Flashlight-Net: a modular convolutional neural network for motor imagery EEG classification

Weidong Dang, Dongmei Lv, Mengxiao Tang, Xinlin Sun,Yong Liu, Zhongke Gao, *Senior Member, IEEE,* and Celso Grebogi

***Abstract*—Brain-computer interface (BCI) establishes an interactive platform by translating brain activity patterns into commands of external devices. BCIs, especially motor imagery (MI)-based BCIs, have injected new vitality into the development of rehabilitation medicine and many other fields. In this work, one convolutional neural network, named as Flashlight-Net model, is proposed for multi-class MI classification. Flashlight-Net model adopts modular design, in which channel fusion module and time domain module ensure the directions of feature extraction, while feature pool module reduces the loss of effective information. Given the multi-frequency nature of the brain, we combine three frequency bands and construct an ensemble Flashlight-Net model. During the model training, by means of transfer learning, pre-training and fine-tuning processes are designed to integrate training samples from multiple subjects. The experimental results on publicly available BCI Competition IV-2a dataset show that the proposed model can achieve good results on all nine subjects, with an average classification accuracy of 81.23% for four classes. All these demonstrate that the proposed Flashlight-Net model can effectively decode multi-channel and multi-class MI signals.**

***Index Terms*—Motor imagery, convolutional neural network, transfer learning, EEG signals, brain-computer interface.**

I. Introduction

BRAIN-computer interface (BCI) builds a pathway between the human brain and external devices by recording the electroencephalogram (EEG) signals and converting them into equipment control instructions [1,2]. Efficient BCI systems have broad application prospects in the fields of smart home and rehabilitation medicine, among others [3,4]. Until now, some commonly used EEG-based BCI paradigms have been developed, such as steady-state visual evoked potentials (SSVEP) [5], which needs visual stimulation, and P300, whose performance depends on the probability of related events [6]. Motor imagery (MI) [7,8] is distinct from these two paradigms. In an MI-based BCI system, EEG signals are evoked by imagining the movement of a specific body part. No external stimuli are needed, which allows MI-based BCI systems to be applicable in a wider range of scenarios. Event related desynchronization (ERD) and event related synchronization (ERS) phenomena can be observed [9] during MI, manifesting as a rapid energy dropping or rising in μ and β bands. These form the basis for distinguishing EEG signals from different MI tasks. However, existing studies [10,11] have confirmed that MI-related EEG signals exhibit obvious nonlinearity and low signal-to-noise ratio (SNR), which makes the classification very difficult.

In the past few decades, a lot of research attempted to extract features manually and then train machine learning-based classifiers for recognizing MI signals. Among them, a popular spatial feature extraction method called common spatial pattern (CSP) [12] was proposed. The basic idea of CSP is to find a spatial filter that maximizes the variations between two classes by solving a matrix decomposition problem. Due to its stable performance, some improved versions of CSP have emerged. For example, Lemm et al. [13] proposed the common space spectral patterns (CSSP), which outperformed the CSP in terms of classification accuracy as well as generalization ability. Via optimizing subject-specific frequency bands, Ang et al. [14] developed the filter bank CSP (FBCSP), which confirmed the importance of frequency information in MI classification. Bashashati et al. [15] used Bayesian optimization to determine the best parameters for FBCSP, and then got an average accuracy of 68.13% on BCI Competition IV-2a dataset. After this, composite FBCSP [16] was proposed, which used a single spatial filter instead of multiple spatial filters. The results showed that composite FBCSP can achieve an average accuracy of 59.6% on BCI Competition IV-2b dataset. Lotte et al. [17] designed a regularized CSP (RCSP), which outperformed CSP by nearly 10% in median classification accuracy, based on BCI Competition IV-2a dataset. Aghaei et al. [18] developed the separable common spatio-spectral patterns (SCSSP), which considered the spectral characteristics of EEG signals and required significantly lower computations than CSP. Moreover, Shahtalebi et al. [19] used error correction output coding (ECOC) classifier to solve the multi-class MI classification problem. The ECOC classifier was combined with a Bayesian framework to calculate the optimized spatio-spectral filter, aiming at extracting the most discriminative feature sets of different classes. Abenna et al. [20] combined common spatial pattern (CSP), decision tree (DT), particle swarm optimization (PSO), and light gradient boosting machine (LGBM) to achieve MI decoding. Jindal

Manuscript received ***; accepted ***. This work was supported in part by the National Natural Science Foundation of China under Grants Nos. 61922062 and 61873181. (Corresponding author: Zhongke Gao)

W. Dang, D. Lv, L. Sun, and Z. Gao, are with the School of Electrical and Information Engineering, Tianjin University, Tianjin, 300072, China (e-mails: weidongdang@tju.edu.cn; zhongkegao@tju.edu.cn; dongmeilv@tju.edu.cn; xinlinsun@tju.edu.cn).

M. Tang is with the School of Electrical and Electronic Engineering, Nanyang Technology University, Singapore, 50 Nanyang Ave, 639798, Singapore (e-mail: MTANG007@e.ntu.edu.sg).

Y. Liu is with the Xinqiao Hospital, Army Medical University, Chongqing, 400037, China (e-mail: lzsht@aliyun.com).

C. Grebogi is with the Institute for Complex Systems and Mathematical Biology, King's College, University of Aberdeen, Aberdeen AB24 3UE, UK (e-mail: grebogi@abdn.ac.uk).

et al. [21] proposed a channel selection and features optimization methodology for MI classification. Despite numerous attempts, these traditional methods are difficult to meet the requirements of BCI systems for high classification accuracy. Novel algorithms and frameworks need to be developed, which are expected to be applicable to multiple subjects at the same time.

Recent years, with the successful applications of deep learning in diverse fields, how to effectively integrate deep learning into MI classification framework has become a timely research topic. Deep learning shows end-to-end structure, and its feature extraction and classification stages are jointly optimized. So far, some deep learning architectures, such as convolutional neural network (CNN) and their improved versions, have been developed for classifying MI-related EEG signals [22-24]. Among them, some of the research use traditional methods to extract features and then input them into the CNN frameworks for further feature optimization and classification. For example, Deng et al. [22] used temporary constrained sparse group Lasso (TCSGL) to improve the EEGNet and enhance the decoding performance of multi-class MI. Olivas-Padilla et al. [25] developed discriminative FBCSP to extract the effective features, and then fed them into a CNN model for MI classification. Similar framework can also be found in Ref. [26]. Such a framework can well inherit the advantages of traditional methods, but it limits the ability of CNNs to learn robust features. Therefore, some researchers have tried to implement MI classification directly through CNN-based deep learning architectures. For instance, Mane et al. [27] proposed a FBCNet model to decode 2-class MI tasks, which used a CNN layer to extract spatial features and a variance layer to extract temporal features. Schirrmeister et al. [28] developed three CNN models including shallow-layer CNN, deep-layer CNN and DenseNet, to conduct the classification of 4-class MI signals. Amin et al. [29] considered the subject-specific frequency bands, and designed one multi-layer CNN architecture for 4-class MI classification.

