Densely Feature Fusion Based on Convolutional Neural Networks for Motor Imagery EEG Classification

DONGLIN LI [ID], JIANHUI WANG, JIACAN XU, AND XIAOKE FANG
School of Information Science and Engineering, Northeastern University, Shenyang 110819, China
Corresponding author: Donglin Li (donglin_li@aliyun.com)

ABSTRACT Electroencephalogram (EEG) signals have been used in the Brain-computer interface (BCI) technology to implement direct communication between the human body and the outside world, which has important application prospects in the fields of cognitive science and medical rehabilitation. In recent years, deep learning technology has achieved remarkable results in the BCI system, especially the using of convolutional neural networks (CNNs) frameworks for the identification and analysis of motor imagery signals. However, practical applications are limited by the complex process of data representation, and the end-to-end method will deteriorate the recognition results. In this paper, we propose a densely feature fusion convolutional neural networks (DFFN). Combining the morphological information of EEG signals, we propose two data representation methods with low complexity, then design and optimize the densely feature fusion network framework for this form of inputs. DFFN considers the correlation between adjacent layers and cross layer features, which reduces the information loss in the process of convolutional operation and considers the local and global characteristics of the network. The simulation results showed that our network improve classification results by 5% in the BCI competition IV-2a data set compare to the ordinary CNNs framework. In order to verify the practical application of the densely feature fusion network framework, we train an adaptive global model method. The results of average classification are close to the baseline approach of the subject-dependent model and better than others.

INDEX TERMS Brain-computer interface (BCI), electroencephalogram (EEG), convolutional neural networks (CNNs), densely feature fusion convolutional neural networks (DFFN).

I. INTRODUCTION

Brain-computer interface (BCI) analyzes and studies brain electrical activity by using non-invasive (scalp electrodes) or invasive techniques (intracranial electrodes). It establishes a non-muscle channel that facilitates direct communication between the human body and external devices [1]. Electroencephalography (EEG) is a comprehensive reflection of cellular activity in the brain, which reflects people's physiological and psychological activities [2]. The EEG-based method is a commonly used non-invasive method. Because of its good real-time performance and low operating cost, it is widely used compared with non-invasive methods such as FMRI, MEG, and PET. The BCI system can be used

in a variety of fields, such as signal processing, cognitive science, medicine and rehabilitation [3], [4]. From the EEG signals [6], brain activity can be detected in several modes which can determine the user's intention, and be used for BCI communication. One of the popular modes is motor imagery (MI). Motor imagery (MI) means that subjects perform the action by using their brain to imagine a certain part of body (such as left hand, right hand, tongue and feet), rather than moving it, meanwhile, the sensorimotor cortex of the brain shows some oscillating activity corresponding to specific imagination [7]. Machine learning technology is often used to classify and identify these MI tasks [8], [9]. People can use the MI-BCI system to control the operation of external devices, such as wheelchair control or neural prosthesis for disabled people, and help healthy people perform demanding tasks, control devices [10], automatic driving [11], epilepsy

The associate editor coordinating the review of this manuscript and approving it for publication was Longzhi Yang.

diagnosis [12] and so on. The low cost, high availability, and the need for any manual assistance of EEG signals have made the intelligent machines built by motor imagery BCI systems increasingly powerful. Therefore, improving the decoding and classification of EEG signals are very important for the future development of artificial intelligence.

The EEG signal motor imagery process has an energy change, represented by a transient and persistent amplitude attenuation and enhancement [13], which is known as event-related desynchronization/event-related synchronization (ERD/ERS) [14]. Currently, BCIs based on ERD/ERS has been attracted wide attention due to its potential application in sports rehabilitation and assistance [15], [16]. However, the ERD/ERS pattern varies in signal pattern, frequency range and location according to individual characteristics [17], [18]. In order to avoid the main problem of this inherent difference, a common spatial pattern (CSP) is proposed and widely applied to MI-BCI [19], [20]. This algorithm tries to find the best spatial filter to maximize the projection scattering difference between two kinds of EEG signals. However, CSP is highly dependent on the covariance of frequency bands and samples, so the filter bank common spatial pattern (FBCSP) [21], discriminant filter bank public space model (DFBCSP) [22] and RCSP [4] are proposed to overcome these problems. At the same time, several excellent CSP variants [23]–[34] have been proposed in the literatures. These methods calculated the relative energy of the filtered channel as the representation of data. This representation form of high-dimensional EEG data can be easily input into linear classifier, such as support vector machine (SVM) and extreme learning machine (ELM). However, most of the above methods only consider two types of information in the space, frequency and time domain of EEG signals. Among the data representation methods, the existing MI-BCI feature extraction methods mainly focus on the extraction of static energy features [35], [36], while ignoring the dynamic nature of signals in the process of motion imagination. Therefore, valuable MI information will be lost in this process, and computational cost will be increased.

Traditional machine learning techniques are combined with the above methods to extract meaningful information from EEG signals. The traditional machine learning method is widely used in the fields of EEG signals, but its performance and accuracy in EEG signals processing are not satisfactory. To improve this situation, researchers began to study the potential of using various deep learning models in the EEG signals analysis [37]. Deep learning model, especially convolutional neural network (CNN), can extract features with higher discrimination and robustness [38]. Other models such as recurrent neural network (RNN) [39], LSTM [40], automatic encoder SAE [41], deep belief network (DBN) are particularly useful in applications with time series. Researches show that deep learning technology has achieved good results in the field of EEG [42], which indicates that features extracted automatically are better than those extracted manually. EEG signals contains artifact, high dimension,

channel correlation and other problems. Therefore, there are still many problems in the classification and recognition of MI EEG signals by using deep learning, the establishment of a framework based on deep learning is a complex and challenging task.

