Biomedical Signal Processing and Control 58 (2020) 101845

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Temporal-spatial-frequency depth extraction of brain-computer interface based on mental tasks

Li Wang [a,*], Weijian Huang [a], Zhao Yang [a], Chun Zhang [b]

[a] School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou 510006, China
[b] School of Electronic Science and Engineering, Southeast University, Nanjing 210096, China

ARTICLE INFO

Article history:
Received 14 October 2019
Received in revised form 3 December 2019
Accepted 1 January 2020

Keywords:
Brain-computer interface (BCI)
Electroencephalogram (EEG)
Temporal-spatial-frequency
Convolutional neural network (CNN)
Long short term memory (LSTM)

ABSTRACT

With the help of brain-computer interface (BCI) systems, the electroencephalography (EEG) signals can be translated into control commands. It is rare to extract temporal-spatial-frequency features of the EEG signals at the same time by conventional deep neural networks. In this study, two types of series and parallel structures are proposed by combining convolutional neural network (CNN) and long short term memory (LSTM). The frequency and spatial features of EEG are extracted by CNN, and the temporal features are extracted by LSTM. The EEG signals of mental tasks with speech imagery are extracted and classified by these architectures. In addition, the proposed methods are further validated by the 2008 BCI competition IV-2a EEG data set, and its mental task is motor imagery. The series structure with compact CNN obtains the best results for two data sets. Compared with the algorithms of other literatures, our proposed method achieves the best result. Better classification results can be obtained by designing the well structured deep neural network.

1. Introduction

With a brain-computer interface (BCI) system, users can directly connect with their surrounding environment via brain activity. The new pathway is independent of users' peripheral nerves and muscles [1]. The BCIs can help some patients (such as amyotrophic lateral sclerosis (ALS)) communicate with the outside world. The control of wearable robot has been a successful application example [2]. As a novel way of communication and entertainment equipment, the BCIs can be also introduced into the daily life of healthy people [3]. Considering some factors such as measuring environment, instrument cost and experimenter safety, the control signals of BCIs are mainly derived from electroencephalography (EEG) signals [4]. The EEG signals are time series with frequency characteristics, and they are produced by the thinking activities of the human brain. The BCIs based on EEG have been gradually developed into a variety of applications. Besides the wearable robot [2], the BCIs can be used to control an embedded web server application in the home environment [3]. Quadcopters and virtual helicopters are successfully controlled by the BCIs based on motor imagery [5].

There are mainly several brain activities to induce and evoke EEG signals for the BCIs [5], e.g. motor imagery, steady state visually evoked potentials (SSVEP) and P300 evoked potentials etc. Motor imagery is an active experimental paradigm, and it allows users to control the equipment as they imagine. When the users perform motor imagery, the EEG signals from the sensorimotor cortex of ipsilateral brain will be enhance. The phenomenon is called event-related synchronization (ERS). At the same time, the EEG signals from the sensorimotor cortex of contralateral brain will be reduced, which is called event-related desynchronization (ERD) [6,7]. With these phenomena, motor imagery has four recognizable operations (right hand, left hand, tongue and foot) [8]. Compared with motor imagery, SSVEP and P300 evoked potentials have higher classification results and information transmission rates [9]. However, SSVEP and P300 evoked potentials are passive experimental paradigms. To produce SSVEP and P300 evoked potentials, additional equipment is needed, so they are very inconvenient to use. In particular, uses are prone to visual fatigue after long-term use. Considering the above reasons, one of the data sets studied in this paper is from motor imagery.

To further improve the practicality and stability of BCIs, some new experimental paradigms have been proposed, such as speech imagery. Based on electrocorticography (ECoG), a new neural decoder is used to control the BCIs by speech imagery [10]. Furthermore, as the EEG-based BCI, vowel speech imagery is designed

* Corresponding author.
E-mail address: wangli@gzhu.edu.cn (L. Wang).

https://doi.org/10.1016/j.bspc.2020.101845
1746-8094/

with reading /a/ and /u/ silently [11]. Various categories of speech imagery are constantly proposed, such as vowels, short words and long words. The difference between sound and word makes the classification more accurate [12]. In our previous research, the speech imagery according to Chinese characters is proposed. The classification results between speech imagery and idle state are satisfactory [13]. We also find that the results of mental tasks can be improved with simultaneously performing speech imagery [14]. The stability of BCIs can be effectively improved by this method, so it is selected as another data set for analysis in this paper.