In addition, it should be mentioned that MI experiments are exhausting. For a single subject, the number of samples is often very limited, which has been plaguing MI research. As an effective cross-dataset problem solution, transfer learning [30] provides good ideas for solving this problem. Some researchers have attempted to apply transfer learning to MI classification. For example, Hang et al. [31] developed a domain adaptation network to transfer the data representation from one subject to another (i.e., one-to-one), and tested on the MI classification tasks. Subsequently, Wei et al. [32] designed a multi-subject transfer learning network. They first merged the training sets of five subjects together, and then train the pre-processing model. Liang et al. [33] used transfer learning to calibrate EEG features in MI classification tasks. They selected the suitable source subjects via a Riemannian geometry alignment algorithm, and then utilized the dataset from source subjects and a few new training data from current user to train the MI classifier. Similar framework can be also found in Ref. [34].

Motivated by the above-described background and progress, this work establishes an end-to-end MI classification framework based on CNN, named as Flashlight-Net model. Firstly, a channel fusion module is designed to achieve the integration of multiple EEG channels. Then, a time domain module based on dilated convolution is constructed, which allows effectively extracting time-domain fluctuation information at different scales. After these two modules, a feature pool module is set to transform and save the features, which are finally input into the classification module for decoding MI tasks. Particularly, considering the frequency-dependent properties, we learn the identifiable MI features from three frequency bands through an ensemble Flashlight-Net model. Moreover, multiple losses and transfer learning are set up to further speed up the model training and improve the model performance. Experimental results on BCI Competition IV-2a dataset show that the proposed method can achieve good results on all nine subjects, with high accuracy and good stability.

The remaining of this work is organized as follows: Section II introduces the detailed design of the proposed model. Section III introduces BCI Competition IV-2a dataset. Section IV presents the training curves and main experimental results.

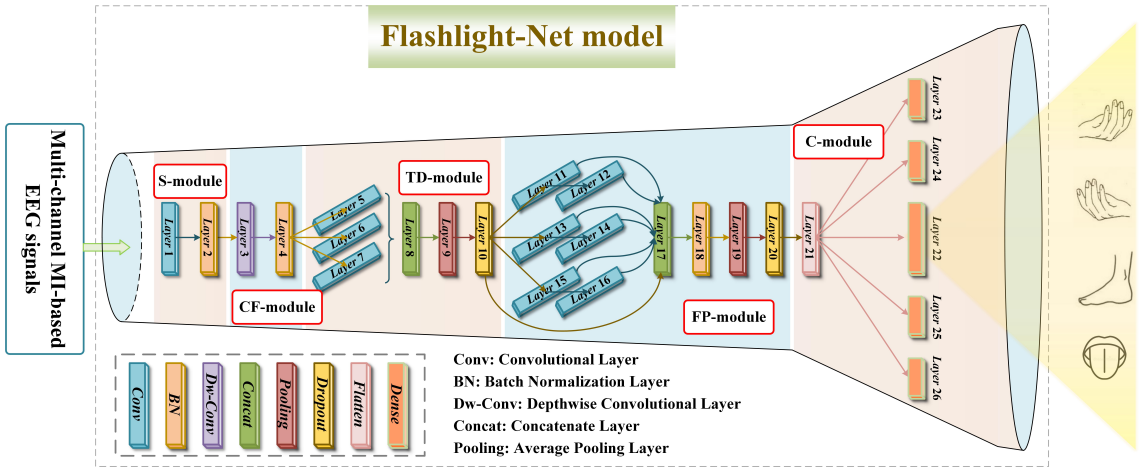


Fig. 1. The proposed Flashlight-Net model for motor imagery classification. Using blocks of different colors to represent different layers, e.g., blue block represents convolutional layer.

Section V discusses the performance of the proposed model, comparing with some variants and some existing studies. Finally, this work is concluded in Section VI.

II. Method

A. Flashlight-Net model

In this work, via adopting modular design, a CNN-based MI classification model is proposed. Multi-channel MI-based EEG signals are used as the input. The model structure is shown in Fig. 1. This model has one input, one classifier output and four auxiliary outputs, thus making it a single-input-multiple-output (SIMO) architecture. Five plug-and-play modules are carefully designed, including smoothing module (S-module, Layers 1-2), channel fusion module (CF-module, Layers 3-4), time domain module (TD-module, Layers 5-10), feature pool module (FP-module, Layers 11-20), and classification module (C-module, Layers 21-26). When using this model, the process of decoding MI signals is like lighting a dark area with a flashlight. So, we name it Flashlight-Net model in this work.

Next, we will illustrate the specific structure of Flashlight-Net model. The detailed parameters are displayed in Table I. In the following description, Layer 1 is simplified to $L1$, and others similarly.

1) Smoothing module (S-module: Layers 1-2): At the beginning of Flashlight-Net model, the input multi-channel EEG signals are smoothed over time to reduce noise by using one temporal convolutional layer ($L1$). We use the matrix $X \in R^{C*T}$ to represent one input sample, where C denotes channels (or recorded electrodes), and T denotes sampling points. In this work, C=22 and T=1000. In $L1$, K_1=10 convolution kernels are designed, with the size of 1*25, corresponding to a sampling time of 0.1 s. For all convolutional layers in this work, 'same padding' operation is used. $L2$ is a batch normalization (BN) layer.

2) Channel fusion module (CF-module: Layers 3-4): After S-module, a CF-module based on depthwise convolutional (Dw-Conv) layer [35] is designed. Multi-channel EEG signals come from different brain electrodes or brain regions. CF-module is designed to integrate these channel-related characteristics. $L3$ is a Dw-Conv layer. Different from general convolutional layer, Dw-Conv layer can reduce the number of parameters due to its sparse connectivity. Dw-Conv layer has two important parameters: one is the kernel, and the other is the depth parameter D. The kernel size in $L3$ is $C * 1 * K_1$. Namely, the features from all C channels are simultaneously integrated. Such a design avoids the impact of channel locations. Particularly, each $C * 1$ kernel slice only performs convolution operation with the corresponding feature map of $L2$. This can be described by the following formula:

$$\sigma = cat(conv(\omega_1^3, \sigma_1^2), conv(\omega_2^3, \sigma_2^2), ..., conv(\omega_{K_1}^3, \sigma_{K_1}^2)), \tag{1}$$

where $cat(\cdot)$ is the concatenate function, σ_2^2 means the second output feature map of $L2$, ω_2^3 represents the weight matrix of the second C*1 kernel slice in $L3$, and others are similar. Then, the above convolution process is performed D times and the results are stacked together. So the output size of $L3$

TABLE I
The parameter values of the proposed Flashlight-Net model.

