Computing and Informatics, Vol. 43, 2024, 735–755, doi: 10.31577/cai_2024_3_735

META-LEARNING FOR EEG MOTOR IMAGERY CLASSIFICATION

Jian Yu, Lili Duan, Hongfei Ji*

College of Electronic and Information Engineering
Tongji University
Shanghai 201804, China
e-mail: {yujian, jhf}@tongji.edu.cn

Jie Li*

College of Electronic and Information Engineering
Tongji University
Shanghai 201804, China
&
Translational Research Center, Shanghai Yangzhi Rehabilitation Hospital
Shanghai Sunshine Rehabilitation Center, Tongji University, Shanghai
e-mail: nijanice@163.com

Zilong Pang

School of Software
Intelligent Data Processing Engineering Research Center of Henan Province
Henan University, Kaifeng, 475004, China
e-mail: jszxpzl@henu.edu.cn

Abstract. A brain-computer interface (BCI) based on motor imagery (MI) enables users to communicate with the computer directly using brain signals. However, due to the low signal-to-noise ratio and significant inter-subject variations, its long calibration time hinders the development of brain-computer interfaces. Meta-learning,

* Corresponding authors

as the current popular framework in few-shot learning, enables the model to quickly adapt to new few-shot tasks by learning a series of similar tasks. In this paper, we propose a novel algorithm that combines pre-training and meta-learning. It can reduce the number of training samples required for the target subject and meanwhile ensure stable and reliable BCI performance. This is achieved by first learning general feature representations from a large number of other subjects' data through pre-training, and then further optimizing the model by task scale using meta-learning based on the pre-training stage. Applications to a dataset confirm the effectiveness of the combination of pre-training and meta-learning. Results indicate that the proposed algorithm outperforms the considered comparable baseline algorithm, early meta-learning frameworks without pre-training, and fine-tuning in transfer learning. In addition, experiments show that the selection of source subjects can further improve the overall performance of the algorithm, and the pre-training stage is crucial for the model to achieve good performance. This study has significant instruction for the application of BCI in the field of medical rehabilitation. It can improve the classification performance of the disabled with poor motor imagination characteristics by using the data of healthy typical subjects and greatly reducing their training time.

Keywords: Brain-computer interface, motor imagery, pre-training, fine-tuning

1 INTRODUCTION

Brain-Computer Interface (BCI) can realize direct communication or control of the external device by the human brain without relying on peripheral nerves and muscle tissues and thereby help those who are paralyzed but cognitively intact [1, 2]. In BCIs, electroencephalogram (EEG) has become one of the most commonly used signals because of its non-invasiveness, convenience, inexpensiveness, and safety. Motor imagery (MI), as an active control paradigm, is widely used in fields like medical rehabilitation and has become a research hotspot in the current BCI field [3, 4]. In order to build a stable and reliable classifier, this autonomously controlled brain-computer interface usually requires subjects to undergo long-term calibration experiments to ensure that easily identifiable EEG signals are generated. However, the entire training process can easily lead to mental fatigue or even to be unbearable. Therefore, reducing the number of training samples required by subjects and meanwhile keeping accuracy in an acceptable range is critical to the success of MI-based BCIs [5, 6, 7].

At present, most of the literature uses transfer learning to solve this problem [8, 9, 10, 11, 12, 13, 14]. In MI-based BCIs, transfer learning can use data from other subjects to facilitate learning for a target subject [15]. For traditional models, many proposed algorithms focus on improving the common spatial patterns (CSP) by modifying the covariance matrix estimation method [8, 9] or the CSP optimization function [10, 11]. For deep models, a common technique is called fine-tuning (FT), which is established on a pre-trained model and then fine-tunes the parameters based on the target subject/session [14].

However, most of the above-mentioned algorithms still require at least 10 samples per class of the target subject to help transfer training [7, 8, 12]. Little research so far has looked at recognizing MI tasks when the target subject has extremely few labeled samples (e.g., 1 or 5 trials per class), especially in the context of deep learning models.

Deep learning has become an effective tool for developing BCI systems in recent years [16, 17]. The good robustness, however, heavily relies on abundant labeled instances. Therefore, it is a great challenge for deep models to extract information from a small number of training samples.

Learning from a few number of samples, called few-shot learning (FSL), is extensively studied in computer vision [18, 19, 20, 21]. Although transfer learning can utilize data from other subjects, its adaptability is limited when the target subject has little training data. Therefore, task-based Meta-Learning in FSL has been widely studied in recent years.

The number of categories to be classified is N, and there are only K labeled samples in each category (where K is generally between 1 to 5), then we call this an *N-way K-shot* FSL problem. A "task" in meta-learning usually takes the form of such an *N-way K-shot*, which contains $N \times K$ training samples (called support set refers to the training set within the task) and $N \times Q$ testing samples (called query set refers to the testing set within the task). The key idea of meta-learning is to treat tasks as samples, and through the learning of a large number of similar tasks, the model can adapt to new tasks quickly and accurately. Previous learning methods learn from a set of labeled data samples, while meta-learning learns from a set of labeled tasks, which elevates the learning level from data to tasks. It provides a new perspective for solving few-shot problems.

Although there have been related studies applying meta-learning to the field of brain-computer interfaces [22, 23], some recent work that takes advantage of both transfer learning and meta-learning has established new state-of-the-art results in the field of computer vision. Inspired by these advanced methods used for image classification [19, 21], in this article, we propose a novel few-shot learning algorithm in the context of MI-BCI, which combines the pre-training stage in fine-tune techniques and the task-based training scheme in meta-learning. Since the algorithm adds a simple and effective meta-learning framework on the basis of pre-training, we call it the Pre-Trained Meta-Baseline (PTMB) model.

See the conceptual comparison of related methods (fine-tune in transfer learning, early meta-learning frameworks without pre-training, and proposed PTMB strategy) in Figure 1. In fine-tune strategy, the model is first pre-trained on large-scale datasets from other subjects, and then fine-tuned using a small amount of data from the target subject. In early meta-learning frameworks, the model is trained on a large number of different tasks formed by other subjects, and then quickly adapted to new tasks through strategies such as learning to fine-tune or metric learning. Our proposed PTMB model combines the above mentioned two strategies. It retains the pre-training module and further uses constructed few-shot tasks for meta-training.