Before EEG signals can be used to control the BCIs, they should be further processed. The process is mainly divided into two steps: feature extraction and feature classification. Similar to the experimental paradigm of motor imagery, the EEG signals can be actively changed when subjects perform mental tasks with speech imagery. Therefore, their signal processing methods can draw lessons from motor imagery. The main difference between the two data sets is the activation of different cerebral cortex. Therefore, the two data sets have different spatial features. To fully test the performance of our proposed methods, we select the two data sets to verify together. As time series, the EEG signals are also distributed with useful features in both spatial and frequency domains. It is helpful to analyze the EEG signals by jointly extracting their temporal, spatial and frequency features. After this, the classification accuracy can be improved [15]. Common spatial pattern (CSP) is a state-of-the-art spatial feature extraction algorithm. Based on CSP, temporal-spatial-frequency feature extraction is developed step by step [16,17]. The frequency characteristics are obtained by wavelet transform, and then they are combined with the spatial features of CSP to improve the classification accuracy [2]. Compared to the single features, the spatial features combined with phase synchronization information also make the classification accuracy higher [18]. After optimizing the filter range by a unified Fisher's ratio, the features extracted by CSP are more efficiently. The classification results are also significantly improved [19]. As an extension of the frequency domain, another famous improved approach of CSP is filter bank common spatial pattern algorithm (FBCSP) [20]. With a bank of band-pass filters, the optimal spatial-frequency features are effectively constructed by FBCSP. The effect of feature extraction of CSP can be further improved by optimizing the frequency-temporal-spatial features at the same time [15]. After selecting the time period and frequency band related to the task, the spatial features are extracted by CSP [17]. The classification effect of the above method is significantly better than the method of only extracting the spatial features or extracting the spatial and frequency features. After feature extraction, feature classification is the final critical step. With unique advantages to classify the features with less data and high dimension, support vector machine (SVM) is often selected to classify the features of EEG [21].

The above traditional methods separate the feature extraction from the feature classification. The matching between two processing steps may not always achieve the best result. Besides, the matching processes are time-consuming and they highly dependent on researchers' experience. Recently, deep neural networks (DNN) have achieved better results than traditional methods in the fields of image classification, video analysis, natural language processing and EEG signals processing [22]. From the research of artificial neural network, the structure of DNN has multiple hidden layers. Without any priori feature extraction and selection, DNN can directly obtain the results from end-to-end learning. Several architectures of DNN have been proposed, including convolutional neural network (CNN), recurrent neural network (RNN), deep Boltzmann machine (DBM) and deep belief network (DBN) etc [23–25]. These architectures usually require a large amount of training data to fit their huge number of hyperparameters. However, limited by the experiments, the amount of EEG data is often relatively small. When P300 evoked potentials are identified by CNN, Batch Normalization is used in the input and convolutional layers to alleviate overfitting [23]. Depthwise and separable convolutions are constructed as EEGNet, and the spatial and frequency features can be extracted from the limited training data [26]. The temporal features of the signals can be extracted by RNN. As an improved version of RNN, long short term memory (LSTM) is more widely used. After the decomposition of discrete wavelet transform, the EEG signals are extracted by a deep BLSTM-LSTM network [27].

The above DNN models only have one type of network. In order to improve the recognition effect of the network, different types of networks are combined together. One of the combinations is CNN and LSTM. Bashivan et al. propose recurrent-convolutional neural networks (RCNNs) with CNN and LSTM [28]. To extract the spectral and spatial features, EEG signals are transformed into 2-D topologypreserving multi-spectral images at first. The robust representations are learned by RCNNs from the sequence of images. Their experimental results illustrate the effectiveness of the method. However, the process of converting to an image takes a certain amount of time, and it is not conducive to the establishment of an online BCI. Xie et al. propose a CNN-LSTM model to decode the finger trajectory from ECoG [29]. For CNN, spatial and temporal filtering is applied by a spatial convolution layer and a temporal convolution layer, respectively. The temporal dynamics of the signals can be captured by LSTM. Compared with the conventional regression methods, their method gives a prediction with higher correlation coeffcient. But the signal their model deals with is ECoG, whose signal-to-noise ratio is much higher than that of EEG. To identify intracortical data, Schwemmer et al. propose a deep neural network decoding framework [30]. The signals are first transformed into 2-D images by wavelet transform. A LSTM layer and a convolutional layer are used to extract the characteristics of images in turn. The results show that deep neural network decoders can be used in the BCI technology. Their method also includes an image conversion process, and the complexity of the model has increased. Zhang et al. propose cascade and parallel CNN-LSTM models [31]. According to the channel distribution, EEG signals are converted to 2-D EEG data meshes. The spatial and temporal features are extracted by CNN and LSTM, respectively. Their method outperforms state-of-the-art models. However, they ignore the frequency features of EEG.

