder of this paper as follows. Section II presents the relevant works especially on the related methods and section III indicates the implementation of our proposed method. In section IV, we describe the multi-brain experimental paradigm. Section V reveals the experimental results. The cross-brain coupling coordination degree is investigated in section VI, followed by the conclusion in section VII.

## II. RELATED WORKS

In what follows, we present a description of the relevant techniques that are involved in our proposed method. Specifically, these techniques include the hypergraph learning, tensor decomposition, and few-shot learning.

### A. Hypergraph Learning

Hypergraph structures are extensively employed in various fields for modeling higher-order data correlations. It not only preserves valuable information, but also models relationships among multiple nodes. This idea was first introduced in [16], which proposed a propagation process over hypergraph structures. The goal of transition inference on hypergraphs is to minimize the discrepancy of labels among strongly connected vertices. Zhou et al. employed the hypergraph theory to partition EEG samples into a predetermined number of clusters, where each cluster corresponds to an emotional category [17]. Vertices within the same cluster share similar emotional characteristics. In [18], Gao et al. proposed a novel seizure detection approach by integrating hypergraph features with machine learning.

### B. Tensor Decomposition

Similar to the matrix factorization, tensor decomposition can be applied to tensor-structured data, which inherently takes advantage of the interactions between multiple modes of the tensor [19]. EEG signals are commonly represented as matrices, and analyzed using methods like time series and spectral analysis, as well as matrix decomposition. Typically, EEG signals exhibit more than the two temporal and spatial patterns and necessitate the tensors representations [20]. Recently, tensor decomposition has proven highly effective in extracting and analyzing features of EEG signals, leading to remarkable outcomes [21], [22]. In this paper, we primarily employ the Tucker decomposition to extract the core feature tensor of a hypergraph generated by EEG data from two subjects.

### C. Few-Shot Learning

Few-shot learning is another important machine learning paradigm, it achieves fast adaptation and transfer abilities of the model by utilizing a limited amount of data to learn new target categories. For example, in [23], a universal few-shot learning framework was proposed, wherein the classifier is demanded to discern novel classes that are absent in the training set, and only sparse examples are accessible for each emerging class. In [24], the authors proposed to

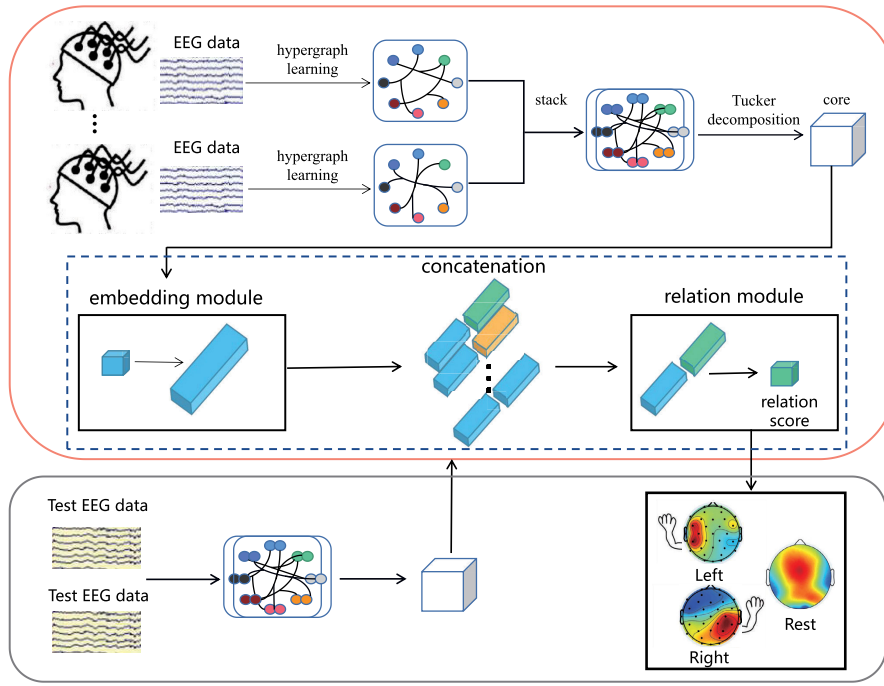


Fig. 1. The overall framework of our proposed method applied to EEG multi-brain MI scenario of three-classification task.

compute the distance between the input data and the prototype representation of each class in the metric space, and subsequently performed classification based on the proximity to each prototype. In [25], a Siamese neural network was proposed for single-shot image recognition, which achieved comparable performance to human recognition abilities. Few-shot learning methods can be applied to EEG signal decoding. In [26], an end-to-end trainable learning paradigm, MLCL, was proposed for decoding emotion recognition from EEG signals. Reference [27] proposed a meta-learning strategy to search for optimal parameters for BCI decoders, which resulted in an increased motor imagery classification accuracy for EEG-based MI-BCI decoding. A novel two-way few-shot network was designed in [28], which was capable of effectively learning representative features for unseen target categories and classifying them with limited MI EEG data.

### III. METHOD

This paper proposes a coupling feature-based few-shot learning model for robust EEG decoding in MI-BCI, given only limited data. In Fig. 1, we present the overall framework of our proposed method consisting of a coupling feature extraction module and a few-shot learning module, which will be applied to the three-class decoding task in multi-brain MI. In the first module, we employ the hypergraph learning and Tucker decomposition to extract the coupling relationship features among multi-brain in performing the same MI task. To handle limited sample size, we adopt the relation network as a few-shot learning module for the subsequent three-class decoding task.

#### A. Hypergraph Learning-Based Coupling Feature Extraction

Specifically, hypergraphs are defined on a finite set  $\mathbf{V}$  as a generalization of graphs to describe high-order relationships

among multiple vertices. Given their capacity to connect any number of vertices, hypergraphs enable direct extraction of complex relationships among vertices connected by an edge, and even between different edges. The following overview introduces the fundamental principles of hypergraphs.

A hypergraph is defined as  $\mathbf{G}(\mathbf{V}, \varepsilon, \mathbf{W})$ , where  $\mathbf{V}$  is a set of vertices,  $\varepsilon$  is a set of hyperedges, and  $\mathbf{W}$  is a diagonal matrix representing edge weights. Hypergraph can be represented by an  $|\mathbf{V}| \times |\varepsilon|$  association matrix  $\mathbf{H}$ , which is defined as

$$\mathbf{H} = h(v, e) = \begin{cases} 1, & \text{if } v \in e; \\ 0, & \text{if } v \notin e. \end{cases} \quad (1)$$

In the hypergraph, the degree of each vertex is defined as  $d(v) = \sum_{e \in \varepsilon} w(e)h(v, e)$ , which is the sum of weights of all edges that contain the vertex. The degree of each edge is defined as  $\delta(e) = \sum_{v \in \mathbf{V}} h(v, e)$ , which is the number of vertices connected by the edge. Moreover, the degrees of a hypergraph can be conveniently represented by means of diagonal matrices. To be specific,  $\mathbf{D}_v$  and  $\mathbf{D}_e$  are diagonal matrices whose diagonal elements correspond to the degrees of vertices and edges, respectively.