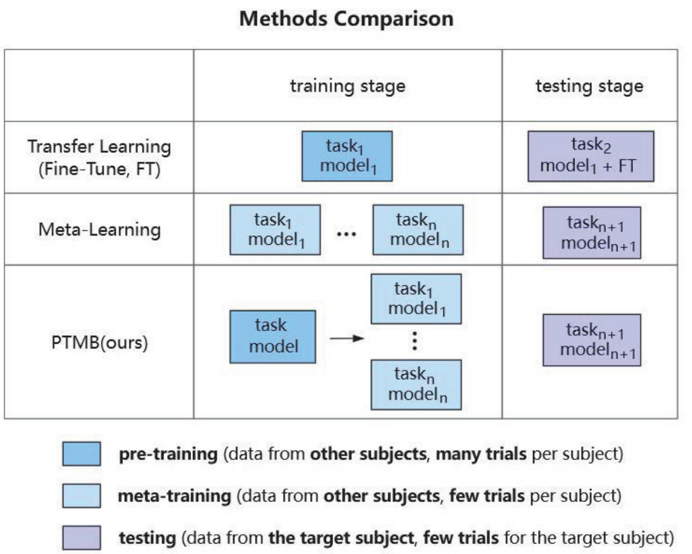

Figure 1. Conceptual comparison of related methods (fine-tune in transfer learning, early meta-learning frameworks, and proposed PTMB)

In general, our PTMB model mainly contains two stages:

1. Pre-training a deep model with a fully connected (FC) layer on source subjects data. Notably, considering that few-shot learning tasks are sampled according to subjects in the subsequent meta-training stage, we regard different subjects as different categories in this stage. Specifically, assuming that there are data from m subjects in the pre-training stage, and they perform N types of MI tasks, then the FC layer of our model will have $m \times N$ outputs representing $m \times N$ categories, instead of N categories in previous methods [14]. Then we remove the FC layer to obtain the feature encoding module.
2. Meta-updating the model parameters by performing *N-way K-shot* tasks with the cosine nearest-centroid method. Finally, given a new subject with a few samples, we predict the label using the abovementioned cosine nearest-centroid metric.

In summary, the main contributions of our work are as follows:

1. We propose a novel PTMB framework applied to EEG-based MI classification, which uses meta-learning on the basis of pre-trained models to further improve performance. Moreover, the proposed framework is a general-purpose one, which

can be applied to different few-shot classification problems of BCIs by simply changing the feature encoding module.

2. In the pre-training stage, we first regard the MI tasks of different subjects as different categories instead of mixing the same MI task data of various subjects. This feature extraction schema implicitly learns the differences between individuals. Besides, owing to the meta-learning method, the model can extract and propagate transferable knowledge from a collection of tasks. Thus, when faced with the target subject, the model shows better generalization performance.

The remainder of this article is arranged as follows. Section 2 elaborates some basic concepts and related work. Section 3 introduces the overall framework of the algorithm and the process of each stage. Section 4 presents the experiments and implementation details. Section 5 shows results and discussion from multiple perspectives. Finally, Section 6 summarizes the complete text.

2 RELATED WORK

In this section, we first introduce the relationship between few-shot learning and meta-learning, as well as some current representative architectures of meta-learning. Then, we discuss the differences and relationships between several related studies and our proposed algorithm.

2.1 Few-Shot Learing and Meta-Learning

Few-shot learning aims to identify new classes from very few labeled examples. Conventional methods such as data augmentation and regularization techniques can only alleviate overfitting, but they do not solve the problem. Besides, this setting challenges the standard "fine-tuning" practice in deep models of transfer learning [18]. Thus, the current few-shot learning methods mostly follow the meta-learning framework, which can be roughly classified into three types: initialization-based methods, memory-based methods, and metric-based methods [21].

1. **Initialization-based methods** address the FSL problem through "learning to fine-tune". One classical approach [18] called MAML aims to learn good model initialization so that the model can converge fast on a limited number of labeled examples within several optimization steps. Ravi and Larochelle [24] not only focus on initialization but also trains an LSTM-based optimizer to help with fine-tuning.
2. **Memory-based methods** [25, 26, 27] mainly focus on training an RNN-based model to store historical information of previous tasks and extract important knowledge required to solve the problem. Then the stored knowledge can be used to assist the learning of new tasks.
3. **Metric-based methods** aim to learn an effective deep representation and then make their prediction based on distance or metric. Vinyals et al. [28] use cosine

similarity to classify query samples. Snell et al. [29] compute squared Euclidean distance to class-mean representation and classify query samples by the nearest-centroid method. Sung et al. [30] learn metric relationships between samples by using relation networks.

In addition to the different variants of these methods, some recent work [19, 21] has achieved better results by exploiting the advantages of both pre-trained classifier and meta-learning.

2.2 Meta-Learning in EEG Classification

Although meta-learning has extensive research on image classification, the literature on its application to EEG is relatively limited. Tynes and Parsapoor [31] employ MAML [18] and prototypical networks [29] to distinguish people with schizophrenia from healthy controls based on their brain activity. Zhu et al. [32] use meta-learning in seizure detection to mitigate the inter-patient seizure pattern variation. All of their methods apply meta-learning frameworks to the field of disease diagnosis based on EEG, rather than the MI-based BCI system we have studied so far.

Li et al. [22] and Duan et al. [33, 34] both draw on the ideas of MAML [18] to train an MI-based EEG decoder that can be fine-tuned easily on a new user with only a few gradient steps. The scenarios where they use meta-learning are similar to ours, but they mainly use initialization-based methods, unlike our metric-based methods, which do not require fine-tuning to update the model and are more convenient for real-time applications. An et al. [23] propose a metric-based algorithm that learns an end-to-end relation net to calculate relation scores between the support data and query data, thus it can classify target subject categories with limited MI EEG data. However, their method does not take advantage of pre-trained classifiers which have proven to generalize better than randomly initialized ones [35].

Therefore, in this paper, we propose a novel few-shot learning method that takes both strengths of pre-trained classifiers and meta-learning strategy, it outperforms the method of fine-tuning pre-trained classifiers or other traditional meta-learning methods.

3 METHODOLOGY

First, we formally define the few-shot learning problem we want to solve in the context of brain-computer interfaces.