Module	Layers	Output size	Details	Number of parameters	Number of MAC operations
	Input	22,1000			
S-module	*L*1:Conv	22,1000,10	kernel 1*25, num 10	260	11440000
	*L*2:BN	22,1000,10		20	880000
CF-module	*L*3:Dw-Conv	1,1000,80	kernel 22*1*10, *D*=8	1020	2040
	*L*4:BN	1,1000,80	Relu	0	160000
TD-module	*L*5:Conv	1,1000,80	kernel 1*125, num 80, *r*5=1, Relu	800000	1600160000
	*L*6:Conv	1,1000,80	kernel 1*125, num 80, *r*6=2, Relu	800000	1600160000
	*L*7:Conv	1,1000,80	kernel 1*125, num 80, *r*7=3, Relu	800000	1600160000
	*L*8:Concat	1,1000,240		0	0
	*L*9:Pooling	1,200,240	pooling 1*5	0	480000
	*L*10:Dropout	1,200,240	rate 0.5	0	0
FP-module	*L*11:Conv	1,200,80	kernel 1*10, num 80, *r*11=1, Relu	192000	77216160
	*L*12:Conv	1,200,80	kernel 1*10, num 80, *r*12=3, Relu	64000	25632000
	*L*13:Conv	1,200,80	kernel 1*20, num 80, *r*13=1, Relu	384000	154400000
	*L*14:Conv	1,200,80	kernel 1*20, num 80, *r*14=3, Relu	128000	51232000
	*L*15:Conv	1,200,80	kernel 1*30, num 80, *r*15=1, Relu	576000	231584160
	*L*16:Conv	1,200,80	kernel 1*30, num 80, *r*16=3, Relu	192000	76832000
	*L*17:Concat	1,200,720		0	0
	*L*18:BN	1,200,720		0	0
	*L*19:Pooling	1,90,200	transform dimension, pooling 8*1	0	288000
	*L*20:Dropout	1,90,200	rate 0.5	0	0
C-module	*L*21:Flatten	18000		0	0
	*L*22:Dense	4	Softmax	72004	144000
	*L*23:Dense	2	Softmax	36002	72000
	*L*24:Dense	2	Softmax	36002	72000
	*L*25:Dense	2	Softmax	36002	72000
	*L*26:Dense	2	Softmax	36002	72000

* 'Conv', 'BN', and 'Concat' mean convolutional layer, batch normalization layer, and concatenate layer, respectively.

is $1*T*(K_1*D)$. D is set as 8 in this work. $L4$ is a BN layer, whose outputs are processed via rectified linear unit (ReLU). Each feature map (with size $1*T$) obtained via CF-module represents a weighted fusion of multi-channel EEG signals.

3) Time domain module (TD-module: Layers 5-10): In TD-module, time-domain fluctuation information is the focus of attention. For this purpose, a dilated convolution group is designed, which consists of three dilated convolutional layers $L5$, $L6$, and $L7$. Dilated convolution has been shown to be suitable for the extraction of temporal features [36-37]. These can be described by the following formula:

$$\begin{cases} \sigma^5 = f(b^5_{k_5} + conv(\omega^5_{k_5}, \sigma^4, r_5)) \\ \sigma^6 = f(b^6_{k_6} + conv(\omega^6_{k_6}, \sigma^4, r_6)) \\ \sigma^7 = f(b^7_{k_7} + conv(\omega^7_{k_7}, \sigma^4, r_7)), \end{cases} \tag{2}$$

where $f(\cdot)$ is the activation function ReLU. The outputs of $L4$ (σ^4) are used as the input. $r5$, $r6$, and $r7$ are the dilation rates, and they are 1, 2, and 3, respectively. In each layer, $K_1 * D$ convolution kernels are designed, with the size of 1*125, corresponding to a sampling time of 0.5 s. Considering the dilation rates, Layers 5-7 can respectively learn the features in the scales of 0.5 s, 1 s and 1.5 s. All outputs of the dilated convolution group are concatenated together ($L8$). An average pooling layer ($L9$) with size of 1*5 is set to smooth the global fluctuation characteristics. The output size of $L9$ is $1*(T/5)*(3*K_1*D)$. At the end of TD-module, a dropout layer ($L10$) is arranged, which randomly removes the output neurons of $L9$ with a rate 0.5.

4) Feature pool module (FP-module: Layers 11-20): In order to make full use of the features learned by CF-module and TD-module, FP-module is constructed, including Layers 11-20. Layers 11-12 can be described by the following formula:

$$\begin{cases} \sigma^{11}_{k_{11}} = f(b^{11}_{k_{11}} + conv(\omega^{11}_{k_{11}}, \sigma^{10}, r_{11})) \\ \sigma^{12}_{k_{12}} = f(b^{12}_{k_{12}} + conv(\omega^{12}_{k_{12}}, \sigma^{11}, r_{12})), \end{cases} \tag{3}$$

where dilation rates r_{11} and r_{12} are 1 and 3, respectively. In Layers 11-12, each has K_1*D convolution kernels, with the size of 1*10. Layers 13-14 and Layers 15-16 are similar to Layers 11-12. The difference is that the kernel size of Layers 13-14 is 1*20, and of Layers 15-16 is 1*30. Layers 11-16 are combined to achieve multi-scale feature extraction. Subsequently, the outputs of Layers 9 and 11-16 are concatenated together ($L17$). The output size of $L17$ is $1*(T/5)*(9*K_1*D)$. This design inherits the advantages of residual learning [38], and can integrate the features of different levels via shortcut connections. $L18$ is a BN layer. We do a dimensional transformation on the outputs of $L18$ (σ^{18}) to make the size change from $1*(T/5)*(9*K_1*D)$ to $1*(9*K_1*D)*(T/5)$. $L19$ is an average pooling layer, whose output size is $1*(9*K_1)*(T/5)$. $L20$ is a dropout layer, with a rate 0.5.

5) Classification module (C-module: Layers 21-26): The outputs of $L20$ are then flatten in $L21$ and used as input to C-module. Five dense layers (i.e., Layers 22-26) are parallelly arranged at the end of Flashlight-Net model, with softmax activation function. Among them, $L22$ has four neurons, corresponding to a four-class MI classification task. And each of Layers 23-26 has two neurons, corresponding to a two-class MI classification task via the OvR (One $vs.$ Rest) strategy. Four two-class MI classification tasks include: left hand $vs.$ others, right hand $vs.$ others, feet $vs.$ others, and tongue $vs.$ others. That is, C-module can produce one main loss and four auxiliary losses during the training process.

B. Ensemble Flashlight-Net model

During MI tasks, ERS and ERD phenomena occur, which are closely related to μ and β bands. In this work, aiming at integrating the multi-band information, we consider three frequency bands, including 4∼38 Hz, 4∼14 Hz (μ band), and 14∼32 Hz (β band), and propose the ensemble Flashlight-Net model. Butterworth filter is used to decompose the frequency bands. The model structure is shown in Fig. 2. Ensemble Flashlight-Net model contains three branches, corresponding to three frequency bands, respectively. For each band, one Flashlight-Net model is used to learn the band-specific information. The outputs of $L22$ in three Flashlight-Net models are directly added together ($L27$). Therefore, ensemble Flashlight-Net model has thirteen outputs, which can produce one main loss and twelve auxiliary losses. Softmax activation function is used to conduct the classification.

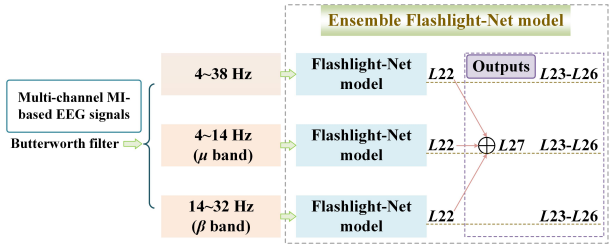


Fig. 2. The model structure of ensemble Flashlight-Net model.