EEG signals can be extracted with three different features, and they come from three domains: time, frequency and space. These three features are more often extracted by conventional methods, but they are rarely extracted by conventional DNN at the same time. The classification accuracy of EEG can be improved by fully extracting the three features. Moreover, different extraction order may lead to different classification results. The existing CNN-LSTM models have too many parameters. When the amount of data is small, these models are easy to overfit. With too many parameters, these models also need more time to train. This limits these models' use for the online BCI. Therefore, a novel DNN architecture should be built to efficiently extract the temporal-spatial-frequency features of the signal. The above content is the motivation of this paper.

There are three advantages of our idea. Firstly, in order to further improve the classification accuracy of EEG signals, novel DNN architectures are proposed to jointly extract the temporal-spatial-frequency features. The frequency and spatial features of EEG signals can be extracted by CNN, and the temporal features are extracted by LSTM. Secondly, to compare the classification effects of different extraction orders, two series and parallel structures with CNN and LSTM are proposed, respectively. It is found that series structures are superior to parallel structures, and the reason is also analyzed. Finally, our best model has fewer parameters. Compared with conventional CNN, the number of parameters of compact CNN is relatively small. It has the advantages of higher classification

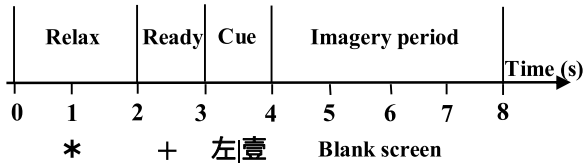

Fig. 1. Timing of a trial of the training paradigm for our own data set.

accuracy and faster training speed. In future studies, it will be beneficial to apply this model to the online BCIs. The above contents are also the contribution of this paper.

The following section introduces the method of acquisition for the EEG signals of two types of data sets, and the architectures of our proposed deep learning models. Section 3 gives the results of analysis and classification. The further analysis about experimental results is discussed in Section 4. Section 5 concludes the paper.

2. Methods

2.1. Data sets

1) Our own data set

There are ten subjects (seven males and three females, right handed students with good health and vision correction) in our own data set [14]. To facilitate subsequent analysis, they are labeled as S1-S10, respectively. Their age ranges from 22 to 28, with the average of 23.6 years. Before the experiment, subjects must rest fully. Coffee, tea or alcohol is not allowed to drink. Seven of them have the experiences about the experiment of speech imagery before, but none of them have attended the experiment about imaging the body rotate or visualizing the Chinese strokes. They sit and finish the experiment with their hands relaxed naturally. By looking at a 22 inch LCD screen, they get hints of the experiment. The Academic Ethics Committee of Southeast University permits the experimental protocol. After comprehending the purpose of the experiment and relevant considerations, all subjects sign Informed Consent.

Without feedback, the experimental paradigm of this data set comes from the first step of the experiment in Reference [14]. It is the mental tasks with speech imagery. As shown in Fig. 1, each trial starts with a fixed asterisk for two seconds, and subjects can have a rest. Then, with a fixation cross, it is a ready period for one second. After that, it is a "Cue" for one second. When the "Cue" is "左 (left)", subjects image rotating their bodies to the left with reading the Chinese character silently. When the "Cue" is "壹 (one)", they visualize writing the strokes with reading it silently. Subjects perform the corresponding task during the imaginary period. With keeping their bodies still and silent, they need to repeat the same task over and over again. Between each trial, they can take short breaks. In each run, the two cues are displayed fifteen times at random. There are five runs in the entire experiment. Therefore, over the course of the experiment, there are 75 imagery periods of each cue.