Vernon j. Lawhern *et al.* [43] studied a compact CNN: EEGNet, which encapsulated various methods for feature extraction of EEG signals for BCI system and constructed a unified standard. Cuntai Sakhavi *et al.* [44] studied the data representation by introducing FBCSP algorithm, and then used the optimized CNN structure for classification. They also discussed and analyzed three types of convolutions method that can be selected for each layer of the network when designing the network: time convolution, channel convolution, and two-dimensional convolution. Brenda E. Olivas-Padilla *et al.* [45] proposed two motor imagery classification methods. Both of the methods introduce variants of DFBCSP to represent the data, and finally compare the classification effects of the modular expert CNNs network and CNN framework. Bashivan *et al.* [46] transformed EEG signals into topological map by using fast Fourier transform (FFT) over a specific time intervals, and then input them into CNN and LSTM frameworks for classification These methods consider the characteristics of the three dimensions of EEG signals, and the classification effect has been improved, but the complex data representation method and deep learning framework increase the computational cost at the same time.

In order to reduce the time and computational burden of data representation, Hauke Dose *et al.* [47] proposed an end-to-end learning model, and used the CNN framework to learn generalized features and dimensionality reduction, while the traditional fully connected layer (FC) was used for classification. This method can learn from the original data with a minimum amount of preprocessing. Current EEG-based BCI systems often extremely depend on subject before they can be used for new users. Zhang *et al.* [48], considered the practicability of BCI system, then established a global model by using transfer learning, and applied the well-trained model directly to new users without specifying the environment. Although these methods solve the corresponding problems, they reduce the classification accuracy.

Recent studies have shown that if convolutional neural networks are used in the field of EEG, multi-layer feature fusion has better performance than models using only the last layer feature. By adding the method of feature fusion [49], the results will be more accurate and effective. Yuan *et al.* [50], proposed the concept of multi-view learning and fusion of multi-view features of EEG signal channel. However, these networks have some shortcomings: First, there are many manually selected modules in the network structure, and the adjusting parameters are complex. Second, the continuous convolutional process of traditional CNN loses the significant information of the input or gradient as the number of layers increases. Finally, the features learned from the shallow framework will be affected by the redundant information of the data set, resulting in low learning efficiency.

Based on the above problems, this paper proposes a novel densely feature fusion network structure. The main content is as follows:

1) In the data representation stage, considering the difference of data distribution and space-frequency characteristics, we propose two time-domain signal representation methods with low complexity to obtain dynamic related information of EEG signals.

2) In the stage of feature extraction and classification, we propose a new feature fusion network. Feature fusion network considers the relevant features of each CNN layer. These features not only contain important information of their own input data, but also integrate information from all previous layers. This network learns the correlate features between the adjacent layer and cross-layer, while reducing the information loss in the process of convolutional operation, and considers the local and global characteristics of the network.

3) Finally, the global model of feature fusion network is trained and adaptive method is added to improve the accuracy of classification. This data representation method and fusion network structure improves the performance of the traditional CNN framework and significantly improves the average accuracy on the BCI competition IV-2a data set [51].

This paper is divided into the following sections. Section 2 describes the used data sets. Section 3 mainly introduces the data representation method and the basic structure of the DFFN. Section 4 compares the performance of the DFFN with other frameworks. Then we verifies the practicability of the DFFN, and finally analyzes the learning results of the network. The conclusions are discussed in Section 5.

II. DATA

In this paper, we use the 2008 BCI competition IV-2a EEG data set recorded from 22 Ag/AgCl electrodes with a 250-Hz sampling rate and band-pass filtered between 0.5 and 100 Hz from 9 subjects. It consists of four different motor imagery tasks including the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). The timing scheme consists of a fixed 2 seconds, a reminder time of 1.25 seconds, followed by a period of a MI of 4 s. The imagined tasks consist of thinking to move both hands and feet and the tongue. The data set consists of two sessions of data, training data and evaluation data. Each session has 288 trials for each of the training and evaluation data (72 trials per class).

Data preprocessing can restore the original appearance of data as much as possible, and will not interfere with the method proposed in this paper due to noise, null value and other problems. The data preprocessing is as follows.

1) We use the half period of MI (0 to 2 s) as the time segment. According to [51], We eliminated the data of the wrong experiment. For the problem of null value, we chose linear interpolation to fill in the missing value.

2) The EEG signals from all recorded channels are filtered using a filter bank with bandpass filters (8-30 HZ).

III. METHOD

In this paper, the main process of data representation is based on CSP, which is the most common methods used in the field of EEG. In Section A, we will review the highlights of the CSP algorithm. And discuss the EEG representation based on CSP. In Section B, we introduce the basic framework and principles of Densely Connected Convolutional Networks (Densenets) [52]. In Section C, we propose a densely feature fusion network based on Densenets and explain the selection of network parameters.

A. EEG REPRESENTATION BASED ON CSP

The CSP algorithm [14], is effective in discriminating two classes of EEG data by maximizing the variance of one class while minimizing the variance of the other class, which is viewed as a spatial filtering algorithm. At present, CSP method and its improvement method occupy a certain proportion in the field of EEG signal classification, and the effect is obvious. The mathematical expression of the CSP function is as follows:

$$J(w) = \frac{w^T X^T X w}{w^T Y^T Y w} = \frac{w^T C_h w}{w^T C_f w}$$

where w is the spatial filter learned by the eigenvectors extracted. C_i is the covariance matrix of the two categories. We use CSP algorithm to extraction a coefficient vector of a spatial filter W_{csp}, and chose the first and last N rows of CSP projection matrix W_{csp}, then the EEG signals is projected as $Z = W^T X$. The features can be obtained as $f = log\left(\sum_{t=0}^{L} Z(t)^2\right)$. However, the CSP algorithm is primarily used to differentiate two categories of data, so in the case of multiple categories, the one-versus-rest strategy needs to be chosen.

we extract the target spatial characteristics as the input of the network. The steps are shown below:

1) After using the one-versus-rest strategy and the CSP algorithm for each class, we get projected EEG signals. Select the first and last N = 2 rows of spatial filters W_{csp} and the number of filter groups $N_w = 4$ are selected (the number of filter groups is related to the classification category), so the size of the projected EEG signal data is 16*500. (Instead of taking a one-vs-rest strategy for classification, we vertically concatenate the four categories of features to form a signal matrix.)