C. Transfer learning strategy

In this work, we aim at achieving EEG-related MI decoding across sessions (or days). We design a transfer learning strategy to integrate training samples from multiple subjects, which allows to eliminate subject-specific interference and obtain effective features directly related to MI tasks. In detail, two processes are designed to train the proposed model, namely, pre-training and fine-tuning. During the pre-training process, the training sets of all nine subjects are merged together for the supervised learning. One subject-free model (i.e., pre-training model) is optimized and obtained. This pre-training model is not for a specific subject, but can provide initial parameters for all subjects in the subsequent fine-tuning. During the pre-training process, based on the Adam optimizer, all kinds of weights and biases are optimized by minimizing the cross-entropy loss function. Learning rate is set to 0.001, and fixed 200 epochs are conducted with batch size 128. During the fine-tuning process, based on the obtained pre-training model, only the training set of the current subject is used to re-optimize the model, and one subject-specific model (i.e., fine-tuning model) is gotten. Fine-tuning model is then directly applied to the testing set of the current subject and MI classification accuracy is obtained. In fine-tuning process, considering the

small sample sizes, we gradually decrease the learning rates and increase the number of epochs. Learning rates are 0.0001, 0.00001, and 0.000001, while the number of epochs is 20, 40, and 60, respectively. Batch size is set as 16. Such a data transfer strategy enlarges the samples available for training by 9 times, and is suitable for actual MI-based BCI system applications.

III. Dataset Description

Publicly available BCI Competition IV-2a dataset [14,15,18] is employed in this work, which has been regarded as a benchmark for 4-class MI research. It has 9 subjects and considers 4 MI tasks, i.e., imagining the movements of left hand, right hand, feet, and tongue. For each subject, the experiments of 2 sessions are conducted on 2 separate days. Each session has 288 trials, with 72 trials for each of the 4 MI tasks. The paradigm of an experimental trial is illustrated in Fig. 3. EEG signals from 22 brain electrodes are recorded. The signals are sampled with 250 Hz and bandpass-filtered between 0.5 Hz and 100 Hz. An additional 50 Hz notch filter is enabled to suppress line noise.

In this work, each trial contributes a sample $X \in \mathbb{R}^{C\times T}$, where $C = 22$ is the number of channels (or electrodes) and $T = 1000$ is the number of data points (from t = 2 s to t = 6 s). Like other studies in this dataset, the first session data are used as the training set, and the second session data are used as the testing set. In this study, we do not arrange a validation set. Because the goal of this study is to achieve MI decoding across sessions.

1) Taking part of the samples from the first session (i.e., training set) as validation set cannot improve the inter-session decoding ability of the model.

2) And it is unreasonable to take part of the samples from the second session (i.e., testing set) as validation set, which would lead to information leakage.

To sum up, each subject has 288 samples as the training set and another 288 samples as the testing set. The shape of one sample is 22*1000.

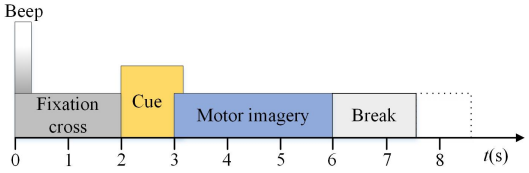


Fig. 3. Timing scheme of an experimental trial.

IV. Results

A. Model training and learning curves

Following the transfer learning strategy in Section II-C, we conduct the model training and plot the learning curves, as shown in Figs. 4-5. During the fine-tuning process, Sub.3, Sub.5 and Sub.9 are selected as examples because they have the highest, lowest and medium MI classification accuracy, respectively. We do not arrange a validation set in this work, so the testing set is used to plot the curves, which is only used to illustrate the characteristics of the proposed model. Specifically, we choose the model corresponding to the last epoch as the optimized model, which is not influenced by the testing set. For example, in Fig. 5(b), 56.94% is used as the testing accuracy instead of 61.81%. Such a training scheme is closer to the actual application, where testing samples are not visible to the training process.

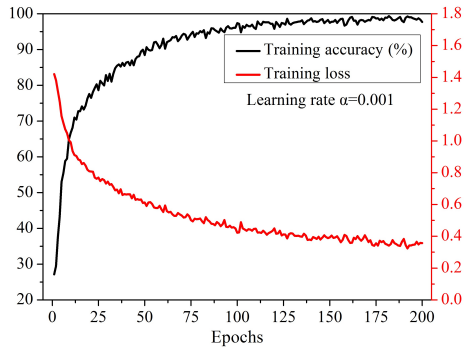


Fig. 4. Training accuracy and training loss in the pre-training process.

As can be seen, during the pre-training process (Fig. 4), as the epoch increases, the model becomes more and more capable of fitting the training samples, and the training accuracy becomes higher and higher, and gradually approaches 100%. During the fine-tuning process (Fig. 5), as the epoch increases, the training accuracy stabilizes at 100%, while the testing accuracy gradually increases. In particular, the change in testing accuracy is closely related to the learning rate. As the learning rate becomes smaller, the testing accuracy becomes more stable. These phenomena remain consistent across the learning curves of all 3 subjects, which indicates that our model and training strategy are highly adaptive for multiple subjects.

B. Four-class MI classification

We run the ensemble Flashlight-Net model 10 times. In each time, we randomly shuffle the training samples of all 9 subjects and randomly initialize the model. Then, the model is optimized through pre-training process to get one pre-training model. On the basis of this pre-trained model, fine-tuning process is performed on each of the 9 subjects. Accordingly, 9 fine-tuning models are obtained, as well as 9 testing accuracies, corresponding to 9 subjects. For each subject, we average the accuracies of 10 times as the final result, as shown in Fig. 6. The final average accuracy of ensemble Flashlight-Net model on all 9 subjects reaches 81.23%.

To further demonstrate the effectiveness of the transfer learning strategy, we also present the results without transfer learning (i.e., using only the fine-tuning process) in Fig. 6. The average accuracy is only 53.78%, which is significantly lower than that when using transfer learning. In such a transfer learning strategy, the training samples of the current subject only account for 1/9 of the total training samples, but play a huge role in model training. That is, using a small number of training samples from the current subject helps to better

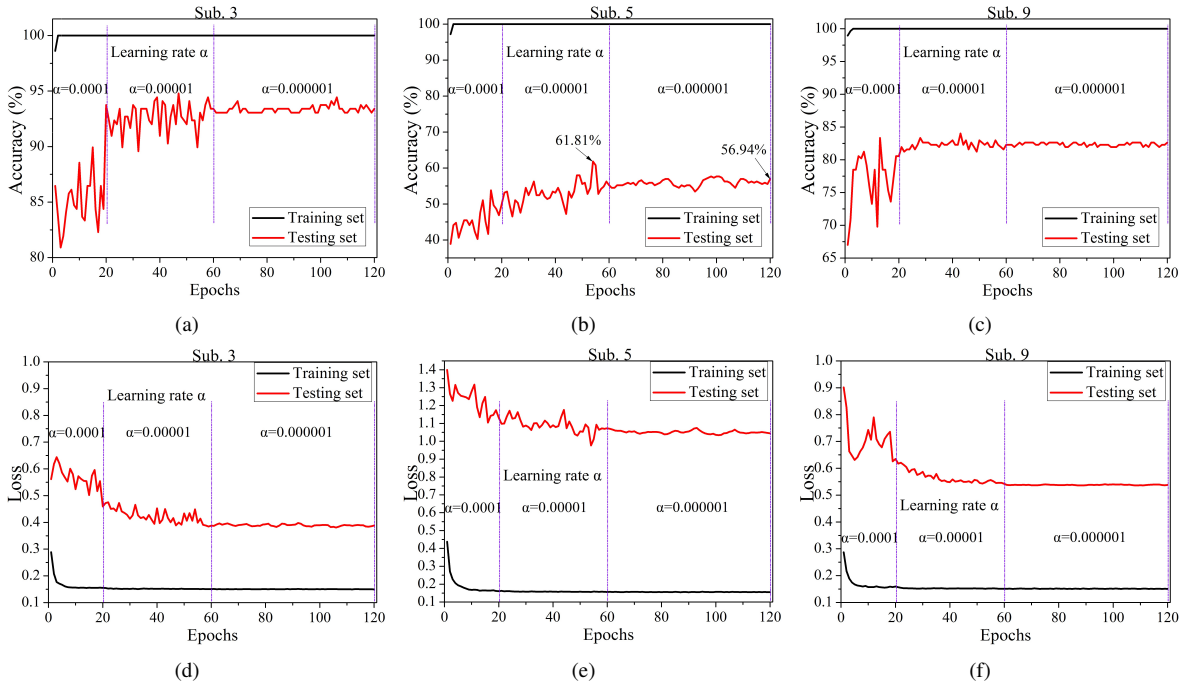


Fig. 5. Training accuracies of (a) Sub.3, (b) Sub.5, (c) Sub.9, and training losses of (d) Sub.3, (e) Sub.5 and (f) Sub.9 in the fine-tuning process.