Let  $\mathbf{X} = [x_1, x_2, \dots, x_N]^T \in \mathbb{R}^{N \times P}$  represents the EEG data collected from a subject, where  $N$  denotes the number of channels,  $x_i \in \mathbb{R}^P$  represents the time series of the  $i$ -th channel, and  $P$  is the total number of time points captured in the EEG acquisition process. Assume that  $\mathbf{X}$  in this study has been subject to a preprocessing step that involves normalization, resulting that each row has a zero mean and a unit Euclidean norm. In our work, each channel is regarded as a vertex in the hypergraph. We construct a hypergraph to model the subject's functional connectivity network (FCN), which is used to capture high-order interaction features, such as coupling relationships, among multiple channels. The specific

approach employed to obtain the hypergraph matrix  $\mathbf{H}$  is described as follows.

By drawing inspiration from recent studies [29], [30], [31], we utilize the sparse coding model (also named Lasso regression in statistics), to construct  $\mathbf{H}$ . Specifically, the time series of each EEG channel is regarded as a response vector that can be estimated by a linear combination of the time series formed by the other  $N - 1$  channels, which generates a representation coefficient vector. That is, for  $x_i$ ,  $i|_{i=1}^N$ , if the coefficient vector is  $\alpha_i$ , we have

$$\min_{\alpha_i} \frac{1}{2} \|x_i - \mathbf{A}_i \alpha_i\|_2^2 + \lambda \|\alpha_i\|_1, \quad (2)$$

where  $\mathbf{A}_i = [x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N] \in \mathbb{R}^{P \times (N-1)}$  denotes the collection of time series of all channels but the  $i$ -th EEG channel. The vector  $\alpha_i \in \mathbb{R}^{N-1}$  quantifies the coefficients associated with the coupling relationship between the  $i$ -th channel and the remaining  $N - 1$  channels. The regularization parameter  $\lambda > 0$  controls the sparsity of vector  $\alpha_i$ , which is usually selected from  $\{0.01, 0.02, \dots, 0.1\}$ .

Through selecting an appropriate value for  $\lambda$  and then optimizing the sparse learning model in equation (2), a hyperedge  $e_i$  can be obtained. This hyperedge  $e_i$  is comprised of a central channel (*i.e.*, the  $i$ -th channel selected each time) and all the other channels in the coefficient vector  $\alpha_i$  that have corresponding positive elements, indicating that they are coupled with the central channel on one hyperedge. It is worth noting that the hyperedge  $e_i$  we constructed excludes the channels with corresponding negative and zero elements in  $\alpha_i$ . This decision is made because such channels have either adverse or negligible effects on the central channel.

We assume that the coupling relationships among channels should exhibit non-negative correlations. Accordingly, we can obtain the coupling relationships (interactions) between the selected central channel and other channels within the same hyperedge, thereby filtering out any insignificant or spurious connections. The methodology employed in this study enables the effective representation of coupling relationships among several channels by leveraging the local information within each hyperedge in the hypergraph. After optimizing the  $N$  sparse coding models, a hypergraph with  $N$  hyperedges is ultimately acquired. It is noteworthy that the sparsity or density of the obtained adjacency matrix  $\mathbf{H} \in \mathbb{R}^{N \times N}$  of the hypergraph is dependent on the increment or decrement of the parameter  $\lambda$ .

Based on the above analysis, hypergraph learning automatically adjusts of the influence of different hyperedges by learning their respective weight values, consequently reducing redundancy within the hyperedges and generating distinctive FCNs. Obviously, when  $\lambda$  is large enough in objective function (2), the corresponding  $\alpha_i$  will be pretty sparse and contain few non-zero elements. In the extreme case, the hyperedge  $e_i$  only comprises the central channel, which is not the intended outcome. Therefore, our hypergraph learning approach, as applied in this paper, excludes such hyperedges.

### B. Tucker Decomposition-Based Feature Compression

Considering two matrices  $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$  and  $\mathbf{B} \in \mathbb{R}^{I_2 \times I_3}$ ,  $\mathbf{A} \times \mathbf{B}$  can be interpreted as a linear transformation that is applied to matrix  $\mathbf{A}$ . Then, the resultant matrix is  $(\mathbf{A} \times \mathbf{B}) \in \mathbb{R}^{I_1 \times I_3}$ .

Since tensors are multi-dimensional arrays with three or more dimensions, we can consider them as high-order matrices. Then, the matrix transformation method can be extended for tensor-represented data, as illustrated below.

In this section, we will delve into the mode- $n$  product of tensors. Specifically, let  $\chi \in \mathbb{R}^{I_1 \times I_2 \times I_3}$  and  $\mathbf{A} \in \mathbb{R}^{J_1 \times I_1}$  be a tensor and a matrix, respectively, such that  $\chi \times_1 \mathbf{A}$  is the desired operation. It follows that the result of the mode-1 product is an  $J_1 \times I_2 \times I_3$  dimensional tensor. We can interpret the mode-1 product as a linear transformation along the first mode of tensor  $\chi$ , mapping the first mode from  $I_1$  to  $J_1$ . Notably, if  $J_1 < I_1$ , this operation corresponds to a dimensionality reduction along the first mode of the tensor  $\chi$ . For a given tensor  $\chi \in \mathbb{R}^{D \times E \times F}$ , its Tucker decomposition can be expressed by

$$\begin{aligned} \chi &\approx \mathbf{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} a_p \circ b_q \circ c_r \\ &= [\mathbf{G}; \mathbf{A}, \mathbf{B}, \mathbf{C}], \end{aligned} \quad (3)$$

where  $\mathbf{G} \in \mathbb{R}^{P \times Q \times R}$  is a core tensor,  $\mathbf{A} \in \mathbb{R}^{D \times P}$ ,  $\mathbf{B} \in \mathbb{R}^{E \times Q}$ , and  $\mathbf{C} \in \mathbb{R}^{F \times R}$  are factor matrices. Mathematically, Tucker decomposition tries to identify a minimal core tensor and a set of matrices that can approximate the original tensor  $\chi$  through their product. The core tensor captures the interaction among components in each dimension, thereby integrating multidimensional information. These factor matrices are also known as factor arrays and generally exhibit orthogonality. Additionally, the factor matrices can capture the dominant features across each dimension of the original tensor. Tucker decomposition is a commonly used method for reducing the dimensionality of high-dimensional data and extracting its features. This approach helps to understand the intrinsic structure and patterns of the data, while also providing an effective means for data compression and representation.