2) we consider two possible representations for the EEG:

a) F1: Hilbert transform [53] is used to extract the envelope of each feature channel, which gives the analytic form of a signal that is complex-valued. Hilbert transform of the filtered EEG signal Z(t) is expressed as Y(t), construct analytic signal $C(t) = Z(t) + jY(t)$. The envelope is calculated from the amplitude of the analytic form. Finally, the envelope is selected as the representation of EEG signal: $a(t) = \sqrt{Z(t)^2 + Y(t)^2}$;

b) F2: We used the logarithmic transformation method in the last step of CSP algorithm, but we did not sum the
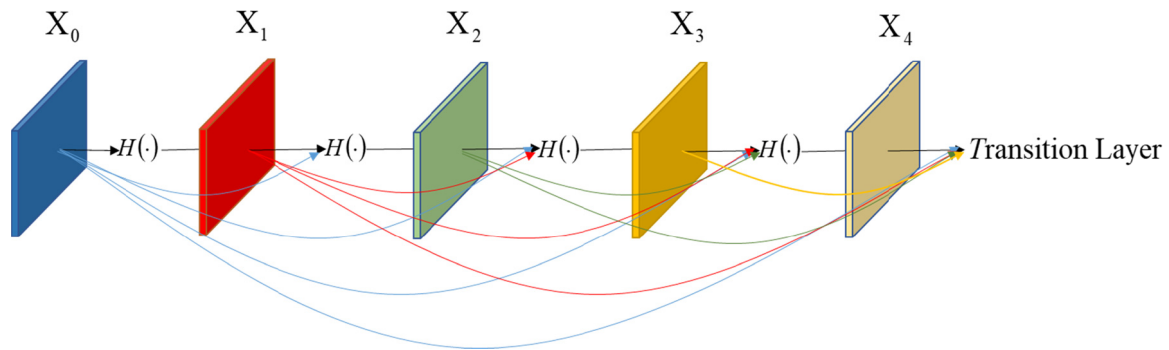
FIGURE 1. This is a five-layer dense block module, which is a mode of connection in Densenets. Each layer fusing with the characteristics of all previous layers.

information at all time points, and changed the formula: $f = log(Z(t)^2)$. Due to the influence of environmental factors, the EEG signal distribution of the subjects will cause complex signal fluctuations during the collect process. There presentation method of logarithmic transformation can make the fluctuation relatively stable.

In the preliminary processing of the input data, the data representation methods with low complexity is used. Therefore, the input data matrix is only the spatial feature obtained by shallow learning, without considering the time-domain feature. Finally, we put this set of feature matrices into the deep learning framework, and conducted in-depth mining and analysis of spatial and temporal features.

B. FEATURE FUSION IN DENSENETS

The architecture of convolutional neural network consists of basic "components" such as input layer, convolution layer, pooling layer, full connection layer, and most architectures follow this process. However, the difference between networks is that, in order to make the network train faster or avoid over-fitting of the network, the network will add linear correction unit Relu, dropout, batch normalization and other methods to change the network architecture.

In this paper, we improve the framework of Densenets to make it more suitable for feature extraction and recognition of EEG signals. We will first explain the basic concepts of Densenets. Then we describe the choice of architecture for Densely Connected Convolutional Networks.

The fusion of features involved in Densenets architecture is a simple connection pattern. In order to ensure maximum information flow between layers of the network, the network directly connects the output feature maps of all layers. Each layer receives additional input from all previous layers and passes its own feature-maps to all subsequent layers, so that the network maintains feedforward characteristics. Figure 1 illustrates this architecture schematically. Traditional convolutional networks have L layers, and Densenets connect each layer together in a feedforward manner. The lth layer has l inputs, it is composed of feature maps of all previous convolution blocks to ensure that features are spliced before being transferred to the next layer. At the same time, the

feature-maps of this layer are passed to all subsequent layers of $L-l$, which is used as the input of all subsequent layers. Finally, new $L(L+1)/2$ connections are introduced into the L-layer network of traditional convolutional networks. The process of feature fusion is to split the feature map together in a certain dimension. At the same time, [52] also observed that dense connection has regularization effect, reducing the over-fitting of small sample data. The proposed deep learning network of EEG signals is rarely targeted at small sample data, so the Densenets-based structure is suitable for EEG signals. Densenets architecture is mainly composed of the dense block layer and the transition layer. Figure 2 illustrates the framework structure of the Densenets network. We will introduce them as flow:

1) DENSE BLOCK

In the dense block, the input of the block is assumed to be X_0. The block has a total of L dense block layers, in which there is a composite algorithm of nonlinear transformation between each layer, defined as $H(\cdot)$: $BN-ReLU-Conv(1 \times 1) -dropout- BN-ReLU-Conv$ (3×3), of which the (1×1) convolution layer before each (3×3) convolution to reduce the number of input feature-maps. Layer l receives feature maps from all preceding layers $X_0, \ldots, X_{l-1}$, as input:

$$X_l = H_l([X_0, X_1, \ldots, X_{l-1}]),$$

where $[X_0, \ldots, X_{l-1}]$ refers to the concatenation of the feature-maps produced in layers $0, \ldots, l-1$.

2) TRANSITION LAYER

In order to avoid excessive computation, transition layer is added into Densenets architecture to reduce the number of feature maps and make the network structure more compact. In order to solve the problem of compatibility of feature fusion and computing efficiency, Densenets modularized the dense layer and divided several dense connection modules. Transition layer is added between modules for connection. This layer is formed by BN-Relu-conv (1×1)-dropout-pooling. If each function H_l produces k feature-maps, $(\theta \times k)$ operation is added before the convolution operation in the
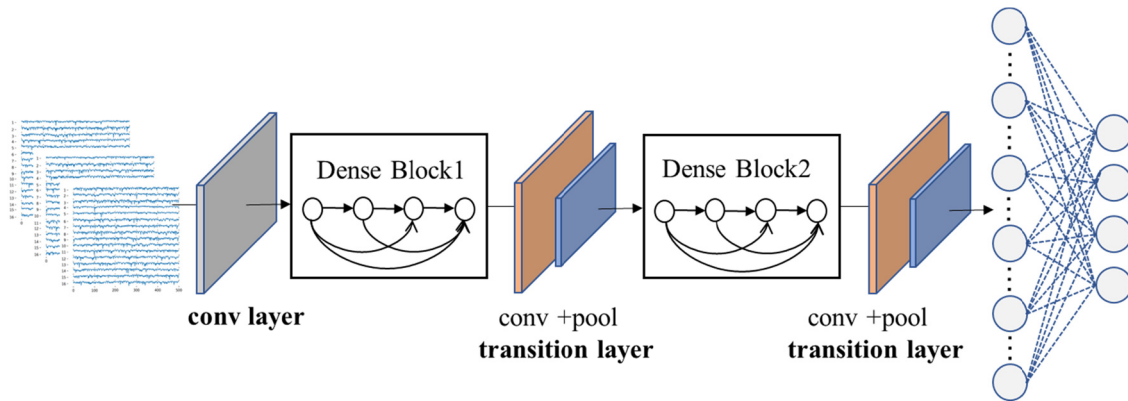
FIGURE 2. This is the basic structure of Densenets network used in this paper. It consists of two groups of Dense blocks, transition layer, and finally two fully connected layers.