optimize the model, compared to not using training samples from the current subject. In fact, this is very relevant to the actual application scenario, where new users often need to do some training before they can use the MI-related BCIs.

In this work, we number each layer of the proposed Flashlight-Net model to present the model structure clearly, as shown in Fig. 1. There are 26 layers in total. In fact, we adopt parallel structures many times and the effective depth of the model is only six. That is ($L1$), ($L3$), ($L5$-7), ($L11$,13,15), ($L12$,14,16), ($L22$-26). The other layers are mainly some pooling layers, BN layers, dropout layers, concatenate layers and flatten layer, which have no trainable parameters. That is, the proposed model has a very clear structure. We have provided the number of parameters and the number of MAC operations in Table I. Total number of parameters is 4.15 M, and total number of MAC operations is 5.44 G. Such a model size is very common in this field. In the subsequent studies, we will try to introduce knowledge distillation to further simplify the model, thus making it more suitable for actual applications.

V. Discussion

A. Model structure analysis

Flashlight-Net model has five modules, and each of them has a specific structure and plays a key role in MI classification. In detail, S-module with one temporal convolutional layer is designed to smooth the MI signals and reduce noise. When performing MI tasks, different brain regions are activated. CF-module fully considers the brain region association information, and performs a diverse and adaptive weighted fusion of multi-channel MI signals. In the TD-module, considering the short and long-term operating characteristics of brain, Layers 5-7 adopt parallel design and learn the time domain features corresponding to the scales of 0.5 s, 1 s and 1.5 s, respectively. Through CF-module and TD-module, Flashlight-Net model carries out targeted feature learning, from two angles of channel and time, respectively. Then, using multi-scale design and residual learning, FP-module can make full use of these acquired features, reducing the loss of effective

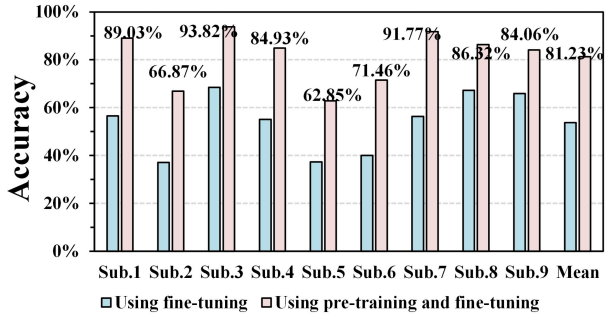


Fig. 6. Average accuracies of the proposed ensemble Flashlight-Net model.

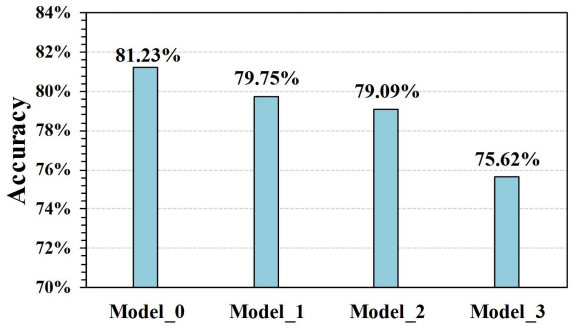


Fig. 7. Accuracies of the ensemble Flashlight-Net model (Model_0) and three comparison models (Model_1-3).

information. C-module is arranged at the end of Flashlight-Net model. Multi-loss design of C-module can provide a larger gradient and a better direction for model parameter optimization.

Based on the ensemble Flashlight-Net model, we conduct some comparative experiments. The details of the comparison models are briefly described as follows:

Model_0: Ensemble Flashlight-Net model.
Model_1: It removes the $L6$ and $L7$ in the TD-module, and the number of convolution kernels in $L5$ is set as $K_1 * D * 3 = 240$.
Model_2: It removes the $L11$, $L12$, $L15$ and $L16$ in the FP-module, and the numbers of convolution kernels in $L13$ and $L14$ are both set as $K_1 * D * 3 = 240$.
Model_3: It removes the $L23$, $L24$, $L25$ and $L26$ in the C-module.

As shown in Fig. 7, the average accuracies of these four models are 81.23%, 79.75%, 79.09%, and 75.62%, respectively. Compared with Model_1, TD-module in ensemble Flashlight-Net model (Model_0) considers both short and long-term operating characteristics of brain, and the accuracy is improved by 1.48%. Compared with Model_2, the multi-scale design of FP-module ensures the full use of features, and the accuracy is improved by 2.14%. In the C-module, different losses focus on different classification tasks. For example, the auxiliary loss after $L24$ corresponds to the classification error between the right hand and the other three MI tasks. By minimizing this loss, the model can effectively extract features that can be used to accurately identify the MI task of the right hand. Compared with Model_3, by adding auxiliary losses in C-module, the accuracy is improved by 5.61%.

B. Multi-band fusion

Ensemble Flashlight-Net model integrates the information from three frequency bands. Here, in order to better explain the contribution of each band, some comparison models are designed. The details of these models are briefly described as follows:

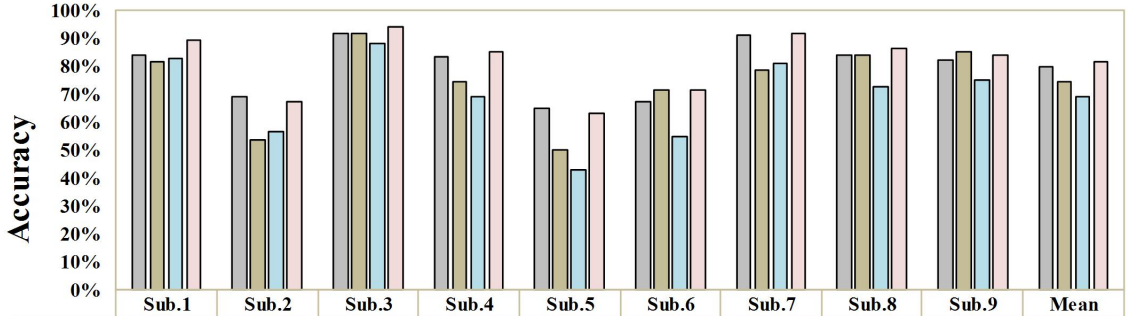
Model_0: Ensemble Flashlight-Net model.
Model_4: Flashlight-Net model using the first band (4∼38Hz).
Model_5: Flashlight-Net model using the second band (μ band).
Model_6: Flashlight-Net model using the third band (β band).