Imagining the body rotation is associated with motor cortex. Silent reading is a kind of language activity. Additionally, with involving both language and motor tasks, it is relatively complicated to imagine writing a Chinese character. Therefore, to record more comprehensive EEG signals, the primary motor area, Broca's area, Wernicke's area, and superior parietal lobule of cerebral cortex should be covered by an electrode cap. With 35 channels, the cap corresponding to the expanded version of international 10–20 system is shown in Fig. 2. To collect EEG signals, the equipment is a SynAmps 2 system (Neuroscan Co., Ltd.). When recording experimental data, it is easy to introduce the interference of electrooculogram (EOG). To record the EOG signals, two bipolar channels are placed around the eyes. In order to reduce the power

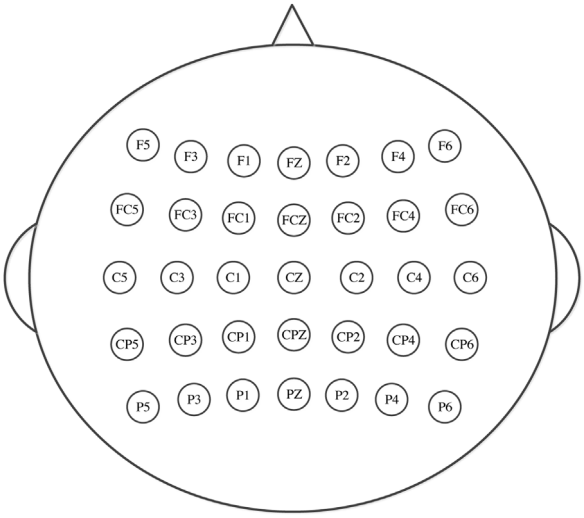

Fig. 2. Channel positions of the EEG setup for our own data set. 35 channels are distributed over the scalp according to the expanded version of international 10–20 system.

frequency signal interference, a grounding channel is attached near the forehead. On the center of the cerebral cortex, a reference channel is attached. The impedance of all channels must remain below 5kΩ. The signal sampling frequency is set as 250 Hz. The band-pass filter of the equipment is set as 0.1−100 Hz.

2) The 2008 BCI competition IV-2a EEG data set

In order to fully verify the algorithms in this paper, the 2008 BCI competition IV-2a EEG data set is also used [32]. It is provided by the Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz. This data set has four types of motor imagery, and they are left hand, right hand, feet and tongue, respectively. The EEG signals are recorded from 22 Ag/AgCl channels, and 3 electrooculogram (EOG) channels are also recorded. The sampling rate is 250 Hz, and the band-pass filter is set as 0.5−100 Hz. 9 subjects participate in the data collection experiment, and they are labeled as A1-A9, respectively. Every experiment has two sessions. The first session is for training, and the second is for testing. Each session has 72 trials for each type of imagination, and it results in 288 samples per session. As shown in Fig. 3, the timing scheme consists of a fixation cross of 2 s, and a cue of 1.25 s. After the cue, it is followed by a period of motor imagery of 3 s. When the cue appears, the subjects begin to imagine until the end of the imaginary period.

2.2. Preprocessing

The two data sets have different acquisition channels and amplifiers, so their preprocessing is a little different. For our own data, the EEG signals from the frontal lobe are more susceptible by EOG interference, so the EOG signals are removed by SCAN 4.5 at first. The useful information of EEG is mainly distributed in θ (4−8 Hz), $\alpha 1$ (8−10 Hz), $\alpha 2$ (10−13 Hz), $\beta 1$ (13−20 Hz) and $\beta 2$ (20−30 Hz) waves [33]. In order to improve the signal-to-noise ratio, the EEG signals of two data sets are both filtered by the 6th order Butterworth band-pass filter. The filter range is 4−35 Hz.

2.3. Deep Neural Networks (DNN)

1) Compact Convolutional Neural Network

CNN is a type of artificial neural networks, and its structure is the multilayer perceptron [23]. Inspired by the working principle of the visual cortex, the convolutional layers are introduced into CNN.