transition layer, where $0 < \theta \leq 1$ is called the compression factor, and the number of feature splicing per time can be changed by changing the size of θ.

C. DESIGNING THE FEATURE FUSION NETWORK ARCHITECTURE

For designing the network, we need to consider the nature of the input. In terms of spatial features, each feature channel has a unique spatial filter (based on the selected eigenvalues in the CSP algorithm) to distinguish one class from others. Since the input of the network is a 2D signal after spatial filtering, and it is already a linear combination feature of the original EEG electrode channels. But usually the EEG channels at this stage have no interaction, and the order in the input matrix does not affect the classification. So we choose a 1D spatial filter whose kernel size is the number of input channels to fuse the information of each channel, instead of small size convolution of channels.

In terms of time-frequency domain features, the data contains 500 sample points, including the main information of EEG signals. Convoluting across time domain sample points will deeply learn morphology hidden in EEG signals. In this paper, time domain convolution is realized by improved Densenets. Densenet has several advantages: they alleviate the vanishing-gradient problem, avoid excessive loss of information in the convolution process, enhance feature propagation, and encourage feature reuse. The two 1D convolution calculation respectively realizes the overall 2D convolution method, which is relatively independent in space and time, and increases the network changes at the same time.

As shown in Figure 2, the input of network first passes the convolution of channels, and then put them into the modified Densenets. In each dense block, there are a certain number of dense block layers. They are responsible for fusing the features of EEG signals at each learning stage, and splicing the learned morphological features of EEG signals together. Then the fusion features are transferred to the transition layer, and the size of the fusion feature-maps are controlled by weight θ, this operation is to avoid excessive fusion data

volume and parameter explosion. After repeating this phase, the feature map is sent to the last two fully connected layers of the network for classification of motor imagery. In the feature fusion framework, each part, such as convolution, dense block and transition layer, selects corresponding parameters through cross validation, and finally displays and discusses in the results.

Compared with the traditional network architecture, the frame parameters we choose can be applicable to each subject and will not cause too much deviation of the maximum accuracy due to individual differences. We only need to adjust the variables θ and m to ensure accurate classification. (m is the variable that controls the number of convolution filters in the first layer, in order to increase the feature-maps of channel fusion). When the sample size of input data becomes large, we just need to adjust the number of the dense block to deepen the network.

The training of the networks is performed with the following configurations.

1) Adam algorithm is selected as the optimizer algorithm of the network, and the parameter is set as the initial value.

2) the loss function algorithm chooses the cross entropy.

3) The learning rate is set as 10e-5.

4) Dropout is added between convolution operations in the compound algorithm H $(\cdot)$ and before the average pooling layer, and the parameter is set to 0.1.

IV. RESULTS

For experiments, the data preprocessing and signal representation that we carried out in Matlab 2017b environment, and we used the 16GB RAM CPU of Intel(R) Core i7-7700hq 2.80Ghz. For deep learning, we used the GeForce GTX 1080 GPU with 8GB of RAM, feature fusion network was implemented using Tensorflow deep learning framework.

A. ARCHITECTURE PARAMETER SELECTION

Due to the differences between individuals, we first selected key parameters for each subject, such as the size and number of the kernel in the dense block layer, the number of the

TABLE 1. Structure of feature fusion network model.

layers	output	Feature Fusion Network
Convolution	1*500@256	$conv: 16*1, stride: 1, filters\_num: m*growth\_rate$
Dense Block1	1*500@448	$[conv: 1*25, stride: 1]*3, \ filters\_num: growth\_rate$
Transition Layer1	1*20@358	$conv: 1*1, \ filters\_num: \theta*448$
		$pooling: 1*25, stride: 1*25$
Dense Block2	1*20@550	$[conv: 1*25, stride: 1]*3, \ filters\_num: growth\_rate$
Transition Layer2	1*1@440	$conv: 1*1, \ filters\_num: \theta*550$
		$pooling: 1*25, stride: 1*20$
Full Connect1	1024	$1024, softmax$
Full Connect2	4	$4, softmax$

dense block layer, the number of the dense block, and the size and stride of the kernel in the pooling layer. Then, under the overall relatively stable framework, we fine-tune the variable parameters, channel convolution kernel multiples, and θ value.

In fact, due to time and computational constraints, we did not choose a complex optimization algorithm to optimize parameters in the parameter space. However, we used cross validation for parameter selection. We choose the size and number of convolution kernel according to the characteristics of input data and the requirement that convolution output is an integer. Too much receptive field leads to inaccurate morphological learning, while too little field leads to too much local information learning. We propose the following points based on the above to select parameters:

1) We chose the size of 2^n as the number of kernel and the maximum size was 128. The number of kernel in channel convolution is m $\times$ 2^n.

2) As the amount of data is small sample, we chose the maximum number of dense block layer up to 5 and the maximum number of dense block up to 3.

3) The size and stride of the pooling layer select the limit of maximum 25.

4) 2^n up to 2048 nodes and 4 nodes are selected as output of the two fully connected layers respectively. The regularization method selects gradient clipping to prevent gradient

explosion by controlling the maximum L2 normal form of gradient. Other parameters are initial parameters.