Assume that there is a target subject and m auxiliary subjects (or called source subjects in transfer learning) performing N motor imagination tasks. The target subject has only K labeled trials per MI task, while the source subject has a relatively large number of labeled trials. Thus, for the target subject, it forms an *N-way K-shot* few-shot problem. The aim of our model is to leverage the data from source subjects to achieve better models for the target subject.

In the remainder of this section, we first introduce the overall framework of our method, and then we describe every essential stage in the proposed framework.

3.1 The Framework of the Approach

The overall framework of the proposed method is shown in Figure 2, which mainly consists of two stages, the pre-training stage, and the meta-learning stage. First, in the pre-training stage, the data from a pool of source subjects is used to learn the feature encoder f_θ, and then the learned feature encoder is used as the initial model parameter of the following meta-learning stage. Subsequently, in each iteration of meta-training, we randomly select a subject from the m subjects and use its data to construct an *N-way K-shot* task (2-way 2-shot is taken as an example in Figure 2). The average feature vector of each class in the support set is calculated as the centroid representing that class, and the query set samples are classified according to the nearest centroid principle using the cosine distance.

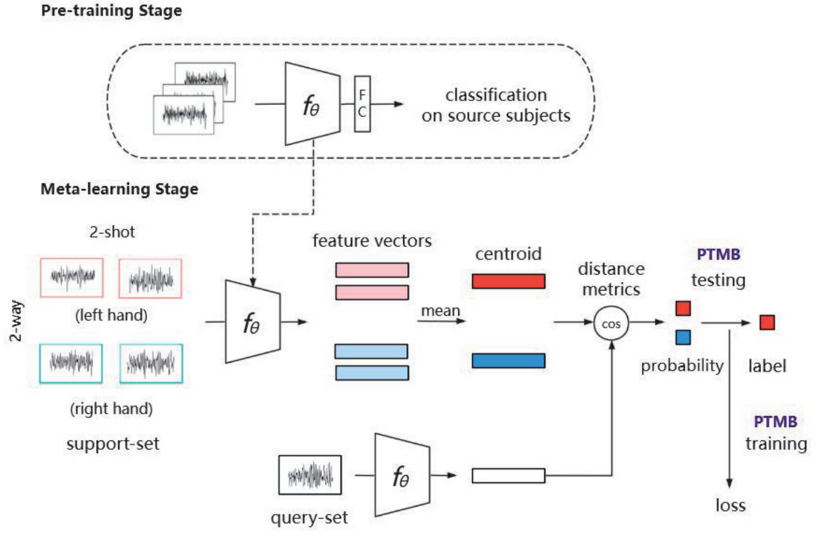

Figure 2. The framework of our proposed approach PTMB

The details of the two stages are given below.

3.2 Pre-Training Stage

The first fundamental step is to pre-train an effective EEG feature encoder based on the large-scale EEG data previously recorded from source subjects.

We adopt a model similar to [16] to extract meaningful features from raw EEG signals. The specific process is as follows. Raw EEG signals formed as a 2D array with the number of electrode channels as the height and the number of time steps as the width are fed into the deep model. The first two layers perform a temporal convolution and a spatial convolution. These steps are analogous to the bandpass filter and spatial filter steps in Filter Bank Common Spatial Patterns (FBCSP) algorithm [36]. After that, we apply a batch normalization layer to reduce the internal covariate shift of the network [37]. Then, a squaring nonlinearity, a mean pooling layer, and a logarithmic activation function followed. Together these steps are analogous to the trial log-variance computation in FBCSP. Subsequently, we add a dropout layer to avoid overfitting, which prevents units from co-adapting too much [38]. Finally, we use a convolutional layer to turn the feature map into a feature vector, and then input it into the FC layer for classification.

Note that in the previous methods that leverage pre-training for MI classification [14, 33], data from all source subjects under each MI category are merged to form a large-scale dataset. However, simply combining data may cause key differences across subjects to cancel out. Thus, we treat the data of different subjects on the same MI task as different categories (that is, given m source subjects perform N motor imagination tasks, we have a total of $m \times N$ categories instead of the traditional N categories). During the pre-training stage, the standard cross-entropy loss function is used to iteratively update the parameters of the network, then the FC layer of the network is removed to obtain the feature encoder, which maps the input EEG signal to the corresponding feature space.

The feature space extracted in this way is shared across subjects. It facilitate the ability to fast adapt to new subjects in the subsequent meta-learning stage.

3.3 Meta-Learning Stage

After getting the pre-trained feature encoder f_θ, we further optimize the model in the meta-learning stage.

Specifically, we sample *N-way K-shot* tasks from source subjects' data. In each task, K support samples and Q query samples in each MI class are selected from one randomly chosen source subject. Unlike the practice of constructing few-shot tasks from the mixed data of all source subjects [33], we treat each source subject as an independent sample, which is in line with the idea of separate the categories of different subjects in the pre-training stage.

In each *N-way K-shot* task, we first obtain the feature vectors (pink, light blue, and white rectangles in Figure 2) of the support set and query set samples through the feature encoder, and then calculate the average feature vector (red and blue rectangles in Figure 2) under each category in the support set as the centroid of the class. Let the average feature vector under category i be w_i, then its formal representation is as follows:

$$w_i = \frac{1}{|K|} \sum_{x \in S_i} f_\theta(x), \tag{1}$$

where x represents a sample in the support set, S_i denotes the subset consisting of K samples under the same category i ($i = 1, \dots, N$) in the support set, and f_θ denotes feature encoder.

Then for a query instance x in the meta-training task, we calculate the probability that the x belongs to class i according to the cosine similarity between its feature vector and the centroid of class i:

$$p(y = i|x) = \frac{\exp(\cos\langle f_\theta(x), w_i\rangle)}{\sum_{j=1}^{N} \exp(\cos\langle f_\theta(x), w_j\rangle)}, \tag{2}$$

where $\cos\langle\cdot,\cdot\rangle$ represents the cosine similarity of two vectors.

For the tasks extracted from the source subjects, we calculate the cross-entropy loss between the predicted probability and the labels in the query set, and update the network weights accordingly.

Finally, after completing the meta-training phase, we can use the model to test the performance of the few-shot task on the target subject.

4 EXPERIMENTS

In this section, we first introduce the data set we use and the preprocessing process, then we explained our implementation details.