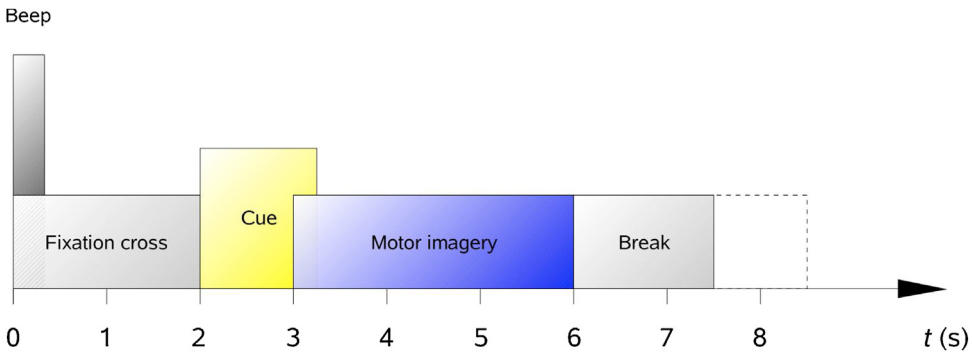

Fig. 3. Timing scheme of the paradigm of the 2008 BCI competition IV-2a EEG data [32].

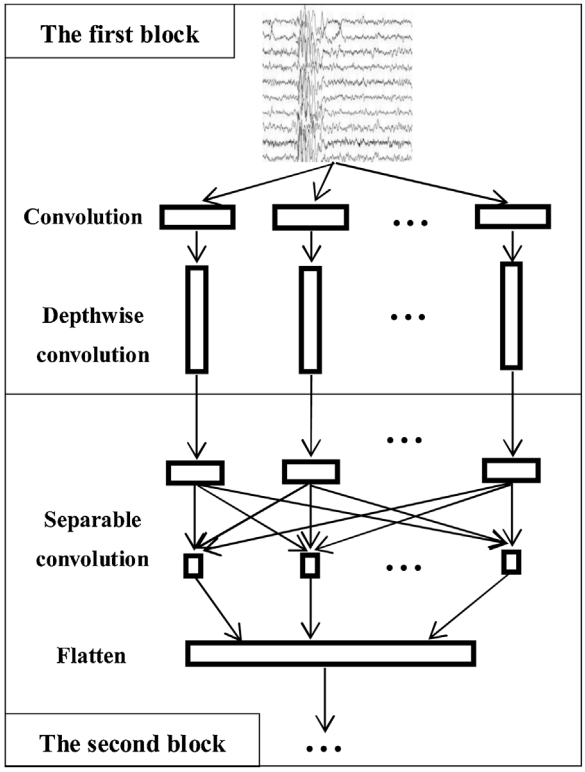

Fig. 4. Overall visualization of the compact CNN architecture.

Weight-sharing and sparse connectivity are the advantages of the convolutional layers. The two advantages can significantly reduce the computational complexity. Unlike images and video having a lot of data to train CNN, the amount of data on EEG signals is very small. For classifying the EEG signals, too many convolutional layers of CNN can easily lead to over-fitting of the training model. Therefore, it is very important to construct a suitable CNN model. Compact CNN is a special CNN with depthwise and separable convolutions, and it has fewer parameters [26]. Additionally, with a little preprocessing, compact CNN can directly extract spatial and frequency features from raw data.

As shown in Fig. 4, the structure of compact CNN has two blocks. In the first block, two convolutional steps are performed in sequence. For ease of software implementation, 2D convolutional functions are used. Eight 2D convolutional filters are fitted to capture frequency information at first, and the filter length is 64. Different band-pass frequencies of EEG are contained in the outputting of feature maps. Depthwise convolutions are used to acquire spatial features of the EEG signals. To obtain a certain number of spatial filters, the depth parameter is set as 2. Therefore, two spatial filters are followed each frequency feature map, and frequency-spatial features of the EEG signals are extracted. Along each feature map dimension, Batch Normalization is applied. The activation function is exponential linear unit (ELU). The expression of ELU is [34]

$$f(x) = \begin{cases} x, x > 0 \\ \alpha(\exp(x) - 1), x \le 0 \end{cases} \quad (1)$$

x is the input eigenvalue, and α is the hyperparameter to control the saturation of ELU for negative net inputs.