The specific parameters of each part of the Feature Fusion Network are shown in Table 1. Parameters are determined by the above method. The number of convolution kernel is determined by the growth_rate and variable parameters, where growth_rate = 64. θ and m in the Table 1 are variable parameters. $\theta = 0.8$ and m = 4 are selected in the example. Each conv layer shown in the Table 1 corresponds to a sequence: BN-Relu-Conv. Take the Dense Block1 as an example. This layer contains three convolution layers. The size of convolution kernel is 1*25, the sliding step is 1*1, and the number of convolution kernel is growth_rate. Finally, the output contains 448 feature maps which size is 1*500.

B. EXPERIMENTAL RESULTS AND COMPARISON OF BASELINE

In the BCI contest IV dataset 2a, each subject had two sessions, one of which is training data and the other is test data. We preprocessed the data as the representation of EEG signals by the above methods. Finally, the densely feature fusion network framework is used for deep learning and classification, and the results are shown in TABLE 2. We evaluated the performance of the proposed method, and compared with the base line method in TABLE 2, the accuracy results is obtained through cross-validation experiments.

TABLE 2. Accuracy for baseline methods and our method.

	FBCSP[56]	SVM+FBCSP[46]	C2CM[46]	CW-CNN[46]	Monolithic network[47]	Modular Network[47]	DFFN(F1)	DFFN(F2)
Subject 1	76.00	82.29	**87.5**	86.11	83.13	84.91	83.20	85.40
Subject 2	56.50	60.42	65.28	60.76	65.45	66.38	65.69	**69.30**
Subject 3	81.25	82.99	90.28	86.81	80.29	84.74	**90.29**	**90.29**
Subject 4	61.00	**72.57**	66.67	67.36	**81.60**	81.36	69.42	71.07
Subject 5	55.00	60.67	62.50	62.50	76.70	**79.22**	61.65	65.41
Subject 6	45.25	44.10	45.49	45.14	**71.12**	70.67	60.74	69.45
Subject 7	82.75	86.11	89.58	**90.63**	84.00	86.12	85.18	88.18
Subject 8	81.25	77.08	83.33	81.25	82.66	83.81	84.21	**86.46**
Subject 9	70.75	75.00	79.51	77.08	80.74	83.04	85.48	**93.54**
AVE	67.75	71.18	74.46	73.07	78.41	**80.03**	76.44	79.90

TABLE 3. Variable parameters.

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	AVE
Accuracy(F1)	83.20	65.69	90.29	69.42	61.65	60.74	85.18	84.21	85.48	76.44
Parameters θ, m	0.8,4	0.6,4	0.8,4	0.6,5	0.9,4	0.8,4	0.8,4	0.8,4	0.6,5	
Accuracy(F2)	83.46	69.30	90.29	71.07	65.41	69.45	88.18	86.76	93.54	79.71
Parameters θ, m	0.8,4	0.6,4	0.8,4	0.6,5	0.9,4	0.6,4	0.8,4	0.8,4	0.6,5	

TABLE 4. Cross-validation accuracy results of the global model.

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	AVE
CRAM[50]	61.02	42.35	73.11	50.43	50.74	51.48	67.26	69.72	66.85	59.10
EEGNet[45]	53.76	39.54	54.88	43.02	51.80	48.96	60.7	61.38	47.82	56.48
leave-one-subject-out	74.36	49.36	82.48	57.14	44.75	55.76	68.58	70.84	73.15	64.38
adaptive of the model	78.17	58.41	87.77	63.68	53.08	59.33	78.45	82.83	86.20	71.99

We use the FBCSP feature extraction algorithm and its derivative methods as baseline. The classification results in the Table 2 are from the original FBCSP paper. In TABLE 2, we also selected the papers that used deep learning method to solve the motor imagery recognition in the field of EEG signals. Including the results of the paper [44], they proposed a new temporal representation of the data, and then used the CNN network framework for classification, and achieved excellent results. Both the Monolithic network and the Modular network [45] use an optimized convolutional neural network that extract features from spectral and spatial domain data by the DFBCSP method. We used the same database and compared the results from the paper with ours.

The classification results of the feature fusion network proposed in this paper are shown in TABLE 2, and the average accuracy is better than the previous methods. The accuracy values of subjects 4 and 7 in the table are slightly less than the baseline method, but overall the accuracy of the other subjects has improved. The results of the two columns on the right are obtained by using different data representations of the same network. It is obvious that the data representation of F2 is more effective. The first four methods of table all use the data representation method of FBCSP, while we choose the basic CSP method as the data representation. Therefore, compared with the previous four methods, DFFN method does not carry out complex manual preprocessing of data at the early stage, but the classification effect of network is obviously higher than them. It indicates that the feature fusion network retains more complete features of EEG signals, does not lose too much intermediate information, and learns more morphologic features. Compared with end-to-end learning network, the classification effect of feature fusion network is more prominent. Monolithic network and Modular network, these two methods are most prominent for the classification of subject4, 5, and 6. However, the online processing of these two methods is cumbersome. In contrast, the DFFN approach is similar to the end-to-end processing and is not very complicated, but the end result is considerable overall. At the same time, the correct rate of the DFFN method in subjects 4, 5, and 6 is slightly improved compared with the conventional method and the deep learning framework.

C. VARIABLE PARAMETER
Due to individual differences, each subject is not necessarily applicable to the same network framework, and affecting the optimal classification accuracy of each subject. In order to solve this problem and improve the network variability and robustness, we finally added variable parameters to the main part of the network framework. In table 3, the principal framework parameters of each subject are uniform, but the ultimate optimal classification accuracy rate is obtained by adjusting variable parameters. In this way, the limitations of network framework caused by individual differences are avoided, the workload of tuning parameters is reduced, and the variability and robustness of the network are increased. Meanwhile, the structure of the network is approximately modular. The number of dense block layer can be adjusted according to experience for the different number of data.

D. GLOBAL MODEL AND ADAPTIVE MODEL
In general, EEG signals are dynamic, the feature extraction performance of motor imagery tasks also have highly subject -dependent characteristics, with which it differs within the same subject and with other subjects. However, the experimental results presented in TABLE 4 show that it is still a promising method to train a unified feature extractor/classifier for different subjects using the proposed framework in practical application, although the effect is reduced.