4.1 Dataset and Preprocessing

To evaluate the performance of our proposed method, we used the dataset from the BCI competition IV 2a [39]. This dataset uses 22 electrodes to collect the EEG signals of 9 subjects. Experiments were recorded at a sampling rate of 250 Hz, and bandpass filtering was performed between 0.5 Hz and 100 Hz. During the experiments, the subjects were instructed to perform one of the four following motor imagery tasks: left hand, right hand, foot, or tongue. Each subject went through two sessions on different days, and 72 trials were recorded for each class in each session. Thus, for each subject, each class contained a total of 144 trial samples. In this paper, only data from left and right-hand motor imagery were used.

For data preprocessing, we first performed band-pass filtering of 4–38 Hz on the acquired continuous EEG signals. Since this frequency range usually contains the frequency components generated when the brain performs motor imagery tasks. Then, we performed exponential moving average (EMA) to calculate the exponential moving average and the corresponding variance within each electrode to normalize the continuous EEG data. Finally, we cut out compute windows, i.e., the inputs for the deep networks. Specifically, the timing scheme of the paradigm is shown in Figure 3. We segmented EEG data of each trial starting 1–4 seconds from the arrow cue (the blue period in Figure 3) since this is the phase where motor imagery is performed and the data is of higher quality. Thus, the input of each trial contains 22 electrodes and 750 sampling points (250 Hz $*$ 3 s).

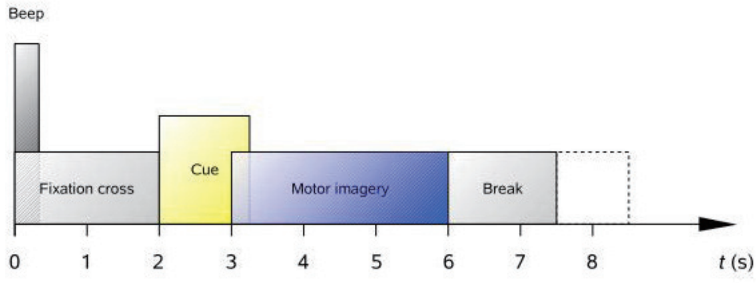

Figure 3. Timing scheme of the paradigm

4.2 Implementation Details

In our experiment, we separate 9 subjects into 8 source subjects providing training data in the pre-training and meta-training stage, and one target subject providing few-shot test samples for evaluating the trained model.

In the pre-training stage, the detailed architecture of our deep model is shown in Figure 4. Note that the batch normalization, dropout layer, and activation functions such as square, logarithmic that do not affect the output shape of each layer are not shown in this figure. As mentioned above, the input of the network is 22 electrodes with 750 sampling points, and the output is a 16-dimensional vector representing the predicted probabilities of 16 categories (8 source subjects * 2 classes of MI tasks). We use an SGD optimizer with a momentum of 0.9, and the learning rate is 0.000625. We train 100 epochs with a batch size of 64 on a single Quadro RTX 5000 GPU.

In meta-training stage, we use the same optimizer settings and learning rate as in the pre-training stage. And we train 30 epochs, each epoch contains 100 *2-way K-shot* ($K = 1, 5$) tasks. In each task, we randomly select one source subject and pick K labeled instances in each class as the support set and 15 instances in each class as the query set. In meta-testing stage, we randomly constructed 200 *2-way K-shot* ($K = 1, 5$) tasks from the data of the target subject, and take the average result as the final few-shot learning accuracy rate.

5 RESULTS AND DISCUSSION

This section verifies the effectiveness of the proposed method on a real EEG dataset. The results are compared with a baseline method and three typical methods. In addition, we provide some extra results analyzed from multiple aspects.

5.1 Results

We compare the performance of our proposed approach (PTMB) with one baseline method (the traditional classification method Regularized CSP, RCSP [40]), two

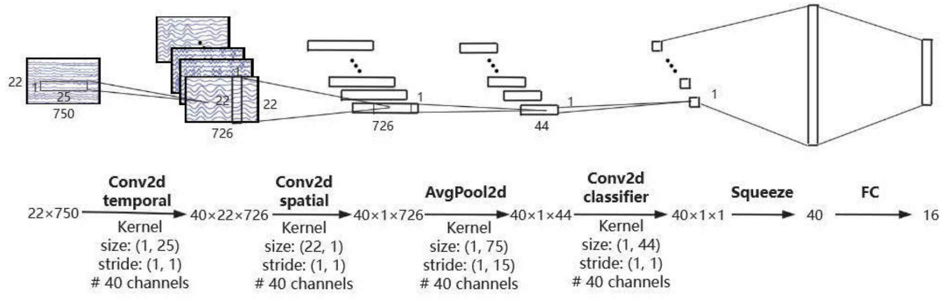

Figure 4. The architecture of our deep model

meta-learning methods (RelationNet proposed in [24] and MAML proposed in [18]), and one transfer learning method (Fine-tune, FT). To get a fair comparison, we use the same data preprocessing process and ensure that all methods use the same amount of data, that is, a small number of samples of the target subject and all data of source subjects. Furthermore, for the latter three contrastive methods involving deep learning, we use the same feature extraction network architecture to guarantee comparability in performance. Four comparison the algorithms are introduced as follows.

1. **RCSP:** This is a baseline approach. It is based on the classical CSP algorithm framework and uses regularization parameters to linearly combine the class average covariance matrix of source and target subjects, so as to achieve effective integration of the two kinds of data. See [40] for details. And then we feed the extracted features to an SVM classifier.
2. **RelationNet:** This is a metric-based meta-learning algorithm. It uses the data of source subjects to construct multiple few-shot tasks and performs few-shot classification by learning the relation scores of query samples and support samples in the task through deep relation networks [24]. It differs from our method in that it does not include a pre-training stage and uses different metric criteria.
3. **MAML:** This is an initialization-based meta-learning algorithm. It aims to train a good set of initialization parameters on a large set of few-shot tasks constructed from source subjects' data. For each task, the model is first trained on the support set and then tested on the query set. Finally, the sum of the test errors in each task is used as the overall training error for training to obtain a set of initialization parameters. It differs from our method in that it still needs to be fine-tuned after training [18].
4. **Fine-tune (FT):** This is a transfer learning method. Unlike task-based meta-learning, the fine-tuning process builds on models pre-trained on source data and uses a relatively small amount of target data for iterative learning.