In the existing MI studies, μ and β bands have received great attention and have been identified as two key bands [26,28]. In this work, three bands are employed. Particularly, the first band (4∼38Hz) covers the next two bands (i.e., μ and β bands). Such a setting not only takes into account the information of these two key bands, but also learns about the coupling relationship between them. 38 Hz can be changed appropriately and has little effect on the average classification results of multiple subjects. Figure 8 displays the classification accuracies of the studied 9 subjects. It is found that each band contains effective information for decoding MI signals, with average accuracies of 79.55% (Model_4), 74.34% (Model_5), and 68.98% (Model_6), respectively. It is also found that there are certain differences in accuracy among subjects. Among them, the accuracy gap (based on Model_0) between Sub.3 and Sub.5 is the largest, reaching 30.97%. However, the maximum difference of Sub.3 under different models is only 5.97%. That is, the worst classification accuracy of Sub.3 is better than the best classification accuracy of Sub.5, showing obvious individual differences, known as subject specificity. Ensemble Flashlight-Net model integrating three bands has obvious advantages in the overall performance (with average accuracies of 81.23%), which plays an important role in improving the stability and robustness of the model.

C. Comparison with existing studies

BCI Competition IV-2a dataset is a benchmark dataset for 4-class MI research. Based on this dataset, many MI classification studies have been carried out. Here we do not consider the studies that combine the data of two sessions together for model training. Some comparable studies and their results are listed in Table II. They all used the BCI Competition IV-2a dataset.

The existing studies can be divided into two frameworks. One is to use an independent machine learning method to conduct the feature extraction, and then to set a classifier



	Sub.1	Sub.2	Sub.3	Sub.4	Sub.5	Sub.6	Sub.7	Sub.8	Sub.9	Mean
Model_4	84.03%	69.10%	91.32%	82.99%	64.58%	67.36%	90.97%	83.68%	81.94%	79.55%
Model_5	81.60%	53.47%	91.67%	74.31%	49.65%	71.53%	78.13%	83.68%	85.07%	74.34%
Model_6	82.64%	56.25%	87.85%	68.75%	42.71%	54.51%	80.56%	72.57%	75.00%	68.98%
Model_0	89.03%	66.87%	93.82%	84.93%	62.85%	71.46%	91.77%	86.32%	84.06%	81.23%

Fig. 8. Accuracies of the ensemble Flashlight-Net model (Model_0) and three Flashlight-Net models (Model_4-6) corresponding to three bands.

TABLE II
MI CLASSIFICATION RESULTS OF THE PROPOSED MODEL VERSUS SOME EXISTING WORKS ON THE BCI COMPETITION IV-2A DATASET.

Work	Year	Description	Accuracy
Bashashati et al.[15]	2016	Bayesian optimization+FBCSP	68.13%
Aghaei et al.[18]	2016	SCSSP	65.03%
Schirrmeister et al.[28]	2017	CNN with cropped training	73.7%
Sakhavi et al.[44]	2017	CNN+Transfer Learning	69.71%
Sakhavi et al.[26]	2018	Modified FBCSP+C2CM	74.46%
Lawhern et al.[39]	2018	EEGNet	69%
Amin et al.[29]	2019	Multi-layer CNNs	75.72%
Amin et al.[40]	2019	CNN layers fusion	74.5%
Azab et al.[41]	2019	Weighted LR + Transfer Learning	75.6%
Li et al.[42]	2019	Channel-Projection Mixed-Scale CNN	74.6%
Xu et al.[43]	2020	RBM+Multi-view feature learning	78.50%
Liang et al. [33]	2020	Riemannian geometry alignment+Transfer Learning	68.5%
Wei et al. [32]	2021	CNN+Transfer Learning	81.8% (5 subjects)
Solórzano-Espíndola et al. [34]	2021	FBCSP+Transfer Learning	74% (3 subjects)
This work	2022	**Ensemble Flashlight-Net model+Transfer Learning**	**81.23%**

to achieve classification [14,15,18,26]. For example, in Refs. [14,15,26], the authors used FBCSP to extract the distinguishable features, and in Ref. [18], SCSSP developed from CSP was utilized. This framework provides a feasible solution for MI classification, but the classification accuracies of related studies are difficult to reach more than 75%.

The other framework is based on data-driven methods, represented by deep learning, especially CNN. These methods are based on large amounts of data and can learn many robust features for classification. Under this research framework, a CNN-based compact architecture named EEGNet [39] was developed, and it achieved an average accuracy of 69%. Schirrmeister et al. [28] developed a shallow CNN model, which increased the average accuracy to 73.7%. Amin et al. [29,40] tried the multi-branch design, and designed one multi-layer CNN architecture and one CNN layer fusion architecture for MI classification. The average accuracies reached 75.72% and 74.50%, respectively. Li et al. [42] proposed the CP-MixedNet model, which obtained the accuracy of 74.6%. Xu et al. [43] used improved deep restricted Boltzmann machine (RBM) network to learn the multi-view features. The average accuracy was 78.50%.

In addition, some studies have introduced transfer learning in MI research. For example, Sakhavi et al. [44] explored the possibility of transferring knowledge by using a CNN model. They trained the model via merging the datasets of Sub.1, Sub.3, Sub.7 and Sub.8, and fine-tuned the model via a new subject. The average classification accuracy of 9 subjects was 69.71%. Similarly, Wei et al. [32] firstly merged the training sets of 5 subjects (i.e., Sub.1, Sub.3, Sub.7, Sub.8 and Sub.9), and then trained the pre-training model. They achieved an average accuracy of 81.8% for these 5 subjects. In this work, the average accuracy of these 5 subjects is improved to 89%. Azab et al. [41] developed a weighted logistic regression based transfer learning algorithm, which can improve the classifier by incorporating data from other users. The average accuracy was 75.6%. Liang et al. [33] selected the suitable source subjects via a Riemannian geometry alignment algorithm, and then utilized the dataset from source subjects and a few new training data from current user to train the classifier. The average accuracy reached 68.5%. Solórzano-Espíndola et al. [34] trained the classifiers via the datasets from multiple subjects but test them only on one. They achieved an average accuracy of 74% (88.97% in our work) on Sub.1, Sub.3 and Sub.9. In fact, some studies [33] have pointed out that simply introducing training samples from other subjects into the current training process may have a negative impact. In this work, the model structure and transfer learning strategy are inseparable. The combination of the two can integrate data from multiple subjects, so as to better extract features directly related to MI tasks, even without the use of additional feature alignment.

VI. CONCLUSION

EEG signals collected at different days show large differences. This phenomenon seriously affects the actual application of BCI systems. How to use the existing EEG signals to train a model, and then directly apply it to test the EEG signals in the following days, has important practical application value. To address this problem, ensemble Flashlight-Net model based on CNN is proposed in this work, which uses EEG signals from different days as training set and testing set, respectively. Ensemble Flashlight-Net model has five well-designed modules and enables the hierarchical feature extraction, ensuring the integrity of effective information. Moreover, ensemble Flashlight-Net model not only considers the information of two key MI-related bands (i.e.,μ and β bands), but also learns about the coupling relationship between them. Experimental analysis confirms that such a setting helps to improve the stability of the model. Furthermore, during the model training, pre-training and fine-tuning are designed. Pre-training process can simultaneously integrate the datasets from multiple subjects and shows strong generalization ability. Through the joint efforts of model structure design, multi-band fusion and transfer learning, the proposed model outperforms existing methods and achieves better results, with an average classification accuracy of 81.23% on BCI Competition IV-2a dataset. All these make the proposed ensemble Flashlight-Net model an advanced solution for decoding multi-class MI signals in BCI systems.