We use the leave-one-subject-out approach, using one set of subject data from the source data as tests and other subjects as training sets to train a global model. The learned knowledge is then transferred to the subject under test for category recognition. This is a method of transfer learning, which can reduce the problem of small sample data and individual differences. In this study, we used the feature fusion network framework, and then selected the eight groups of subjects as the training data of the global model, and the other subject as the test object. The global model accuracy values in the table are 13% lower on average than the classification rate dependent on subject, but the model is more practical. The proposed model is compared with the recently published CRAM, which extracted spatio-temporal features of EEG
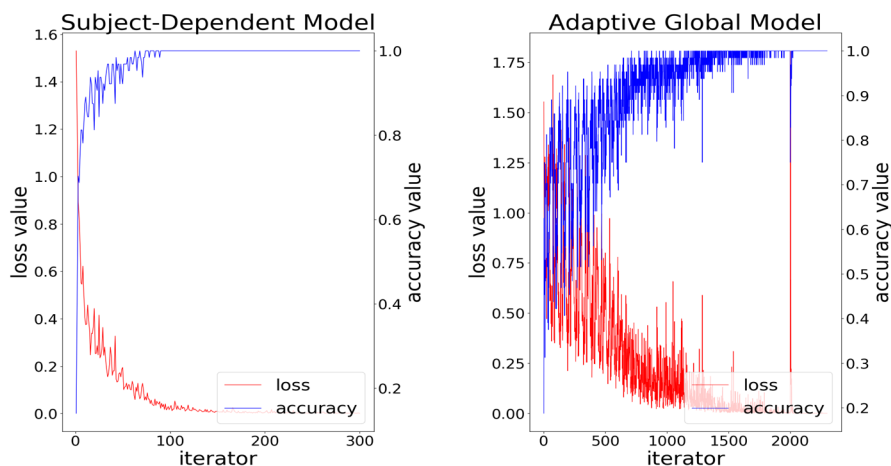
FIGURE 3. Training process of Subject-Dependent model and adaptive global model.

signals sliced at different times, and finally classified and identified them by using the convolution recursive attention model (CRAM), and achieved excellent results compared with end-to-end EEG-Net methods. The end-to-end approach is superior in complexity to our feature fusion network, but loses nearly 10% accuracy. In our algorithm of input EEG data representation, in order to minimize the complexity, the basic algorithm CSP is selected, and finally the accuracy is improved.

The method of transferring learning can avoid retraining the new subject data for a long time, but the results of the experiment will be worse due to the inevitable differences in the data obtained from the experiment. To solve this problem, we use a small sample set for adaptive fine-tuning of the model on the basis of transfer learning. As can be seen from the table, the fine-tuned global model shows a good improvement in classification accuracy and a slight improvement over the baseline method, although the effect is lower than that of the subject-dependent method. Although MI EEG signals have highly subject-dependent characteristics, variable parameters in the model can improve this. In general, the feature fusion network using theme adaptation technology has achieved good results. There are many practical ways to improve the accuracy of classification. For example, the method based on transfer learning is adopted to transfer data sets with similar data distribution, making the global model more accurate.

E. ANALYSIS RESULTS AND FEATURE VISUALIZATION

To analyze the classification results of our proposed model, we selected the global model and the subject-dependent model of subject9. As shown in Figure 3, we draw the loss of two models of Subject9 and the variation of accuracy of the training process, two kinds of loss converges to 0, but the subject-dependent model converges quickly and oscillates little, the global model do opposite. As subjects have different perceptions of motor imagery, data differences are caused, so the global model loss osillates seriously. Although the

oscillation occurs in the training process, the convergence is finally successful. In the 2000 iterations, adaptive adjustment is added to the training, loss and training accuracy have a mutation process, which quickly re-converge in subsequent learning.

As shown in Figure 4, the output of the last transition layer from the DFFN is reduced in dimension and visualization by t-SNE, t-SNE visualization can clearly show the clustering situation of the results, so as to facilitate our analysis and hypothesis of the results, and its scatter diagram is finally drawn as shown. In order to prove the validity of the proposed model, we analyze the final features of the proposed network before fully connected layer for classification. For subject-dependent model, as shown in Figure 4(a), the boundaries of the four categories features are obvious, and the four clusters of the same category are distributed together, which proves that the network learns the morphological characteristics of EEG signals and separates the data of different categories well. For global model, as shown in Figure 4(b), the distribution of feature clusters can be clearly observed, but the fifth cluster contains 4 types of category features similar to each other appears in the middle, which is the difference between the classification results of the global model and the subject-dependent model. However, for the two models, the characteristic data of all category exists overlap, possibly because the MI signals obtained in EEG experiments are not obvious, and the features are not enough to distinguish the movements of the foot and tongue. In the experiment, the electrodes in the brain region were not comprehensive enough to collect signals from the feet and tongue, leading to a small difference between the two signals. As can be seen from the figure, these four categories have different degrees of overlap, which may be caused by the experimental process and individual differences.

V. DISCUSSION

The deep learning algorithm has an important application prospect in the field of EEG. Its unique learning method

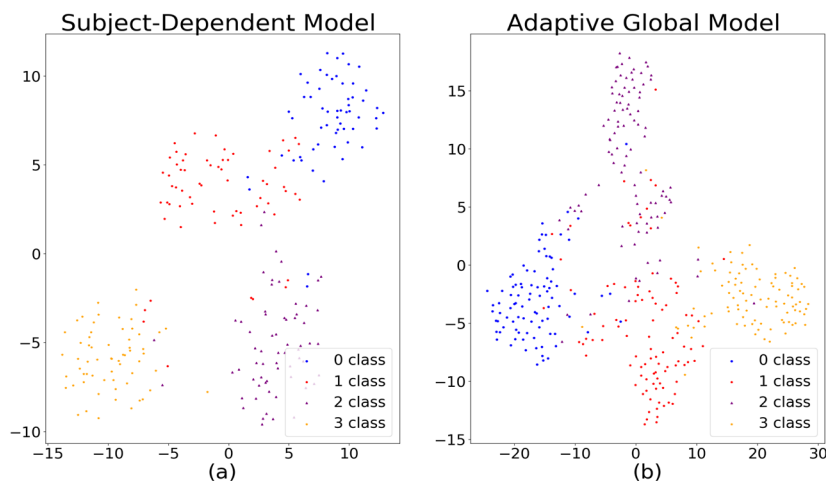Subject-Dependent Model   Adaptive Global Model

FIGURE 4. The t-SNE dimensionality reduction visualization of network characteristics in subject-dependent model and adaptive global model.

and efficiency make it have a negligible position in the field of EEG.