The sampling rate of the EEG signals is reduced by an average pooling layer with size $(1, 4)$. In order to prevent over-fitting in the training process, dropout technique is used at the last step of the first block. The dropout probability is set as 0.5. In the second block, the first step is separable convolutions. A separable convolution is made up of a depthwise convolution (size $(1, 32)$) followed by eight pointwise convolutions. The relationship within and across feature maps can be decoupled, and then they are optimally merged as outputs. Batch Normalization, average pooling layer (size $(1, 8)$) and dropout technique (the dropout probability is 0.5) are also executed in sequence. ELU is selected as the activation function in this block, too. After a flatten layer, the eigenvalues are input to a fully connected (FC) layer at last. The final extracted eigenvalues can be used for classification or for further feature extraction of other neural networks. The specific manipulation depends on the needs, so it is indicated by an ellipsis at the bottom of Fig. 4. When it comes to classify directly, the activation function is softmax. The input eigenvector $\boldsymbol{z}$ can be squashed between 0 and 1 by softmax function [35]:

$$output_j = \frac{\exp(z_j)}{\sum_{k=1}^{K} \exp(z_k)}, j = 1, 2, \ldots, K \quad (2)$$

The number of units in the output of model is K, and it is equal to the number of classes. As a classifier, the description of the architecture of compact CNN is shown in Table 1.

2) Shallow CNN

Besides the compact CNN, other structures of CNN can also extract the frequency and spatial features of the EEG signals. To compare the effects of feature extraction of different CNN, a shallow CNN is used [36]. The shallow CNN is inspired by the FBCSP, and it consists of two layers of convolution with 40 units. The fist layer is a temporal convolution, and it is proposed to band-pass filter the EEG signals of each channel. The second layer is a spatial filter, and it achieves the convolution across the channels. The kernel size of temporal convolution is 25, and a larger kernel size means a larger range of transformations. After the convolution layers, it is followed by Batch Normalization, a squaring nonlinearity and an average pooling layer. The activation function is logarithmic. Dropout technique is used, and the dropout probability is also

Table 1
The architecture of compact CNN as a classifier. Our own data set is selected as an example.

Layer	#flters	Input	Size	#params	Output	Activation
Reshape		35 × 200	(35, 200, 1)		35 × 200 × 1	
Conv2D	8	35 × 200 × 1	(1, 64)	512	35 × 200 × 8	Linear
BatchNorm		35 × 200 × 8		140	35 × 200 × 8	
DepthwiseConv2D	16	35 × 200 × 8	(35, 1)	560	1 × 200 × 16	Linear
BatchNorm		1 × 200 × 16		4	1 × 200 × 16	
Activation		1 × 200 × 16			1 × 200 × 16	ELU
AveragePool2D		1 × 200 × 16	(1, 4)		1 × 50 × 16	
Dropout		1 × 50 × 16			1 × 50 × 16	
SeparableConv2D	8	1 × 50 × 16	(1, 16)	640	1 × 50 × 8	Linear
BatchNorm		1 × 50 × 8		4	1 × 50 × 8	
Activation		1 × 50 × 8			1 × 50 × 8	ELU
AveragePool2D		1 × 50 × 8	(1, 8)		1 × 6×8	
Dropout		1 × 6×8			1 × 6×8	
Flatten		1 × 6×8			48	
Dense	96	48		98	2	Softmax

Table 2
The architecture of shallow CNN as a classifier. The 'Square' is given as $f(x)=x^2$, and 'Log' is $f(x)=\log(x)$. Our own data set is selected as an example.

Layer	#flters	Input	Size	#params	Output	Activation
Reshape		35 × 200	(35, 200, 1)		35 × 200 × 1	
Conv2D	40	35 × 200 × 1	(1, 25)	1040	35 × 176 × 40	Linear
Conv2D	40	35 × 176 × 40	(35, 1)	56000	1 × 176 × 40	Linear
BatchNorm		1 × 176 × 40		4	1 × 176 × 40	
Activation		1 × 176 × 40			1 × 176 × 40	Square
AveragePool2D		1 × 176 × 40	(1, 35), stride (1, 7)		1 × 21 × 40	
Activation		1 × 21 × 40			1 × 21 × 40	Log
Dropout		1 × 21 × 40			1 × 21 × 40	
Flatten		1 × 21 × 40			840	
Dense	2	840		1682	2	Softmax

set as 0.5. The output is flattened to 1-D and then sent to a fully connected layer. The final extracted eigenvalues can be also used for classification or for further feature extraction of other neural networks. When it comes to classify, the activation function is also softmax. All the computational steps are embedded in the shallow CNN, so all parameters of each layer can be jointly optimized. As a classifier, the description of the architecture of shallow CNN is shown in Table 2.