Below, we utilize the core tensor derived from the Tucker decomposition for EEG decoding task. The element-wise representation of tensor  $\chi$  is

$$\begin{aligned} x_{def} &\approx \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} a_{dp} b_{eq} c_{fr}, \\ d &= 1, \dots, D, e = 1, \dots, E, f = 1, \dots, F. \end{aligned} \quad (4)$$

If  $P, Q, R$  are respectively less than  $D, E, F$ , the core tensor can serve as the compressed form of the original tensor, which achieves dimensionality reduction. Equivalently, the Tucker decomposition constitutes a high-order implementation of the principal component analysis technique for dimensionality reduction.

Following the application of hypergraph learning on the EEG data, two hypergraphs were obtained that depict the intricate brain coupling relationships of paired-group subjects during performing cognitive tasks. Both hypergraphs are concatenated to produce a three-dimensional tensor, which is subjected to Tucker decomposition across three dimensions, *i.e.*, subject, channel and channel. As a result, the core tensor is extracted to capture the coupling relationships between the two subjects, which will be employed in the subsequent classification module.



### C. Relation Network-Based Few-Shot Learning

The Relation Network is comprised of two modules, *i.e.*, the embedding module and the relation module. To be precise, the embedding module is trained by the core tensors obtained by hypergraph learning and Tucker decomposition, whereas the relation module is trained on the merged feature map of the train and the test core tensors. In particular, the training set is  $S = \{(x_i, y_i)\}_{i=1}^m$  ( $m = K \times C$ ), where  $K$  denotes the number of labeled samples and  $C$  represents the number of categories. Meanwhile, the test set  $T = \{(x_j, y_j)\}_{j=1}^n$  ( $n = R \times C$ ) has  $R$  samples, which is obtained by subtracting  $K$  from the total number of samples. Subsequently, the relation score  $r_{i,j}$  (*i.e.*, a scalar value between 0 and 1) is generated by the network to reflect the similarity between the train and test core tensors. Finally, the category with the highest relation score  $r_{i,j}$  is selected as the ultimate classification outcome.

The selected loss function for the model is the mean square error loss

$$\operatorname{argmin} \sum_{i=1}^m \sum_{j=1}^n (r_{i,j} - 1(y_i == y_j))^2, \quad (5)$$

which performs regression of the relation score  $r_{i,j}$  against the ground truth. Specifically, matched pairs have a similarity value 1, while mismatched pairs have a similarity value 0.

The embedding module consists of four convolutional blocks, each among which comprises a 64-filter  $3 \times 3$  convolution, a batch normalization layer, and a rectified linear unit (ReLU) nonlinearity layer. To enable additional convolution of the output feature maps in the relation module, the first two convolutional blocks in the embedding module include an extra  $2 \times 2$  max-pooling layer.

The structure of the relation module consists of two convolutional blocks and two fully-connected layers. Each convolutional block employs a 64-filter  $3 \times 3$  convolution, followed by a batch normalization, ReLU non-linear activation, and  $2 \times 2$  max-pooling. In consideration of the training and inference time const for the deep model, the fully-connected layers have been set to 8 and 1 dimensions, respectively, and both of them utilize ReLU activation functions. To make the relation scores  $r_{i,j}$  be suitable for classification, the output layer utilizes a sigmoid activation function.

## IV. EXPERIMENTS

This section describes the multi-brain MI paradigm for EEG data acquisition and the setup for the EEG decoding method.

### A. The Multi-Brain MI Experiment Design

16 healthy participants were recruited and grouped into eight pairs for the purpose of EEG data collection. All participants were inexperienced with the BCI system, and had received detailed instructions regarding the experimental protocol prior to the commencement of this study. To optimize the experimental experience for participants, adjustments were made to the distance between the chair and LCD monitor based on their feedback, as illustrated in Fig. 2(a). In the experiment, two participants performed the MI task simultaneously and two connected Neuroscan amplifier were used to record the EEG signals according to international 10-20 system. The sampling rate of the signal is 1000 Hz. It was down-sampled

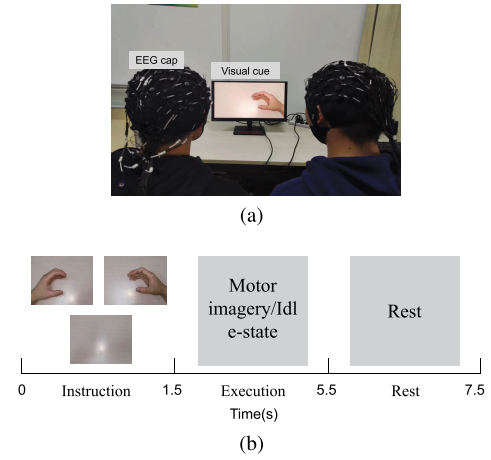


Fig. 2. Multi-brain MI experimental environment and design. (a) Environmental configuration of the multi-brain MI EEG data acquisition. (b) Experiment protocol in each trial including the instruction for 1.5 s, the MI execution for 4 s, and resting state for 2 s.

to 100 Hz in this study. Signal impedance was maintained at or below  $15K\Omega$  throughout the experiment.

The purpose of this multi-brain MI experiment was to decipher distinct user intentions within a three-class MI task, including left hand, right hand motor imagery and idle state. There were five sessions in the experiment and each session consisted of 75 trials. Each trial lasted 7.5s, comprising video clip cue for 1.5 seconds, the MI task for 4s, and the rest duration for 2s. Fig. 2(b) showed the protocol of our experiment. A short training phase was used to participants to familiarize them with the experiment before the formal sessions.

### B. Model Setup of the Few-Shot Learning Framework

Initially, we present the  $C$ -way  $K$ -shot problem in the domain of few-shot learning. In this scenario, assuming that  $K$  instances per class are arbitrarily picked from the dataset as the training set, and the remaining instances belonging to each class constitute the testing set. The intended few-shot problem is dubbed as  $C$ -way  $K$ -shot.