In this research, the first consideration is to simplify the representation of network input data, which effectively reduces the burden of data preparation. At the same time, the characteristics of the spatial and temporal domains of EEG signals are preserved, and the loss of information is reduced. The representation method of logarithmic transformation performs better in the subsequent network, which makes the fluctuation of EEG signal relatively stable. It avoids the influence of redundant morphological characteristics on network learning, and increases the learning efficiency.

Then this paper mainly proposes the application of feature fusion deep learning network in EEG signals. The feature fusion network does not learn the morphological features of EEG signals from the depth and breadth of the architecture, but strengthens the network through reusing or fusing feature. The feature loss in each part of the network and the number of parameters are reduced. The variable parameters in the densely feature fusion network can increase the input changes of the subsequent fusion layer and improve the learning efficiency. The final experimental results also prove that the feature fusion network framework can effectively distinguish four categories of motor imagery by learning MI EEG signals. Compared to SVM and other CNN frameworks, we can produce better results. Our analysis shows that our results do learn something from the input EEG signals, demonstrating that feature fusion structures make sense for MI EEG signals. We observed the clustering effect of learning features after t-SNE visualization and verified the above results.

In order to verify the applicability of the feature fusion network, we study the global model. Most CNN frameworks usually adopt intra-disciplinary classification. However, in practice, the data of labeled subjects are very small, and the training model of small samples is not accurate. Therefore, in the actual problem, the global model of BCI system is always

the priority. In this study, we used a framework of feature fusion networks to compare the two construction methods of the global model. Adaptive of the model has better accuracy, which can approach the accuracy of the subject-dependent baseline method, while the accuracy of the interdisciplinary transfer learning baseline method is difficult to reach 70%. This shows that, unlike ordinary transfer learning, the adaptive model method can effectively overcome the above problems without losing time optimality.

The current framework of the network has three caveats. Firstly, the representation of EEG signals and the optimization of network structure are independent and not related to each other. Secondly, the selection of network architecture parameters for deep learning requires cross validation to select the most appropriate results, which is a complicated and tedious process. In addition, although the Densenet network architecture has the advantage of reducing over-fitting, using a small sample of each class to train a highly concentrated network will still have a problem of over-fitting. Later, a large amount of relevant data can be selected for pre-training to avoid this problem. If two independent algorithms can be combined in the future and there is an excellent theoretical support for parameter selection, the feature fusion network will be more perfect.

In general, the deep learning framework of feature fusion has great prospects in the application of MI-EEG signals, which may have important applications in controlling the prediction of brain-computer interface. In the future, we will also apply the feature fusion method in other EEG studies to verify the robustness of our method and finally improve the feature fusion network.

REFERENCES

[1] K. A. Jellinger, "Reading EEGs: A practical approach," *Eur. J. Neurol.*, vol. 4, no. 17, p. e34, Apr. 2010.

[2] F. Speranza, W. J. Tam, R. Renaud, and N. Hur, "Effect of disparity and motion on visual comfort of stereoscopic images," *Proc. SPIE*, vol. 6055, pp. 94–103, Jan. 2006.

[3] D. J. McFarland, C. W. Anderson, K.-R. Müller, A. Schlogl, and D. J. Krusienski, "BCI meeting 2005-workshop on BCI signal processing: Feature extraction and translation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 135–138, Jun. 2006.

[4] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011.

[5] K. P. Thomas, C. Guan, C. T. Lau, A. P. Vinod, and K. K. Ang, "A new discriminative common spatial pattern method for motor imagery brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 11, pp. 2730–2733, Nov. 2009.

[6] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. H. Lopes da Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, 2006.

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

[8] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1991–2000, Aug. 2008.

[9] 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," *Frontiers Neurosci.*, vol. 6, no. 1, p. 39, 2012.

[10] T. Carlson and J. R. del Millan, "Brain-controlled wheelchairs: A robotic architecture," *IEEE Robot. Autom. Mag.*, vol. 20, no. 1, pp. 65–73, Mar. 2013.

[11] L. Tonin, T. Carlson, R. Leeb, and J. R. del Millán, "Brain-controlled telepresence robot by motor-disabled people," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug./Sep. 2011, pp. 4227–4230.

[12] J.-B. Schiratti, J. Le Douget, M. Le van Quyen, S. Essid, and A. Gramfort, "An ensemble learning approach to detect epileptic seizures from long intracranial EEG recordings," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 856–860.

[13] N. F. Ince, F. Goksu, A. H. Tewfik, and S. Arica, "Adapting subject specific motor imagery EEG patterns in space–time–frequency for a brain computer interface," *Biomed. Signal Process. Control*, vol. 4, no. 3, pp. 236–246, 2009.

[14] C. Park, D. Looney, N. U. Rehman, A. Ahrabian, and D. P. Mandic, "Classification of motor imagery BCI using multivariate empirical mode decomposition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 1, pp. 10–22, Nov. 2012.

[15] F. Pichiorri, G. Morone, M. Petti, J. Toppi, I. Pisotta, M. Molinari, S. Paolucci, M. Inghilleri, L. Astolfi, F. Cincotti, and D. Mattia, "Brain–computer interface boosts motor imagery practice during stroke recovery," *Ann. Neurol.*, vol. 77, no. 5, pp. 851–865, 2015.

[16] M. Tavakolan, X. Yong, X. Zhang, and C. Menon, "Classification scheme for arm motor imagery," *J. Med. Biol. Eng.*, vol. 36, no. 1, pp. 12–21, Jan. 2016.

[17] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzerb, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 6, pp. 642–651, Dec. 1997.

[18] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI)," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 447–456, Dec. 2000.

[19] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[20] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Dec. 2008.

[21] S.-H. Park, D. Lee, and S.-G. Lee, "Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 498–505, Feb. 2018.

[22] H. Higashi and T. Tanaka, "Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1100–1110, Apr. 2013.

[23] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Müller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 11, pp. 2274–2281, Nov. 2006.