All the above methods are applied to the dataset from the BCI competition IV 2a. Since many other algorithms [7, 8, 12] use at least 10 or more samples per class of target objects and pay little attention to performance with fewer samples, we focus on evaluating the performance of all algorithms under the settings of $K = 1$ and $K = 5$. Specifically, 1 or 5 samples of each class in each subject were randomly selected and performed 200 times in repetition. And the average classification accuracies for all 9 subjects are listed in Tables 1 and 2.

Algorithm	Subjects									Overall	
	sub1	sub2	sub3	sub4	sub5	sub6	sub7	sub8	sub9	Mean	Std
RCSP	81.87	50.58	79.93	55.95	50.99	54.58	61.29	66.56	79.10	64.54	12.82
RelationNet	89.33	**54.81**	72.34	61.91	50.46	58.74	51.62	63.37	81.35	64.88	13.50
MAML	89.21	52.73	73.05	**63.87**	50.24	**59.35**	51.27	65.89	80.27	65.10	13.58
Fine-tune	89.84	51.87	80.39	61.47	**52.49**	55.14	54.22	**67.72**	80.15	65.92	14.31
PTMB	**90.74**	51.28	**81.38**	62.33	51.84	54.23	**62.12**	67.63	**85.32**	**67.43**	14.96

Table 1. Classification accuracies (%) of all subjects when only 1 trial per a class was available for training from the target subject

Algorithm	Subjects									Overall	
	sub1	sub2	sub3	sub4	sub5	sub6	sub7	sub8	sub9	Mean	Std
RCSP	84.16	51.46	83.75	59.20	52.61	57.09	65.06	71.14	80.85	67.26	13.20
RelationNet	93.74	**56.20**	76.42	62.21	51.34	**66.71**	55.25	72.12	82.21	68.47	13.99
MAML	90.13	54.11	80.13	**65.25**	51.93	59.86	60.41	72.97	84.27	68.78	13.71
Fine-tune	93.21	52.14	82.22	61.52	**53.31**	61.68	56.39	73.23	86.67	68.93	15.37
PTMB	**96.02**	52.24	**88.45**	63.74	52.23	59.79	**70.79**	**73.76**	**88.32**	**71.70**	15.72

Table 2. Classification accuracies (%) of all subjects when only 5 trials per a class were available for training from the target subject

When the target subject has only one training trial per a category (Table 1, $K = 1$), the proposed PTMB algorithm combining pre-training and meta-learning has the most significant improvement over the baseline method (significance level $\alpha = 0.05$, p value $= 0.031 < 0.05$), reaching 67.43 %, a 2.89 % improvement over the baseline method. And it also outperformed the results of RelationNet, MAML, and FT by an average of 2.55 %, 2.33 %, and 1.51 %. Similarly, when the target subject has only five training trials per a category (Table 2, $K = 5$), the proposed PTMB algorithm is also significantly improved than the baseline method (significance level $\alpha = 0.05$, p value $= 0.006 < 0.05$). And it achieved the highest average result, reaching 71.70 %, which is 4.44 %, 3.23 %, 2.92 %, and 2.77 % higher than baseline, RelationNet, MAML, and Fine-tune, respectively.

5.2 The Effect of Choosing Good Subjects

Looking deeper in Table 1 and Table 2, we can observe that for those good subjects (such as sub1, 3, and 9) which can achieve high accuracy even under the baseline method, the proposed PTMB method that combines pre-training and meta-learning tends to achieve optimal results among multiple comparison methods. While for the remaining subjects with poor performance, the proposed method is inferior to the comparative method in most cases.

From the following aspects thinking we can investigate its reason. First, the high baseline accuracy of good subjects may indicate that their EEG data itself is highly separable, and the pre-training stage that has a large number of samples enables our deep learning model to learn basic features with strong discrimination. What is more, the meta-learning stage makes our model adapt to few-shot tasks, so it is better than other methods. While for the second point, our method has a lower accuracy in bad subjects, which may indicate that in addition to the low separability of their EEG data itself, some other bad subjects' data makes the feature extracted unobvious and ultimately result in unsatisfactory results. So, is it possible to use only the data of good subjects to allow the model to learn more pure features and thereby improve the accuracy of bad subjects? To verify this conjecture, we train the model using good subjects 1, 3, 9 and compare the results with the model trained in all other subjects.

The results of $K = 5$ are reported in Table 3, for the results changing more obviously under this condition. It can be seen from the table that for the bad subjects where EEG patterns are not obvious, choosing good subjects can improve the accuracy, it indicates that reducing the amount of data but improving the data quality brings better results. While for the good subjects with high data separability, the dropping performance may indicate that the reduction in data volume has a greater impact than the improvement in data quality. Therefore, in general, data volume and data quality are both important factors, but they have different effects due to different types of subjects.

In addition, we also tested the performance of all algorithms under the condition of choosing good source subjects. Table 4 shows the classification accuracies of 9 subjects of all algorithms when $K = 5$. Comparing it with Table 2, it can be found that after the selection of good source subjects, the performance of different algorithms on different target subjects increases or decreases, but in general, the proposed PTMB algorithm still achieves the highest average performance. As a result, it reaches 73.13 %, which is 7.73 %, 4.40 %, 3.89 %, and 4.76 % higher than baseline, RelationNet, MAML, and Fine-tune, respectively.

We also performed paired t-tests for significance analysis on the performance of different algorithms under this condition, and the results are shown in Table 5. It can be observed from the table that the paired t-test p-value results of the proposed PTMB algorithm relative to all other comparison algorithms are all less than 0.05, thus indicating that the proposed algorithm significantly outperforms all other algorithms.

Subject	Not choose	Choose good subjects
sub2	52.24	55.43
sub4	63.74	69.27
sub5	52.23	55.82
sub6	59.79	64.89
sub7	70.79	76.97
sub8	73.76	78.19
Mean	62.09	**66.76**

Subject	Not choose	Choose good subjects
sub1	96.02	93.14
sub3	88.45	80.37
sub9	88.32	84.09
Mean	**90.93**	85.87

Table 3. Under the condition of $K = 5$, classification accuracies (%) of choosing good subjects and not choosing good subjects. Subjects are divided into bad subjects (6 of them, shown at the top of the table) and good subjects (3 of them, shown at the bottom of the table, trained using good subjects excludes the data from themselves).