In the experimental setup, we implemented leave-one-subject-out cross-validation, where among 8 groups of participants, signals from 7 groups of subjects were used as training dataset, and the rest was used for testing. During the training process,  $K$  samples are randomly selected for training, while all remaining samples were used for validation.

To assess the effectiveness of our proposed approach comprehensively, we validated the approach on situations with different numbers of few shots (*i.e.*,  $K = \{1, 5, 10, 15\}$ ). Given that EEG is non-stationary and its statistical characteristics fluctuate across trials and over time [32], [33], the EEG decoding performance may vary due to the different training sets. Therefore, we carried out the identical experiment ten times and obtained the average accuracies for performance assessment. Especially, analysis of variance (ANOVA) was performed to evaluate whether or not there were significant differences between dual-brain and single-brain modes. The experimental data were filtered to 8-30Hz. In addition, we also explored the performance of the approach in the different

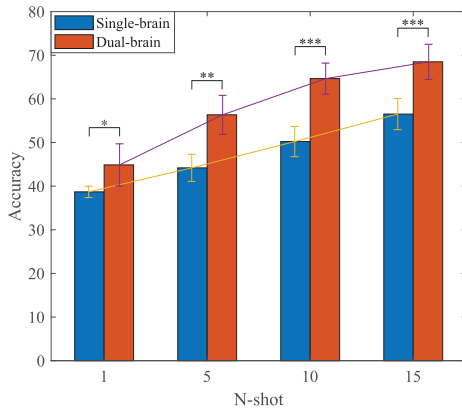


Fig. 3. Accuracy comparisons between the single-brain mode and the dual-brain mode for the three-class MI decoding task. Asterisks represent statistical significance levels (\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ).

frequency ranges and different lengths of data segments and compared the performance between these cases.

## V. RESULTS

In this section, we show the results of model evaluations and performance comparisons. Specifically, the error-bar value used in Fig.3-Fig.7 is the standard deviation.

### A. Performance of Single-Brain and Dual-Brain Modes

We first compare the experimental performance between the multi-brain MI paradigm and the single-brain. Specifically, the multi-brain paradigm used here is the dual-brain paradigm. The Fig. 3 shows that the average classification accuracy of the multi-brain mode is nearly 10% higher than that of the single-brain mode in the 1-shot case. More prominently, around 15% performance improvement was achieved by the multi-brain mode in the 5-shot, the 10-shot, and 15-shot cases. As shown in Fig. 4, the dual-brain outperforms the single-brain for all paired groups in all cases of 1-shot, 5-shot, 10-shot and 15-shot. The dual brain advantages of the fifth, seventh and eighth groups gradually emerged as the number of shots increased, with the improved accuracy of dual-brain than single-brain modes ranging from 1.57% to 15.99%. This obviously demonstrates the capability of capturing the inter-brain coupling relationships in the multi-brain mode, which is beneficial for improving the EEG decoding accuracy and leads to superior performance compared to the popular single-brain mode in motor imagery. The underlying rationale is also intuitive. The multi-brain MI involves brain-to-brain coupling relationships and enhances traditional MI by promoting the mutual recognition, collaboration, and synchronization between users. As a result, our study demonstrates that multi-brain MI has greater classification accuracy than single-brain mode.

### B. Comparison to the State-of-the-Art Methods

We conduct experiments by comparing our proposed EEG decoding method with state-of-the-art methods, including the EEG Convolutional Neural Network (EEGNet) [8], the Filter Bank Common Spatial Patterns (FBCSP) [7], the Graph Convolutional Network (GCN) [34], the Prototypical

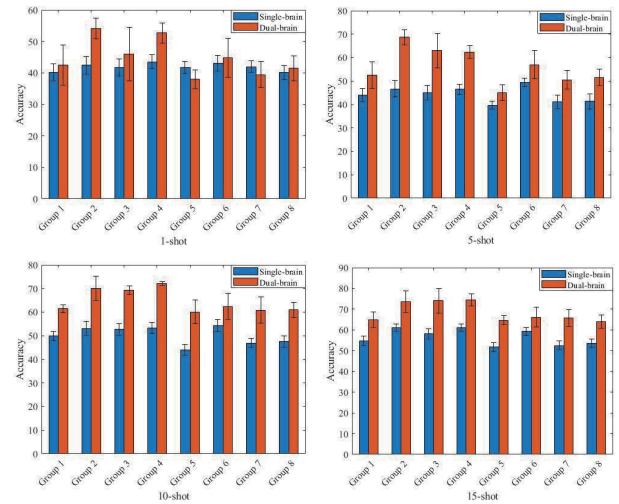


Fig. 4. Accuracy comparisons for each group between the single-brain and the dual-brain modes.

TABLE I  
PERFORMANCE COMPARISON AMONG METHODS  
IN THE DIFFERENT CASES OF SHOTS (%)

	1-shot	5-shot	10-shot	15-shot
EEGNet	32.43±1.21	35.21±1.45	37.50±0.74	39.63±1.39
FBCSP	34.18±1.89	39.23±1.98	43.55±3.39	49.02±3.39
GCN	35.00±0.61	40.26±1.62	46.10±3.72	53.69±2.57
Proto. Nets	38.49±3.66	46.22±3.98	51.33±3.40	59.37±3.31
Siamese Nets	39.67±3.13	47.17±2.61	52.26±3.37	60.07±3.33
MAML-BCI	42.23±1.33	51.53±2.11	60.77±3.01	63.13±2.17
DA-RelationNet	42.17±3.37	52.43±3.73	61.67±2.67	64.11±1.71
Single-brain	38.69±1.31	44.19±3.11	50.22±3.47	56.52±3.57
Non-hyper.	41.84±1.16	49.77±2.17	54.88±3.43	60.37±3.55
Non-tucker	42.08±3.20	51.56±4.04	58.08±4.45	62.57±4.43
Ours	<b>44.87±4.82</b>	<b>56.35±4.47</b>	<b>64.45±3.57</b>	<b>68.49±4.02</b>

Networks (Proto. Nets) [24], the Siamese Neural Network (Siamese Nets) [25], the Model-Agnostic Meta-Learning for EEG Motor Imagery Decoding in Brain-Computer-Interfacing (MAML-BCI) [27] and the two-way few-shot network (DA-RelationNet) [28].

Generally, EEGNet is a compact convolutional neural network model, which was originally proposed for MI EEG decoding. FBCSP is a common spatial pattern-based spatial filtering technique, which is implemented by feature selection combined with frequency band segmentation. GCN basically incorporates graphs into convolutional neural network to better model the data correlations, which was originally proposed for solving semi-supervised learning. Prototypical Networks and Siamese Neural Network are few-shot learning methods. MAML-BCI is a meta-learning deep architecture with three processing stages, designed to optimize the parameters of BCI decoders so that they can quickly generalize to different subjects. DA-RelationNet is a dual attention relation network that can generalize on unseen subjects by using few-shot learning and a FT strategy for EEG-based MI classification.