REFERENCES

[1] M. Hamedi, H. Salleh Sh, and A. M. Noor, "Electroencephalographic motor imagery brain connectivity analysis for BCI: a review," *Neural Comput.*, vol. 28, pp. 999-1041, Jun. 2016.

[2] L. H. He, D. Hu, M. Wan, Y. Wen, K. M. von Deneen, and M. C. Zhou, "Common Bayesian network for classification of EEG-based multiclass motor imagery BCI," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, pp. 843-854, Jun. 2016.

[3] M. K. Al-Qaderi and A. B. Rad, "A brain-inspired multi-modal perceptual system for social robots: an experimental realization," *IEEE Access*, vol. 6, pp. 35402-35424, 2018.

[4] A. Vuckovic, P. Sara, and P. Finda, "Unimanual versus bimanual motor imagery classifiers for assistive and rehabilitative brain computer interfaces," *IEEE Trans. Neur. Sys. Reh.*, vol. 26, pp. 2407-2415, Dec. 2018.

[5] Z. K. Gao, W. D. Dang, M. X. Liu, W. Guo, K. Ma and G. R. Chen, "Classification of EEG signals on VEP-based BCI systems with broad learning," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, pp. 7143-7151, 2021.

[6] Y. Q. Li, J. H. Pan, F. Wang, and Z. L. Yu, "A hybrid BCI system combining P300 and SSVEP and its application to wheelchair control," *IEEE Trans. Biomed. Eng.*, vol. 60, pp. 3156-3166, 2013.

[7] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, et al., "Sparse group representation model for motor imagery EEG classification," *IEEE J. Biomed. Health. Inform.*, vol. 23, pp. 631-641, Mar. 2019.

[8] J. Feng, E. Yin, J. Jin, R. Saab, I. Daly, X. Wang, et al., "Towards correlation-based time window selection method for motor imagery BCIs," *Neural Networks*, vol. 102, pp. 87-95, Jun. 2018.

[9] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin. Neurophysiol.*, vol. 110, pp. 1842-1857, 1999.

[10] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Trans. Cybern.*, vol. 49, pp. 3322-3332, Sep. 2019.

[11] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, p. 016003, Feb. 2017.

[12] H. X. Wang, Q. Tang, and W. M. Zheng, "L1-Norm-Based common spatial patterns," *IEEE Trans. Biomed. Eng.*, vol. 59, pp. 653-662, Mar. 2012.

[13] S. Lemm, B. Blankertz, G. Curio, and K. R. Muller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE. Trans. Biomed. Eng.*, vol. 52, pp. 1541-1548, Sept. 2005.

[14] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Front. Neurosci.*, vol. 6, p. 39, 2012.

[15] H. Bashashati, R. K. Ward, and A. Bashashati, "User-customized brain computer interfaces using Bayesian optimization," *J. Neural Eng.*, vol. 13, p. 026001, Apr. 2016.

[16] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Composite filter bank common spatial pattern for motor imagery-based brain-computer interface," *2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, pp. 1-5, 2011.

[17] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *IEEE. Trans. Biomed. Eng.*, vol. 58, pp. 355-362, 2011.

[18] A. S. Aghaei, M. S. Mahanta, and K. N. Plataniotis, "Separable common spatio-spectral patterns for motor imagery BCI systems," *IEEE Trans. Biomed. Eng.*, vol. 63, pp. 15-29, Jan. 2016.

[19] S. Shahtalebi and A. Mohammadi, "Ternary ECOC classifiers coupled with optimized spatio-spectral patterns for multiclass motor imagery classification," *in 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017, pp. 2231-2236.

[20] S. Abenna, M. Nahid, and A. Bajit, "Motor imagery based brain-computer interface: improving the EEG classification using delta rhythm and LightGBM algorithm," *Biomed. Signal Process. Contr.*, vol. 71, p. 103102, 2022.

[21] K. Jindal, R. Upadhyay, and H. S. Singh, "A novel EEG channel selection and classification methodology for multi-class motor imagery-based BCI system design," *Int. J. Imaging Syst. Technol.*, vol. 71, pp. 1318–1337, 2022.

[22] X. Deng, B. X. Zhang, N. Yu, K. Liu, and K. W. Sun, "Advanced TSGL-EEGNet for motor imagery EEG-based brain-computer interfaces," *IEEE Access*, vol. 9, pp. 25118-25130, 2021.

[23] O. Y. Kwon, M. H. Lee, C. T. Guan, and S. W. Lee, "Subject-independent brain-computer interfaces based on deep convolutional neural networks," *IEEE Trans. Neur. Net. Lear.*, vol. 31, pp. 3839-3852, Oct. 2020.

[24] M. Y. Xu, J. F. Yao, Z. H. Zhang, R. Li, B. R. Yang, C. Y. Li, et al., "Learning EEG topographical representation for classification via convolutional neural network," *Pattern Recogn.*, vol. 105, Sep. 2020.

[25] B. E. Olivas-Padilla and M. I. Chacon-Murguia, "Classification of multiple motor imagery using deep convolutional neural networks and spatial filters," *Appl. Soft Comput.*, vol. 75, pp. 461-472, Feb. 2019.

[26] S. Sakhavi, C. T. Guan, and S. C. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Trans. Neur. Net. Lear.*, vol. 29, pp. 5619-5629, Nov. 2018.

[27] R. Mane, N. Robinson, A. P. Vinod, S. W. Lee, and C. T. Guan, "A multi-view CNN with novel variance layer for motor imagery brain computer interface," *in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 2950-2953.

[28] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.*, vol. 38, pp. 5391-5420, Nov. 2017.

[29] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, and M. S. Hossain, "Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion," *Future Gener. Comp. Sy.*, vol. 101, pp. 542-554, Dec. 2019.

[30] L. Wen, L. Gao, and X. Y. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, pp. 136-144, Jan. 2019.

[31] W. Hang, W. Feng, R. Du, et al., "Cross-subject EEG signal recognition using deep domain adaptation network," *IEEE Access*, vol. 7, pp. 128273-128282, 2019.

[32] X. Wei, P. Ortega, and A. A. Faisal, "Inter-subject deep transfer learning for motor imagery EEG decoding," *in 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2021, pp. 21-24.

[33] Y. Liang and Y. Ma, "Calibrating EEG features in motor imagery classification tasks with a small amount of current data using multisource fusion transfer learning," *Biomedical Signal Processing and Control*, vol. 62, p. 102101, Sept. 2020.

[34] C. E. Solórzano-Espíndola, E. Zamora, and H. Sossa, "Multi-subject classification of motor imagery EEG signals using transfer learning in neural networks," *in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021, pp. 1006-1009.

[35] A. Karatzoglou, N. Schnell, and M. Beigl, "Applying depthwise separable and multi-channel convolutional neural networks of varied kernel size on semantic trajectories," *Neural Comput. Appl.*, vol. 32, pp. 6685-6698, Jun. 2020.

[36] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *in Proc. Int. Conf. Learn. Representations*, 2016.