Algorithm	Subjects									Overall	
	sub1	sub2	sub3	sub4	sub5	sub6	sub7	sub8	sub9	Mean	Std
RCSP	81.87	50.71	79.96	60.86	51.95	58.31	61.54	62.69	80.73	65.40	12.29
RelationNet	88.03	**57.35**	77.17	60.92	54.16	64.39	64.05	71.53	80.97	68.73	11.43
MAML	89.43	54.28	**82.39**	68.17	50.89	59.27	66.32	70.16	82.29	69.24	13.33
Fine-tune	85.37	53.46	78.29	61.96	53.28	61.43	63.07	71.98	**86.48**	68.37	12.73
PTMB	**93.14**	55.43	80.37	**69.27**	**55.82**	**64.89**	**76.97**	**78.19**	84.09	**73.13**	12.79

Table 4. Under the condition of selecting good source subjects, classification accuracies (%) of all subjects of all different algorithms when $K = 5$

5.3 The Importance of Pre-Training

Choosing good subjects or not can affect the results, which makes us realize the importance of pre-training. So, we train our PTMB model from scratch and compare it with the one with pre-training. The results are shown in Table 6. From the

Methods contrast	p-value
PTMB vs. RCSP	0.003
PTMB vs. RelationNet	0.018
PTMB vs. MAML	0.017
PTMB vs. Fine-tune	0.016

Table 5. Under the condition of selecting good source subjects ($K = 5$), the p-value under the paired t-test (significance level $\alpha = 0.05$) of different algorithms

table, we observe that the proposed PTMB with pre-training achieves higher accuracies when compared to PTMB trained from scratch. Our results indicate that the combination of pre-training and meta-learning can effectively improve performance. One potential reason is that pre-trained classifiers can provide extra transferability from source subjects to the target subject in the meta-learning stage.

	Training Scheme	Mean Acc.
1-shot	From scratch	64.68
	Pre-training	**67.43**
5-shot	From scratch	68.55
	Pre-training	**71.70**

Table 6. Mean classification accuracies (%) of nine subjects under two different training schemes

5.4 The Effect of Choosing Different Metrics

In addition to the importance of pre-training, the choice of different metrics will also affect the results. Therefore, we choose two of the most commonly used distance measures, Euclidean distance and cosine similarity, to conduct experiments. The results are shown in Table 7. It can be observed from the table that the average accuracy using cosine similarity is slightly higher than that using Euclidean distance. But overall, the difference is not significant, which indicates that these two metrics have less influence than subject selection and pre-training in the proposed framework. In future work, other different metrics can be tried to further improve the accuracy of the proposed algorithm.

	Metrics	Mean Acc.
1-shot	Euclidean	66.21
	Cosine Similarity	**67.43**
5-shot	Euclidean	71.31
	Cosine Similarity	**71.70**

Table 7. Mean classification accuracies (%) of nine subjects under two different metrics

6 CONCLUSIONS

In this paper, we present a novel few-shot learning method that performs well on target subjects even with limited samples by taking the strengths of both classification pre-training and meta-learning. The model mainly contains two stages: First, pre-training a deep network using the data from source subjects to obtain the feature encoder. Second, meta-updating the model parameters by performing auxiliary few-shot tasks to enable the model to generalize better on new few-shot tasks from target subjects. Validated using a public dataset from the international BCI competition, the results show that our model outperforms the baseline method, early meta-learning frameworks, and fine-tune in transfer learning. Moreover, we explore the effect of choosing good subjects and find that the performance depends on both data volume and data quality, suggesting that both factors are noteworthy in future research. In addition, the experiments demonstrate that the pre-training phase is crucial for the model to achieve good performance. And the effect of choosing different metrics is also explored. The study is most significant for the rehabilitative potential of BCIs, because the disabled people with poor motor imagination characteristics can use healthy typical subjects' data to improve their classification performance.

To sum up, our studies show that the proposed approach is an effective method in classifying MI-based EEG signals, especially in the case of lacking the data of target subjects. It shows the great potential to design key technologies in the field of rehabilitation.

Data Availability. A detailed description and download link of the public dataset BCI competition IV 2a used to support the findings of this study are available at `https://www.bbci.de/competition/iv/#dataset2a`.

Conflicts of Interest. The authors declare that there are no conflicts of interest regarding the publication of this paper.

Funding Statement. This work was supported by the Research Project of China Disabled Persons' Federation – On Assistive Technology, Shanghai Shenkang Hospital Development Center's Second Round of "Three-Year Action Plan to Promote Clinical Skills and Clinical Innovation in Municipal Hospitals" Research Physician Innovation Transformation Capability Training Program (No. SHDC-2023CR001), and the Youth Cross Fund General Project (Medical Unit), Henan University (No. S23060Y).

REFERENCES

[1] Wolpaw, J. R.—McFarland, D. J.—Vaughan, T. M.: Brain-Computer Interface Research at the Wadsworth Center. IEEE Transactions on Rehabilitation Engineering, Vol. 8, 2000, No. 2, pp. 222–226, doi: 10.1109/86.847823.

[2] Wolpaw, J. R.—Birbaumer, N.—McFarland, D. J.—Pfurtscheller, G.—Vaughan, T. M.: Brain-Computer Interfaces for Communication and Control. Clinical Neurophysiology, Vol. 113, 2002, No. 6, pp. 767–791, doi: 10.1016/S1388-2457(02)00057-3.

[3] Duan, L.—Li, J.—Ji, H.—Pang, Z.—Zheng, X.—Lu, R.—Li, M.—Zhuang, J.: Zero-Shot Learning for EEG Classification in Motor Imagery-Based BCI System. IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 28, 2020, No. 11, pp. 2411–2419, doi: 10.1109/TNSRE.2020.3027004.

[4] PFURTSCHELLER, G.—LOPES DA SILVA, F. H.: Event-Related EEG/MEG Synchronization and Desynchronization: Basic Principles. Clinical Neurophysiology, Vol. 110, 1999, No. 11, pp. 1842–1857, doi: 10.1016/S1388-2457(99)00141-8.

[5] LOTTE, F.: Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-Based Brain-Computer Interfaces. Proceedings of the IEEE, Vol. 103, 2015, No. 6, pp. 871–890, doi: 10.1109/JPROC.2015.2404941.