The parameters of the above compared methods were set according to the original papers.

Table I presents the results for the three-class MI classification task (*i.e.*, left hand, right hand, and idle state). EEGNet exhibited the lowest performance in small-sample tasks. For

TABLE II  
THREE-CLASS MI TASK DECODING ACCURACY(%) IN  
TERMS OF DIFFERENT FREQUENCY BANDS

	1-shot	5-shot	10-shot	15-shot
$\alpha$ - $\alpha$	43.00 $\pm$ 4.05	52.47 $\pm$ 2.77	62.05 $\pm$ 2.21	67.15 $\pm$ 2.91
$\alpha$ - $\beta$	42.26 $\pm$ 3.32	52.32 $\pm$ 2.15	62.01 $\pm$ 2.46	66.74 $\pm$ 2.76
$\beta$ - $\alpha$	43.21 $\pm$ 2.86	52.51 $\pm$ 1.92	61.67 $\pm$ 1.83	66.61 $\pm$ 2.11
$\beta$ - $\beta$	43.35 $\pm$ 1.93	51.43 $\pm$ 1.91	59.69 $\pm$ 1.62	65.33 $\pm$ 1.98

example, it achieved an accuracy of about 40% only when  $K = 15$ . When given the same number of training samples, FBCSP and GCN achieved classification accuracies of 49.02% and 53.69%, respectively. Proto. Nets and Siamese Nets Proto. Nets exhibited similar performance in terms of classification accuracy, achieving approximately 60% accuracy when  $K = 15$ . MAML-BCI and DA-RelationNet achieved accuracy rates of 63.13% and 64.11%, respectively, with  $K = 15$ . In contrast, our proposed method attained a significantly higher decoding accuracy of 68.49%, improving the performance by around 5%.

### C. The Impacts of Frequency Bands and Time Windows

This section performs experiments with EEG data corresponding to different frequency bands and time windows, to identify the best one. Since the motor imagery task mainly correlated to the  $\alpha$  (8-12 Hz) and  $\beta$  (13-30 Hz) frequency bands, We tested the classification accuracies based on the four possible configurations of these two bands, and showed the results in Table II. From these results, we noticed that the differences among the classification accuracies of the four configurations were negligible when  $K$  equals to 1 and 5. However, as  $K$  increased, the  $\alpha$  frequency data for both participants resulted in the highest classification accuracy. This might be due to that the MI features primarily lie within the  $\alpha$  frequency band.

Besides the different frequency bands, we tested the experimental data with different time window periods, that is, we fix the entire frequency band and then respectively divide the MI task time windows (*i.e.*, the 4 seconds from 1.5s to 5.5s) into four segments, *i.e.*, [0,1], [1,2], [2,3], and [3,4]. The classification accuracy results corresponding to these different time segments are presented in Table III. Obviously, the highest accuracy was achieved when the time window is set as [1,2], *i.e.*, the time window from the first second to the second second after the onset of MI and its accuracies corresponding to different numbers of training samples are shown in the second row. These findings indicate that there might be a startup time required for the synchronization between the two participants, which occurs after the second second of the execution.

Furthermore, in order to explore the proper duration for motor imagery task analysis, we conducted the experiment on the decoding accuracy between data with 0-1s, 0-2s, 0-3s, and 0-4s time window cases. Results shown in the Table IV, it can be observed that the 0-2s exhibits the better decoding accuracy through different number of few-shots, which indicates that the proper duration for the MI task lies within the first two seconds.

TABLE III  
THREE-CLASSIFICATION MI TASK DECODING ACCURACY(%) IN TERMS  
OF DIFFERENT TIME WINDOWS(S)

	1-shot	5-shot	10-shot	15-shot
0-1	45.09 $\pm$ 4.07	55.19 $\pm$ 3.03	62.34 $\pm$ 3.40	68.19 $\pm$ 2.89
1-2	46.28 $\pm$ 3.71	56.40 $\pm$ 3.17	64.75 $\pm$ 2.84	69.44 $\pm$ 3.36
2-3	45.66 $\pm$ 3.64	55.32 $\pm$ 3.42	62.34 $\pm$ 3.62	66.94 $\pm$ 3.68
3-4	44.02 $\pm$ 4.05	53.87 $\pm$ 3.57	61.12 $\pm$ 3.88	66.50 $\pm$ 3.54

TABLE IV  
THREE-CLASSIFICATION MI TASK DECODING ACCURACY(%) IN TERMS  
OF DIFFERENT LENGTH OF TASK DURATIONS(S)

	1-shot	5-shot	10-shot	15-shot
0-1	45.09 $\pm$ 4.07	55.19 $\pm$ 3.03	62.34 $\pm$ 3.40	68.19 $\pm$ 2.89
0-2	46.87 $\pm$ 3.68	56.88 $\pm$ 3.53	64.90 $\pm$ 3.51	69.65 $\pm$ 3.93
0-3	46.13 $\pm$ 3.81	56.36 $\pm$ 3.89	64.52 $\pm$ 3.64	68.90 $\pm$ 4.28
0-4	44.87 $\pm$ 4.82	56.35 $\pm$ 4.47	64.45 $\pm$ 3.57	68.49 $\pm$ 4.02

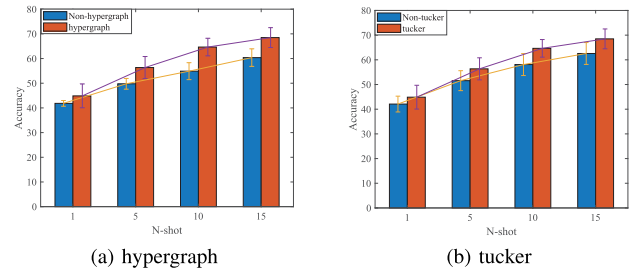


Fig. 5. Investigation on the impacts of hypergraph learning and Tucker-based tensor decomposition.

### D. Ablation Experiments

In this section, we verify the effectiveness of each component of our proposed EEG decoding method by ablation experiments. We evaluate its performance by reserving or deleting a sub-module. Fig. 5(a) shows us the classification accuracies with and without the hypergraph learning, and in Fig. 5(b), a comparison of recognition accuracies with and without the Tucker decomposition is provided. From both experiments, we can see that there are only slight differences between using and not using a certain technique, in the case of  $K = 1$ . Nevertheless, as the sample size increases, our proposed method significantly improves the decoding performance by increasing 5-10%.