[37] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, arXiv:1803.01271.

[38] C. L. Guo, C. Y. Li, J. C. Guo, R. M. Cong, H. Z. Fu, and P. Han, "Hierarchical features driven residual learning for depth map super-resolution," *IEEE Trans. Image Process.*, vol. 28, pp. 2545-2557, May 2019.

[39] 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, p. 056013, Oct. 2018.

[40] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain, "Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification," *IEEE Access*, vol. 7, pp. 18940-18950, 2019.

[41] A. M. Azab, L. Mihaylova, K. K. Ang, and M. Arvaneh, "Weighted transfer learning for improving motor imagery-based brain-computer interface," *IEEE Trans. Neur. Sys. Reh.*, vol. 27, pp. 1352-1359, Jul. 2019.

[42] Y. Li, X. R. Zhang, B. Zhang, M. Y. Lei, W. G. Cui, and Y. Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding," *IEEE Trans. Neur. Sys. Reh.*, vol. 27, pp. 1170-1180, Jun. 2019.

[43] J. Xu, H. Zheng, J. Wang, D. Li, and X. Fang, "Recognition of EEG signal motor imagery intention based on deep multi-view feature learning," *Sensors (Basel)*, vol. 20, Jun. 2020.

[44] S. Sakhavi and C. Guan, "Convolutional neural network-based transfer learning and knowledge distillation using multi-subject data in motor imagery BCI," *in 2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2017, pp. 588-591.

Supplementary Material to "Flashlight-Net: a modular convolutional neural network for motor imagery EEG classification"

Weidong Dang, Dongmei Lv, Mengxiao Tang, Xinlin Sun,Yong Liu, Zhongke Gao, *Senior Member, IEEE,* and Celso Grebogi

***Abstract*—In this supplementary material, we test the proposed ensemble Flashlight-Net model on the Physionet EEG Motor Movement/Imagery Dataset. The experimental results show that the proposed model achieves an average classification accuracy of 70.95% for four classes. Six state-of-the-art methods are employed for performance comparison.**

***Index Terms*—Motor imagery, convolutional neural network, brain-computer interface.**

I. Dataset description

Physionet EEG Motor Movement/Imagery Dataset [1,2] is another publicly available dataset for 4-class MI research. Four subjects are discarded due to variability in the number of trials, resulting in 105 subjects to be finally used. The subjects conducted both motor movement and MI tasks, however, in this work we solely focus on the classification of MI. EEG signals from 64 brain electrodes are recorded. The sampling frequency is 160 Hz. Each subject participates in three runs for MI of left fist (L) against right fist (R), and three runs for MI of both fists (B) against both feet (F). One run lasts 120 s and consists of 14 MI trials, which results in 21 trials per class per subject. In this work, each trial contributes a sample $X \in \mathbb{R}^{C\times T}$, where C = 64 is the number of channels (or electrodes) and T = 640 is the number of data points. For each subject, we set 75% of the samples in each class as the training set and the remaining samples as the testing set. We refer to this dataset as Physionet MI dataset in subsequent writing.

II. Four-class MI classification

In this section, we test the proposed ensemble Flashlight-Net model and training strategy on the Physionet MI dataset. Compared with the BCI Competition IV-2a dataset, the sample size of the Phyionet dataset varies from $X \in \mathbb{R}^{22\times 1000}$ to $X \in \mathbb{R}^{64\times 640}$. Therefore, some convolution kernel parameters of the proposed model are adjusted accordingly, as follows:
1) In the channel fusion module (CF module), the kernel size of convolutional layer $L3$ is changed from 22*1*10 to 64*1*10.
2) In the time domain module (TD module), the kernel sizes of three convolutional layers ($L5$-7) are both changed from 1*125 to 1*80.
3) In the feature pool module (FP module), the kernel sizes of convolutional layers $L11$-12 are both changed from 1*10 to 1*8. The kernel sizes of convolutional layers $L13$-14 are both changed from 1*20 to 1*16. The kernel sizes of convolutional layers $L15$-16 are both changed from 1*30 to 1*24.

Following the transfer learning strategy described in Section II-C, we conduct the model training and testing. The epochs during the fine-tuning process are adjusted to 10, 20, and 20, respectively.

Experimental results are shown in Table I. The average classification accuracy of ensemble Flashlight-Net model on all 105 subjects reaches 70.95%. To further demonstrate the effectiveness of the transfer learning strategy, we also test the results without transfer learning (i.e., using only the fine-tuning process). The average accuracy is only 25.61%, which confirms that integrating multiple subject information through pre-training is a very effective solution for MI decoding.

Additionally, six state-of-the-art methods are employed for performance comparison, including MWFFConvnet [3], HS-CNN [4], EEGNet [5], DeepConvnet [6], ATCNet [7], and MShallowConvNet [8]. All these methods have been successfully applied for MI classification. During the training process, these methods employ the same training strategies, as described in Section II-C. The corresponding results are shown in Table I. It is found that the proposed model exhibits certain advantages in the MI decoding based on the Physionet MI dataset.

W. Dang, D. Lv, L. Sun, and Z. Gao, are with the School of Electrical and Information Engineering, Tianjin University, Tianjin, 300072, China (e-mails: weidongdang@tju.edu.cn; zhongkegao@tju.edu.cn; dongmeilv@tju.edu.cn; xinlinsun@tju.edu.cn).

M. Tang is with the School of Electrical and Electronic Engineering, Nanyang Technology University, Singapore, 50 Nanyang Ave, 639798, Singapore (e-mail: MTANG007@e.ntu.edu.sg).

Y. Liu is with the Xinqiao Hospital, Army Medical University, Chongqing, 400037, China (e-mail: lzsht@aliyun.com).

C. Grebogi is with the Institute for Complex Systems and Mathematical Biology, King's College, University of Aberdeen, Aberdeen AB24 3UE, UK (e-mail: grebogi@abdn.ac.uk).

References

[1] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, and J.R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE. Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034-1043, 2004.

[2] A. Goldberger, L. Amaral, L. Glass, et al., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[3] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif and M. S. Hossain, "Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification," *IEEE Access*, vol. 7, pp. 18940-18950, 2019.

TABLE I
AVERAGE CLASSIFICATION ACCURACY OVER ALL 105 SUBJECTS ON THE PHYSIONET MI DATASET.

Method	MWFFConvnet[3]	HS-CNN[4]	EEGNet[5]	DeepConvnet[6]	ATCNet[7]	MShallowConvNet[8]	Ensemble Flashlight-Net
Average accuracy	58.34%	64.69%	65.92%	67.79%	68.48%	69.28%	**70.95%**

[4] G. H. Dai, J. Zhou, J. H. Huang, and N. Wang, "HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification," *J. Neural Eng.*, vol. 17, no.1 p. 016025, Feb. 2020.

[5] 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, p. 056013, Oct. 2018.

[6] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.*, vol. 38, pp. 5391-5420, 2017.

[7] H. Altaheri, G. Muhammad, and M. Alsulaiman, "Physics-inform attention temporal convolutional network for EEG-based motor imagery classification," *IEEE Trans. Ind. Inform.*, vol. 19, pp. 2249-2258, 2023.

[8] S. J. Kim, D. H. Lee, and S. W. Lee, "Rethinking CNN architecture for enhancing decoding performance of motor imagery-based EEG signals," *IEEE Access*, vol. 10, pp. 96984–96996, 2022.