[24] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005.

[25] W. Wu, X. Gao, B. Hong, and S. Gao, "Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL)," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 6, pp. 1733–1743, Jun. 2008.

[26] H. Zhang, Z. Y. Chin, K. K. Ang, C. Guan, and C. Wang, "Optimum spatio-spectral filtering network for brain–computer interface," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 52–63, Jan. 2011.

[27] R. Tomioka, G. Dornhege, G. Nolte, B. Blankertz, K. Aihara, and K.-R. Müller, "Spectrally weighted common spatial pattern algorithm for single trial EEG classification," Dept. Math. Eng., Univ. Tokyo, Tokyo, Japan, Tech. Rep., 2006, vol. 40.

[28] J. Meng, L. Yao, X. Sheng, D. Zhang, and X. Zhu, "Simultaneously optimizing spatial spectral features based on mutual information for EEG classification," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 1, pp. 227–240, Jan. 2015.

[29] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1865–1873, Jun. 2011.

[30] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain–computer interfacing," *J. Neural Eng.*, vol. 9, no. 2, 2012, Art. no. 026013.

[31] W. Samek, M. Kawanabe, and K.-R. Müller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 50–72, 2013.

[32] W. Wu, Z. Chen, X. Gao, Y. Li, E. N. Brown, and S. Gao, "Probabilistic common spatial patterns for multichannel EEG analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 639–653, Mar. 2015.

[33] F. Qi, Y. Li, and W. Wu, "RSTFC: A novel algorithm for spatio-temporal filtering and classification of single-trial EEG," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3070–3082, Dec. 2015.

[34] 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, no. 1, pp. 15–29, Jan. 2016.

[35] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry," *Expert Syst. Appl.*, vol. 95, pp. 201–211, Apr. 2018.

[36] J. Jiang, Z. Zhou, E. Yin, Y. Yu, Y. Liu, and D. Hu, "A novel Morse code-inspired method for multiclass motor imagery brain–computer interface (BCI) design," *Comput. Biol. Med.*, vol. 66, pp. 11–19, Nov. 2015.

[37] R. M. Mehmood, R. Du, and H. J. Lee, "Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors," *IEEE Access*, vol. 5, pp. 14797–14806, 2017.

[38] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[39] L. Sun, B. Jin, H. Yang, J. Tong, C. Liu, and H. Xiong, "Unsupervised EEG feature extraction based on echo state network," *Inf. Sci.*, vol. 475, pp. 1–17, Feb. 2019.

[40] N. Michielli, U. R. Acharya, and F. Molinari, "Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals," *Comput. Biol. Med.*, vol. 106, pp. 71–81, Mar. 2019.

[41] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, Feb. 2017, Art. no. 016003. doi: 10.1088/1741-2560/14/1/016003.

[42] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 758–769, Apr. 2018.

[43] V. Lawhern, A. J. Solon, N. 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, Nov. 2016, Art. no. 056013. doi: 10.1088/1741-2552/aace8c.

[44] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.

[45] 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. 2018.

[46] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," 2015, *arXiv:1511.06448*. [Online]. Available: http://arxiv.org/abs/1511.06448

[47] H. Dose, J. S. Møller, H. K. Iversen, and S. Puthusserypady, "An end-to-end deep learning approach to MI-EEG signal classification for BCIs," *Expert Syst. Appl.*, vol. 114, pp. 532–542, Dec. 2018. doi: 10.1016/j.eswa.2018.08.031.

[48] D. Zhang, L. Yao, K. Chen, and J. Monaghan, "A convolutional recurrent attention model for subject-independent EEG signal analysis," *IEEE Signal Process. Lett.*, vol. 26, no. 5, pp. 715–719, May 2019.

[49] 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, Jan. 2019.

[50] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-view deep learning framework for EEG seizure detection," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 83–94, Sep. 2019.

[51] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, no. 1, p. 55, 2012.

[52] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. CVPR*, Honolulu, Hawaii, Jul. 2017, pp. 4700–4708.

[53] M. Le van Quyen, J. Foucher, J. P. Lauchaux, E. Rodrigues, A. Lutz, J. Martinerrie, and F. J. Varela, "Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony," *J. Neurosci. Methods*, vol. 111, no. 2, pp. 83–98, 2001.

[54] H. Bashashati, R. Ward, and A. Bashashati, "User-customized brain computer interfaces using Bayesian optimization," *J. Neural Eng.*, vol. 13, no. 2, 2016, Art. no. 026001. doi: 10.1088/1741-2560/13/2/026001.
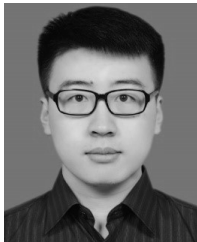
DONGLIN LI was born in Liaoning, China, in 1994. He received the B.S. degree in automation from the Shenyang University of Technology, Shenyang, China, in 2017. He is currently pursuing the Ph.D. degree in control science and engineering with the School of Information Science and Engineering, Northeastern University, Shenyang. His research interests include deep learning, pattern recognition, intelligent control, and brain–computer interface technology.

JIANHUI WANG was born in Liaoning, China, in 1957. She received the B.S., M.S., and Ph.D. degrees in electrical engineering from Northeastern University, Shenyang, China, in 1982, 1986, and 1999, respectively, where she is currently a Full Professor and a Doctoral Supervisor with the School of Information Science and Engineering. Her research interests include complex industrial process modeling, control and optimization, rehabilitation robots, and intelligent control. Her current research interests include intelligent control theory and its applications.

JIACAN XU was born in Liaoning, China, in 1988. She received the B.S. degree in automation from the Shenyang University of Technology, Shenyang, China, in 2010, and the M.S. degree in biomedical engineering from Northeastern University, Shenyang, China, in 2012, where she is currently pursuing the Ph.D. degree in control theory and control engineering. Her research interests include deep learning, pattern recognition, intelligent control, and brain–computer interface technology.

XIAOKE FANG received the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2005, where she is currently a Lecturer with the School of Information Science and Engineering. She was a Visiting Scholar of science and engineering with The University of Manchester, U.K., in 2006. Her research interests include intelligent control theory and its applications and advanced control technology in networks.

. . .