Furthermore, we have performed ablation experiments to demonstrate the effectiveness of each component in our proposed methods, *i.e.*, the hypergraph learning and the Tucker decomposition. It is obvious that the combination of both components leads to performance improvements of 8.12% (*i.e.*, 68.49% vs. 60.37%) and 5.92% (*i.e.*, 68.49% vs. 62.57%), respectively.

To more intuitively show the results on each paired group, we provide the ablation experimental results in Fig. 6 and Fig. 7, which demonstrate that our proposed method was consistently effective for all the eight groups. In more detail, the results of the second, third and fourth groups are superior to those of the remaining groups. As shown in Fig. 6, the performance of these three groups increased by

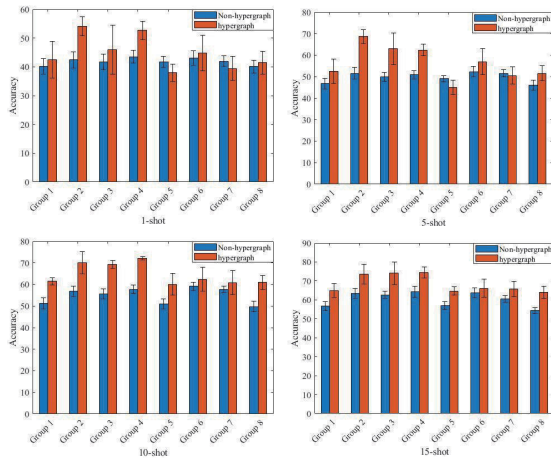


Fig. 6. Investigation on the impacts of hypergraph learning in each of the eight groups.

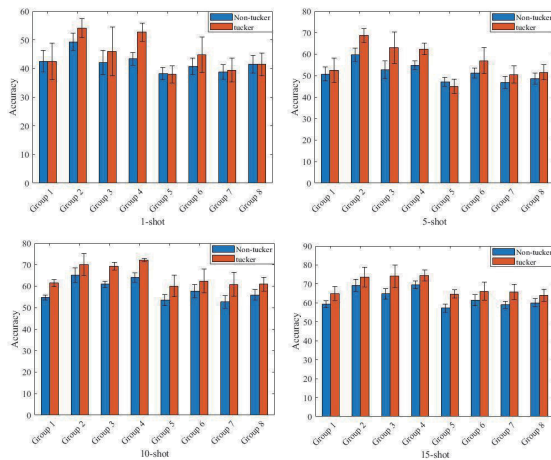


Fig. 7. Investigation on the impacts of Tucker decomposition-based feature compression in each of the eight groups.

15 percentage points with the help of the hypergraph learning technique. Additionally, when  $K = 10$ , the improvement was almost 20%. Similarly, in the Tucker decomposition ablation experiment, we have around 10% performance improvements in the best three groups. Furthermore, the hypergraph learning in our proposed method brings more performance improvements than the Tucker decomposition. This coincides our understanding to the underlying decision-making mechanism in multi-brain MI. That is, because hypergraph learning mainly exploits the coupling relationship between two subjects during synchronous experiments, it should play a more important role than Tucker decomposition, which is mainly responsible for feature compression.

To provide more insights into the learned hypergraph between subjects, in Fig. 8, we visualize the inter-brain and intra-brain hypergraph connections of subjects during motor imagery and idle states. Notably, a greater number of inter-brain connections were observed during the task. In contrast, the idle state is characterized by almost non-existent inter-brain connections and has primarily the intra-brain connections. The more connections mean that there are more information flow and synchronization. During the idle state trials, the paired participants can imagine any different

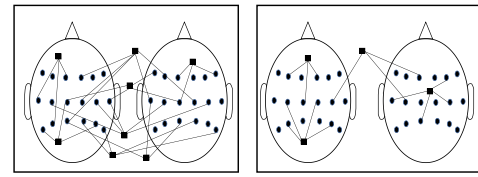


Fig. 8. The visualization of brain hypergraph connections during task (left subfigure) and idle (right subfigure) states.

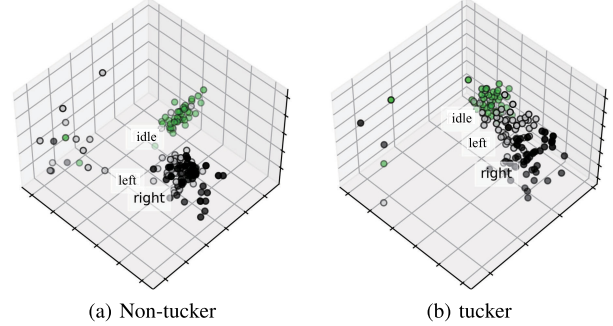


Fig. 9. Illustration to the impact of Tucker decomposition by low-dimensional data visualization.

motor imagery task independent content therefore the synchronization becomes lower. This disparity allows for better differentiating the task and idle states, leading to improved decoding performance. This figure intuitively explains why the hypergraph learning techniques can bring higher classification accuracy in multi-brain MI decoding.

In Fig. 9, we visualize the low-dimensional data representations with and without the Tucker decomposition. In Fig. 9(a), where Tucker decomposition techniques were not utilized, it is evident that there was inadequate differentiation between the features of the left and right hands. Nonetheless, in Fig. 9(b), the features respectively corresponding to the left and right hands became more separable after utilizing the Tucker decomposition techniques.

## VI. DISCUSSION

In this section, we first introduce a quantitative metric for measuring the cross-brain coupling coordination degree, which is then evaluated on our collected multi-brain MI EEG data.

### A. Cross-Brain Coupling Coordination Degree

To further investigate the impact on decoding accuracy, we adopt a new metric in this paper inspired from the field of economics, the cross-brain Coupling Coordination Degree (CCD). This metric serves as an analytical tool for assessing the level of coordinated development of phenomena.

Through a scholarly inquiry into the definitions and models of coupling in other disciplines such as economics and engineering, we have the following definition of the coupling degree between two brains:

$$C_i = 2 \left\{ \frac{u_i(1)u_i(2)}{(u_i(1) + u_i(2))^2} \right\}^{1/2}, \quad (6)$$

where  $u_i(1)$ ,  $u_i(2)$  are the local efficiency  $E_l$  of brain 1 and 2 at the  $i$ -th moment, respectively.