[6] KRAULEDAT, M.—SCHRÖDER, M.—BLANKERTZ, B.—MÜLLER, K. R.: Reducing Calibration Time for Brain-Computer Interfaces: A Clustering Approach. In: Schölkopf, B., Platt, J., Hoffman, T. (Eds.): Advances in Neural Information Processing Systems 19 (NIPS 2006). MIT Press, 2006, pp. 753–760, `https://proceedings.neurips.cc/paper_files/paper/2006/file/275d7fb2fd45098ad5c3ece2ed4a2824-Paper.pdf`.

[7] AZAB, A. M.—MIHAYLOVA, L.—ANG, K. K.—ARVANEH, M.: Weighted Transfer Learning for Improving Motor Imagery-Based Brain-Computer Interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 27, 2019, No. 7, pp. 1352–1359, doi: 10.1109/TNSRE.2019.2923315.

[8] DAI, M.—ZHENG, D.—LIU, S.—ZHANG, P.: Transfer Kernel Common Spatial Patterns for Motor Imagery Brain-Computer Interface Classification. Computational and Mathematical Methods in Medicine, Vol. 2018, 2018, Art. No. 9871603, doi: 10.1155/2018/9871603.

[9] SAMEK, W.—MEINECKE, F. C.—MÜLLER, K. R.: Transferring Subspaces Between Subjects in Brain-Computer Interfacing. IEEE Transactions on Biomedical Engineering, Vol. 60, 2013, No. 8, pp. 2289–2298, doi: 10.1109/TBME.2013.2253608.

[10] JAYARAM, V.—ALAMGIR, M.—ALTUN, Y.—SCHOLKOPF, B.—GROSSE-WENTRUP, M.: Transfer Learning in Brain-Computer Interfaces. IEEE Computational Intelligence Magazine, Vol. 11, 2016, No. 1, pp. 20–31, doi: 10.1109/MCI.2015.2501545.

[11] SAMEK, W.—KAWANABE, M.—MÜLLER, K. R.: Divergence-Based Framework for Common Spatial Patterns Algorithms. IEEE Reviews in Biomedical Engineering, Vol. 7, 2014, pp. 50–72, doi: 10.1109/RBME.2013.2290621.

[12] ZANINI, P.—CONGEDO, M.—JUTTEN, C.—SAID, S.—BERTHOUMIEU, Y.: Transfer Learning: A Riemannian Geometry Framework with Applications to Brain-Computer Interfaces. IEEE Transactions on Biomedical Engineering, Vol. 65, 2018, No. 5, pp. 1107–1116, doi: 10.1109/TBME.2017.2742541.

[13] YAIR, O.—BEN-CHEN, M.—TALMON, R.: Parallel Transport on the Cone Manifold of SPD Matrices for Domain Adaptation. IEEE Transactions on Signal Processing, Vol. 67, 2019, No. 7, pp. 1797–1811, doi: 10.1109/TSP.2019.2894801.

[14] WEI, C. S.—KOIKE-AKINO, T.—WANG, Y.: Spatial Component-Wise Convolutional Network (SCCNet) for Motor-Imagery EEG Classification. 2019 9$^{\text{th}}$ International IEEE/EMBS Conference on Neural Engineering (NER), 2019, pp. 328–331, doi: 10.1109/NER.2019.8716937.

[15] WU, D.—XU, Y.—LU, B. L.: Transfer Learning for EEG-Based Brain-Computer Interfaces: A Review of Progress Made Since 2016. IEEE Transactions on Cognitive and Developmental Systems, Vol. 14, 2022, No. 1, pp. 4–19, doi:

10.1109/TCDS.2020.3007453.

[16] Schirrmeister, R. T.—Springenberg, J. T.—Fiederer, L. D. J.—Glasstetter, M.—Eggensperger, K.—Tangermann, M.—Hutter, F.—Burgard, W.—Ball, T.: Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization. Human Brain Mapping, Vol. 38, 2017, No. 11, pp. 5391–5420, doi: 10.1002/hbm.23730.

[17] Lawhern, V. J.—Solon, A. J.—Waytowich, N. R.—Gordon, S. M.—Hung, C. P.—Lance, B. J.: EEGNet: A Compact Convolutional Neural Network for EEG-Based Brain-Computer Interfaces. Journal of Neural Engineering, Vol. 15, 2018, No. 5, Art. No. 056013, doi: 10.1088/1741-2552/aace8c.

[18] Finn, C.—Abbeel, P.—Levine, S.: Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In: Precup, D., Teh, Y. W. (Eds.): Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research (PMLR), Vol. 70, 2017, pp. 1126–1135, `http://proceedings.mlr.press/v70/finn17a/finn17a.pdf`.

[19] Sun, Q.—Liu, Y.—Chua, T. S.—Schiele, B.: Meta-Transfer Learning for Few-Shot Learning. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 403–412, doi: 10.1109/CVPR.2019.00049.

[20] Chen, W. Y.—Liu, Y. C.—Kira, Z.—Wang, Y. C. F.—Huang, J. B.: A Closer Look at Few-Shot Classification. International Conference on Learning Representations (ICLR 2019), 2019, doi: 10.48550/arXiv.1904.04232.

[21] Chen, Y.—Liu, Z.—Xu, H.—Darrell, T.—Wang, X.: Meta-Baseline: Exploring Simple Meta-Learning for Few-Shot Learning. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 9042–9051, doi: 10.1109/ICCV48922.2021.00893.

[22] Li, D.—Ortega, P.—Wei, X.—Faisal, A.: Model-Agnostic Meta-Learning for EEG Motor Imagery Decoding in Brain-Computer-Interfacing. 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER), 2021, pp. 527–530, doi: 10.1109/NER49283.2021.9441077.

[23] An, S.—Kim, S.—Chikontwe, P.—Park, S. H.: Few-Shot Relation Learning with Attention for EEG-Based Motor Imagery Classification. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020, pp. 10933–10938, doi: 10.1109/IROS45743.2020.9340933.

[24] Ravi, S.—Larochelle, H.: Optimization as a Model for Few-Shot Learning. International Conference on Learning Representations (ICLR 2017), 2017, `https://openreview.net/forum?id=rJY0-Kcll`.