3) Long Short Term Memory

RNN is a kind of recursive neural network. Taking sequence data as input, it gets recursion in the direction of sequence evolution. All circulation units are linked by chains to form a closed loop. However, it is a challenge to train RNN by gradient descent algorithm. For an earlier network, RNN is insensitive to inputs. To address the long-term dependence issue, LSTM is proposed as an improved version of RNN [29]. LSTM can more effectively extract the temporal features of the EEG signals. Three control cells are added in LSTM, and they are input gate, output gate and forget gate, respectively. They can be described by these equations [29]:

$$i = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{3}$$

$$f = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{4}$$

$$o = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{5}$$

$$\tilde{c} = W_c x_t + U_c h_{t-1} + b_c \tag{6}$$

$$c_t = f \odot c_{t-1} + i \odot \tilde{c} \tag{7}$$

$$h_t = o \odot c_t \tag{8}$$

$$\sigma(x) = 1/(1 + \exp(-x)) \tag{9}$$

To control each gate, W, U and b are set as learnable parameters. x, h, i, f, o and c are input, output, input gate, forget gate, output gate and memory cell state, respectively. $\odot$ represents element-wise product. t represents the data as the time series. As the eigenvalues enter the model, the cells of LSTM can judge them. The eigenvalues that conform to the rules are left behind, while the eigenvalues that do not conform are forgotten. Based on this principle, the problem of long-term dependence in RNN can be solved. The constructed LSTM network has two stacked layers in this paper. In each layer, the number of units corresponds to the number of sliding time windows. The two layer structure is a compromise between performance and resource. The output of the last time step of LSTM is fed into the fully connected layer to get the classification results. The activation function is softmax.

4) Series Convolutional Recurrent Neural Network

After extracting the temporal-spatial-frequency features, the useful information of EEG signals can be fully exploited. The goal of our research is to improve classification accuracy with comprehensive feature extraction. Therefore, how to combine these three features is the key. In order to achieve the research goal, series and parallel convolutional recurrent neural network frameworks are designed and compared. The series structure is shown in Fig. 5. The frequency and spatial features of the filtered EEG signals are first extracted by CNN, and then the sequences of the extracted features are fed into LSTM to extract temporal features. The output of the last time step of LSTM layers is transported to a fully connected layer. A softmax classifier obtains the final prediction at last. To form the series convolutional recurrent neural network with LSTM, compact CNN and shallow CNN are respectively used as the CNN module. They are labeled as series compact convolutional recurrent neural network (SCCRNN) and series shallow convolutional recurrent neural network (SSCRNN), respectively.

5) Parallel Convolutional Recurrent Neural Network

The series structure can extract different features step by step, and the effect of temporal feature extraction is affected by the previous step. To independently extract the temporal features, a parallel convolutional recurrent neural network is also proposed