Inspired by the graph theory, the shortest path length is employed to determine the most efficient route between two vertices. However, in cases where two nodes in the network are not connected, the calculation yields an infinite shortest path length, which lacks practical significance. To mitigate this issue, the concepts of global efficiency  $E_g$  and local efficiency  $E_l$  are introduced in graph theory, *i.e.*,

$$E_g = \frac{1}{N(N-1)} \sum_{i \neq j=1}^N \frac{1}{l_{ij}}, \quad (7)$$

$$u_i = E_l = \frac{1}{N} \sum_{i=1}^N E_g(i), \quad (8)$$

where  $N$  is the total number of vertices and  $l_{ij}$  is the shortest path length between vertices  $i$  and  $j$ . The global efficiency  $E_g$  demonstrates an inverse relationship with the shortest path length. As the shortest path length decreases, the global efficiency increases, promoting faster information transmission between vertices. Local efficiency  $E_l$  is an average of all global efficiency values of nodes and serves as a parameter evaluating the information transmission abilities among vertices in localized network regions. Both global efficiency  $E_g$  and local efficiency  $E_l$  are bounded between 0 and 1, where 0 indicates no connectivity and 1 signifies optimal connectivity.

The  $T_i$  is known as the comprehensive coordination variable, which can be calculated as

$$T_i = a \times u_i(1) + b \times u_i(2), \quad (9)$$

where  $a$ ,  $b$  are two coefficients to be determined. Due to the equal significance granted to both brains in multi-brain MI, both  $a$  and  $b$  are assigned 0.5.

To comprehensively present the coupling degree, intra and inter the single-brain 1 and the single-brain 2, this study integrates the coupling degree  $C$  with the coordination degree  $T$ . Specifically, for the  $i$ -th moment, we have the definition of CCD metric  $D_i$  as

$$D_i = \sqrt{C_i \times T_i}, \quad (10)$$

Similarly,  $D_i$  also has the value range of [0,1], where a larger value indicates a higher level of internal single brain and between the two brains coupling coordination, whereas a smaller value indicates weaker coupling coordination.

Consequently, CCD objectively reflects the level of coordination development intra and inter each single brain in the paired group, effectively avoiding abnormal situations where the information transmission efficiency level is low, but their coordination is high. The cross-brain coupling coordination degree model is characterized by simplicity, comprehensive, operability, and visual analysis.

### B. Evaluation of the CCD Metric

Table V provides an analysis to the cross-brain CCD metric across the eight groups in our multi-brain MI experiments. The results demonstrate that the second, third and fourth groups (*i.e.*, the bold numbers) display the most effective cross-brain coupling coordination degree among the eight groups. This outcome is in line with our findings in section V-D, indicating that the information flow for intra single brain as well as the inter paired brains have positive impact on the decoding accuracy.

TABLE V  
CROSS-BRAIN CCD VALUES OF THE EIGHT GROUPS IN OUR MULTI-BRAIN MI EXPERIMENTS

	1	2	3	4	5	6	7	8
$C$	0.5883	<b>0.7126</b>	<b>0.6678</b>	<b>0.6358</b>	0.6113	0.5973	0.7809	0.5656
$T$	0.5883	<b>0.7126</b>	<b>0.6678</b>	<b>0.6358</b>	0.6113	0.5973	0.7809	0.5656
$D$	0.7802	<b>0.8368</b>	<b>0.8217</b>	<b>0.8131</b>	0.7946	0.7782	0.7801	0.7692

### C. Performance of Multi-Brain Brain Computer Interface From the View of Cognitive Neuroscience

Brain computer interface research has made tremendous progress in recent years. However, it remains challenging to transfer its results from the lab to the marketplace. Herein, one of the big challenges is underdeveloped paradigm which is invented about 30 years ago such as motor imagery [35]. To break the shortcomings of traditional BCI such as low stability, poor performance, researchers also try to propose a new scheme as hybrid BCI (hBCI), which is implemented by combining two or more kinds of EEG signals (*i.e.* motor imagery and P300), another combining EEG and other signals (*i.e.* Electromyography (EMG) signal) [36]. In this paper, we propose a new multi-brain experimental design and the corresponding EEG decoding method for motor imagery. For such multi-brain computer interface, it can be taken as a kind of hBCI paradigm, which combines the various kind of EEG signals as well as the interactions from multi-brains.

Our study demonstrated that the better classification performance was obtained when the dual-brain mode was compared to the single-brain mode. Previous studies on social interactions based on hyperscanning reveal that the synchronization between multiple brains is enhanced in the scenario of collaborative tasks since the participants share the same strategy and objective [11], [37]. An intuitive reason is that inter-brain synchrony is positively correlated with performance metrics. Moreover, alpha band was relevant to the social interactions [38] and effectively in motor imagery task, which might explain why the decoding accuracy obtained from  $\alpha$ - $\alpha$  is higher than  $\alpha$ - $\beta$ ,  $\beta$ - $\alpha$  and  $\beta$ - $\beta$  cases.

## VII. CONCLUSION

In this paper, we proposed an EEG-based multi-brain MI decoding method embedding the coupling relation feature extraction and few-shot learning to efficiently learn representative multi-brain features and classify them with limited amount of EEG data. A comprehensive multi-brain MI-BCI decoding study was conducted by employing the hypergraph learning, Tucker decomposition and relation network. Our experimental results validated the superiority of the multi-brain experimental paradigm over the single-brain paradigm. Notably, the results confirmed the effectiveness and importance of the coupling relationships between participants to the classification of MI. We also proposed a cross-brain CDC metric to quantify the relationships between brains. The results reiterate the contribution of coupling relationship to the classification performance, showing a positive correlation between the coupling strength and classification performance. Finally, the results demonstrate that our proposed method still works well even in the case of the limited samples (*i.e.*, few shots). This study could provide a solution for high-performance BCI involving multiple participants.