[25] Munkhdalai, T.—Yu, H.: Meta Networks. In: Precup, D., Teh, Y. W. (Eds.): Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research (PMLR), Vol. 70, 2017, pp. 2554–2563, `http://proceedings.mlr.press/v70/munkhdalai17a/munkhdalai17a.pdf`.

[26] Santoro, A.—Bartunov, S.—Botvinick, M.—Wierstra, D.—Lillicrap, T.: Meta-Learning with Memory-Augmented Neural Networks. In: Balcan, M. F., Weinberger, K. Q. (Eds.): Proceedings of the 33rd International Conference on Machine Learning. Proceedings of Machine Learning Research (PMLR),

Vol. 48, 2016, pp. 1842–1850, http://proceedings.mlr.press/v48/santoro16.pdf.

[27] Mishra, N.—Rohaninejad, M.—Chen, X.—Abbeel, P.: A Simple Neural Attentive Meta-Learner. 6th International Conference on Learning Representations (ICLR 2018), 2018, doi: 10.48550/arXiv.1707.03141.

[28] Vinyals, O.—Blundell, C.—Lillicrap, T.—Kavukcuoglu, K.—Wierstra, D.: Matching Networks for One Shot Learning. In: Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., Garnett, R. (Eds.): Advances in Neural Information Processing Systems 29 (NIPS 2016). Curran Associates, Inc., 2016, pp. 3630–3638, https://proceedings.neurips.cc/paper_files/paper/2016/file/90e1357833654983612fb05e3ec9148c-Paper.pdf.

[29] Snell, J.—Swersky, K.—Zemel, R.: Prototypical Networks for Few-Shot Learning. In: Guyon, I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.): Advances in Neural Information Processing Systems 30 (NIPS 2017). Curran Associates, Inc., 2017, pp. 4077–4087, https://proceedings.neurips.cc/paper_files/paper/2017/file/cb8da6767461f2812ae4290eac7cbc42-Paper.pdf.

[30] Sung, F.—Yang, Y.—Zhang, L.—Xiang, T.—Torr, P. H. S.—Hospedales, T. M.: Learning to Compare: Relation Network for Few-Shot Learning. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 1199–1208, doi: 10.1109/CVPR.2018.00131.

[31] Tynes, M.—Parsapoor, M.: Meta-Learning on Spectral Images of Electroencephalogram of Schizophrenics. CoRR, 2021, doi: 10.48550/arXiv.2101.12208.

[32] Zhu, Y.—Saqib, M.—Ham, E.—Belhareth, S.—Hoffman, R.—Wang, M. D.: Mitigating Patient-to-Patient Variation in EEG Seizure Detection Using Meta Transfer Learning. 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), 2020, pp. 548–555, doi: 10.1109/BIBE50027.2020.00095.

[33] Duan, T.—Shaikh, M. A.—Chauhan, M.—Chu, J.—Srihari, R. K.—Pathak, A.—Srihari, S. N.: Meta Learn on Constrained Transfer Learning for Low Resource Cross Subject EEG Classification. IEEE Access, Vol. 8, 2020, pp. 224791–224802, doi: 10.1109/ACCESS.2020.3045225.

[34] Duan, T.—Chauhan, M.—Shaikh, M. A.—Chu, J.—Srihari, S.: Ultra Efficient Transfer Learning with Meta Update for Cross Subject EEG Classification. CoRR, 2020, doi: 10.48550/arXiv.2003.06113.

[35] Erhan, D.—Courville, A.—Bengio, Y.—Vincent, P.: Why Does Unsupervised Pre-Training Help Deep Learning? In: Teh, Y. W., Titterington, M. (Eds.): Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research (PMLR), Vol. 9, 2010, pp. 201–208, http://proceedings.mlr.press/v9/erhan10a/erhan10a.pdf.

[36] Ang, K. K.—Chin, Z. Y.—Zhang, H.—Guan, C.: Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008, pp. 2390–2397, doi: 10.1109/IJCNN.2008.4634130.

[37] Ioffe, S.—Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: Bach, F., Blei, D. (Eds.): Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research (PMLR), Vol. 37, 2015, pp. 448–456, http://proceedings.mlr.press/v37/ioffe15.pdf.

[38] Srivastava, N.—Hinton, G.—Krizhevsky, A.—Sutskever, I.—Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, Vol. 15, 2014, No. 56, pp. 1929–1958, http://jmlr.org/papers/v15/srivastava14a.html.

[39] Brunner, C.—Leeb, R.—Müller-Putz, G.—Schlögl, A.—Pfurtscheller, G.: BCI Competition 2008 – Graz Data Set A. 2008, doi: 10.21227/katb-zv89.

[40] Lu, H.—Plataniotis, K. N.—Venetsanopoulos, A. N.: Regularized Common Spatial Patterns with Generic Learning for EEG Signal Classification. 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 6599–6602, doi: 10.1109/IEMBS.2009.5332554.

Jian Yu is an Engineer with the Department of Laboratory Center, School of Electronic and Information Engineering, Tongji University, Shanghai, China. He received his Ph.D. degree in circuits and systems from the South China University of Technology, Guangzhou, China, in 2010. His research interests include Internet of Things, machine learning, pattern recognition, and embedded systems.

Lili Duan received her B.Sc. degree in software engineering from the Nantong University, China, in 2019. Now, she is a graduate student in the Department of Computer Science and Technology, Tongji University, Shanghai, China. Her research interests include machine learning and brain-computer interface.

Hongfei Ji is Associate Professor with the Department of Computer Science and Technology, Tongji University, Shanghai, China. He received his Ph.D. degree in software engineering from the Tongji University, Shanghai, China, in 2015. His research interests include machine learning, pattern recognition, brain-computer interface, and cognitive neuroscience.

Jie Li is Associate Professor with the Department of Computer Science and Technology, Tongji University, Shanghai, China. She received her Ph.D. degree in computer science from the Shanghai Jiao Tong University, Shanghai, China, in 2010. Her research interests include machine learning, brain-computer interface, brain-like computation, and cognitive neuroscience.

Zilong Pang is Lecturer with the School of Software, Henan University, Kaifeng, China. He received his Master's degree in applied mathematics in computer science and his Ph.D. from the School of Electronic and Information Engineering, Tongji University, Shanghai, China. His research directions mainly include brain-computer interface, artificial intelligence in medical imaging, machine vision.