## REFERENCES

- [1] Y. Li, L. Guo, Y. Liu, J. Liu, and F. Meng, "A temporal-spectral-based squeeze-and-excitation feature fusion network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1534–1545, 2021.
- [2] A. Vuckovic, S. Pangaro, and P. Finda, "Unimanual versus bimanual motor imagery classifiers for assistive and rehabilitative brain computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 12, pp. 2407–2415, Dec. 2018.
- [3] J. Cantillo-Negrete, R. I. Carino-Escobar, P. Carrillo-Mora, D. Elias-Vinas, and J. Gutierrez-Martinez, "Motor imagery-based brain-computer interface coupled to a robotic hand orthosis aimed for neurorehabilitation of stroke patients," *J. Healthcare Eng.*, vol. 2018, pp. 1–10, Apr. 2018.
- [4] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces—A review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.
- [5] Y. Peng, F. Qin, W. Kong, Y. Ge, F. Nie, and A. Cichocki, "GFIL: A unified framework for the importance analysis of features, frequency bands, and channels in EEG-based emotion recognition," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 3, pp. 935–947, Sep. 2022.
- [6] A. M. Azab, L. Mihaylova, K. K. Ang, and M. Arvaneh, "Weighted transfer learning for improving motor imagery-based brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 7, pp. 1352–1359, Jul. 2019.
- [7] K. Keng Ang, Z. Yang Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intelligence)*, Jun. 2008, pp. 2390–2397.
- [8] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.
- [9] M. Dai, D. Zheng, R. Na, S. Wang, and S. Zhang, "EEG classification of motor imagery using a novel deep learning framework," *Sensors*, vol. 19, no. 3, p. 551, Jan. 2019.
- [10] O. George, R. Smith, P. Madiraju, N. Yahyasoltani, and S. I. Ahamed, "Motor imagery: A review of existing techniques, challenges and potentials," in *Proc. IEEE 45th Annu. Comput., Softw., Appl. Conf. (COMPSAC)*, Jul. 2021, pp. 1893–1899.
- [11] N. Sinha, T. Maszczyk, Z. Wanxuan, J. Tan, and J. Dauwels, "EEG hyperscanning study of inter-brain synchrony during cooperative and competitive interaction," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2016, pp. 004813–004818.
- [12] D. Zhang, "Computational EEG analysis for hyperscanning and social neuroscience," in *Computational EEG Analysis*. Singapore: Springer, 2018, pp. 215–228.
- [13] A. Czeszumski et al., "Hyperscanning: A valid method to study neural inter-brain underpinnings of social interaction," *Frontiers Hum. Neurosci.*, vol. 14, p. 39, Feb. 2020.
- [14] A. Solon, S. Gordon, J. McDaniel, and V. Lawhern, "Collaborative brain-computer interface for human interest detection in complex and dynamic settings," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2018, pp. 970–975.
- [15] M. R. Short, J. C. Hernandez-Pavon, A. Jones, and J. L. Pons, "EEG hyperscanning in motor rehabilitation: A position paper," *J. NeuroEng. Rehabil.*, vol. 18, no. 1, p. 98, Jun. 2021.
- [16] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19, 2006, pp. 1–8.
- [17] Z. Liang, S. Oba, and S. Ishii, "An unsupervised EEG decoding system for human emotion recognition," *Neural Netw.*, vol. 116, pp. 257–268, Aug. 2019.
- [18] X. Gao et al., "A seizure detection method based on hypergraph features and machine learning," *Biomed. Signal Process. Control*, vol. 77, Aug. 2022, Art. no. 103769.
- [19] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.
- [20] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of EEG signals: A brief review," *J. Neurosci. Methods*, vol. 248, pp. 59–69, Jun. 2015.
- [21] L. T. Thanh, N. T. A. Dao, N. V. Dung, N. L. Trung, and K. Abed-Meraim, "Multi-channel EEG epileptic spike detection by a new method of tensor decomposition," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016023.
- [22] Y. Jonmohamadi, S. Muthukumaraswamy, J. Chen, J. Roberts, R. Crawford, and A. Pandey, "Extraction of common task features in EEG-fMRI data using coupled tensor-tensor decomposition," *Brain Topography*, vol. 33, no. 5, pp. 636–650, Sep. 2020.
- [23] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1199–1208.
- [24] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
- [25] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *Proc. ICML Deep Learn. Workshop*, 2015, vol. 2, no. 1, pp. 1–30.
- [26] T. Duan et al., "Meta learn on constrained transfer learning for low resource cross subject EEG classification," *IEEE Access*, vol. 8, pp. 224791–224802, 2020.
- [27] D. Li, P. Ortega, X. Wei, and A. Faisal, "Model-agnostic meta-learning for EEG motor imagery decoding in brain-computer-interfacing," in *Proc. 10th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, May 2021, pp. 527–530.
- [28] S. An, S. Kim, P. Chikontwe, and S. H. Park, "Dual attention relation network with fine-tuning for few-shot EEG motor imagery classification," *IEEE Trans. Neural Netw. Syst.*, early access, Jun. 28, 2023, doi: [10.1109/TNNLS.2023.3287181](https://doi.org/10.1109/TNNLS.2023.3287181).
- [29] M. Wang, X. Hao, J. Huang, W. Shao, and D. Zhang, "Discovering network phenotype between genetic risk factors and disease status via diagnosis-aligned multi-modality regression method in Alzheimer's disease," *Bioinformatics*, vol. 35, no. 11, pp. 1948–1957, Jun. 2019.
- [30] H. Guo, Y. Li, Y. Xu, Y. Jin, J. Xiang, and J. Chen, "Resting-state brain functional hyper-network construction based on elastic net and group lasso methods," *Frontiers Neuroinform.*, vol. 12, p. 25, May 2018.
- [31] Y. Li et al., "Multimodal hyper-connectivity of functional networks using functionally-weighted LASSO for MCI classification," *Med. Image Anal.*, vol. 52, pp. 80–96, Feb. 2019.
- [32] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *J. Neural Eng.*, vol. 16, no. 5, Oct. 2019, Art. no. 051001.
- [33] S. Cole and B. Voytek, "Cycle-by-cycle analysis of neural oscillations," *J. Neurophysiology*, vol. 122, no. 2, pp. 849–861, Aug. 2019.
- [34] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent.*, 2017, pp. 1–14.
- [35] M. Xu, F. He, T.-P. Jung, X. Gu, and D. Ming, "Current challenges for the practical application of electroencephalography-based brain-computer interfaces," *Engineering*, vol. 7, no. 12, pp. 1710–1712, Dec. 2021.
- [36] J. Ju, A. G. Feleke, L. Luo, and X. Fan, "Recognition of drivers' hard and soft braking intentions based on hybrid brain-computer interfaces," *Cyborg Bionic Syst.*, vol. 2022, Jan. 2022, doi: [10.34133/2022/9847652](https://doi.org/10.34133/2022/9847652).
- [37] V. Wikström et al., "Inter-brain synchronization occurs without physical co-presence during cooperative online gaming," *Neuropsychologia*, vol. 174, Sep. 2022, Art. no. 108316.
- [38] P. Léné et al., "Is there collaboration specific neurophysiological activation during collaborative task activity? An analysis of brain responses using electroencephalography and hyperscanning," *Brain Behav.*, vol. 11, no. 11, p. e2270, Nov. 2021.