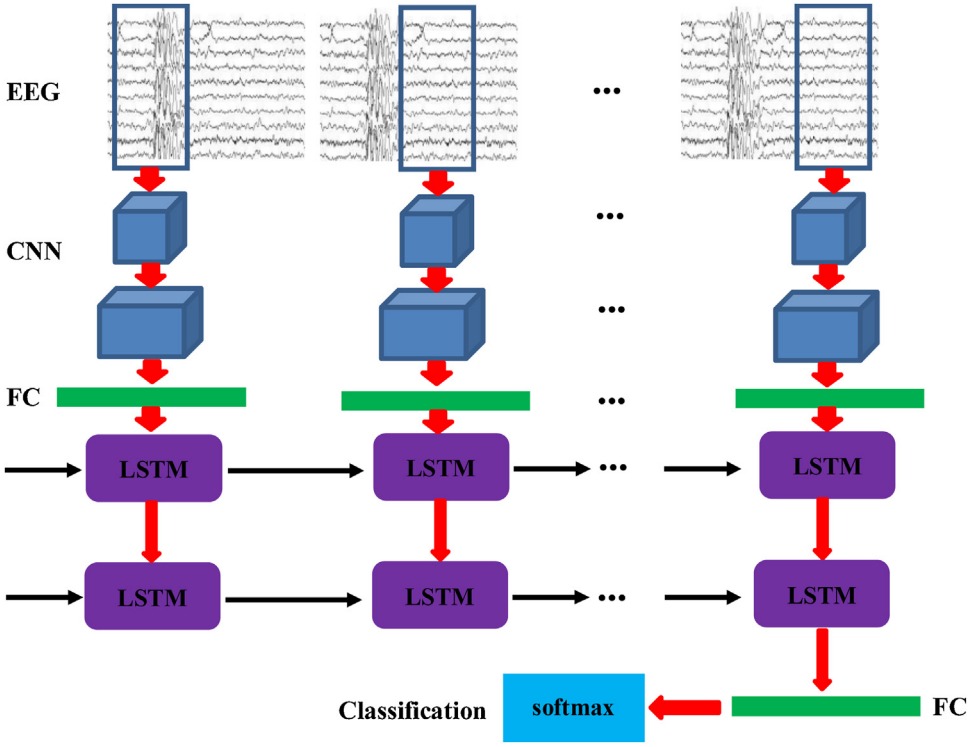

Fig. 5. Series convolutional recurrent neural network architecture.

in Fig. 6. The parallel structure also contains CNN and LSTM. Unlike the series structure, it extracts the features in parallel. After passing through their own full connection layers, the two extracted features are combined together. Finally, they are sent to a softmax classifier. To form the parallel structure with LSTM, compact CNN and shallow CNN are also respectively used as the CNN module. They are labeled as parallel compact convolutional recurrent neural network (PCCRNN) and parallel shallow convolutional recurrent neural network (PSCRNN), respectively.

3. Results

Our own data has two kinds of signals, and the imagery period is four seconds for each kind. Therefore, there are four seconds of EEG signals for analysis, and they come from 4 s to 8 s of every trial in Fig. 1. The four seconds data has 1000 points, and they are evenly divided into five temporal segments without overlap. Therefore, each data segment has 200 points. There are 75 imagery periods for each task. After interception, each task has 375 pieces of data. The 2008 BCI competition IV-2a data set has four types of motor imagery. For more reasonable analysis of the proposed algorithms, these four types of imagination are grouped in pairs at first. They can be combined into six groups, including left hand vs right hand, left hand vs feet, left hand vs tongue, right hand vs feet, right hand vs tongue, and feet vs tongue. After analyzing the two classifications, the best algorithm will be selected to further analyze the four classifications. The final results will be compared with the algorithms of other literatures. In the public data set, data loss occurs at individual channels in very few trials of some subjects. To keep the number of trials consistent for each subject, the channels lost data are removed, rather than the trial being removed. For the 2008 BCI competition IV-2a EEG data set, the subjects begin to imagine as soon as the cue appears. After imaging their bodies movement, four seconds data is generated. The data is also evenly divided into five temporal segments without overlap. Because the two data sets have the same sampling rate, each segment of the public data set also has 200 points. Finally, each task has 720 pieces of data.

After band-pass filtering, the EEG signals are directly analyzed by DNN. With Python 3.5.2 using the Keras API in TensorFlow, all models of DNN are trained on one NVIDIA GTX 1080 Ti GPU. The models are trained with an Adam optimizer for 200 epochs, and the learning rate is set as 1×10^{-5}. For 2 and 4 classification problems, loss functions are set as binary crossentropy and categorical crossentropy, respectively. Batch size of all models is 64. The EEG signals have spatial, temporal and frequency features, so it is helpful to improve the classification results by jointly extracting these three features. Four novel architectures (SCCRNN, SSCRNN, PCCRNN and PSCRNN) with the series and parallel structures are proposed to extract the three features, respectively. To compare with the proposed architectures, the classification results of LSTM, compact CNN and shallow CNN are also calculated. Additionally, the results of a traditional method with CSP and SVM are obtained. After calculating by 10×10 cross-validation of the above eight methods, the results of our own data set are shown in Fig. 7.

As shown in Fig. 7, the classification results between the two mental tasks are calculated for ten subjects. Significant differences between the results are analyzed by paired samples t-test. For the eight algorithms, SCCRNN has the highest average accuracy with 85.6 %. It is 2.7 % higher than the compact CNN ($p < 0.05$). It indicates that time features extracted by LSTM have a contribution to the classification results. There is no significant difference between the results of compact CNN and CSP-SVM ($p > 0.05$). The average result of SCCRNN is 4.0 % higher than CSP-SVM ($p < 0.01$). Compared to conventional methods, better results can be obtained by constructing appropriate deep learning models. There are two reasons for the results: firstly, higher classification accuracy can be obtained by jointly optimizing feature extraction and classification algorithms; secondly, only the spatial features of EEG signals are extracted by CSP, and the frequency and temporal features are not necessarily optimal results. Different DNN structures also have dif-

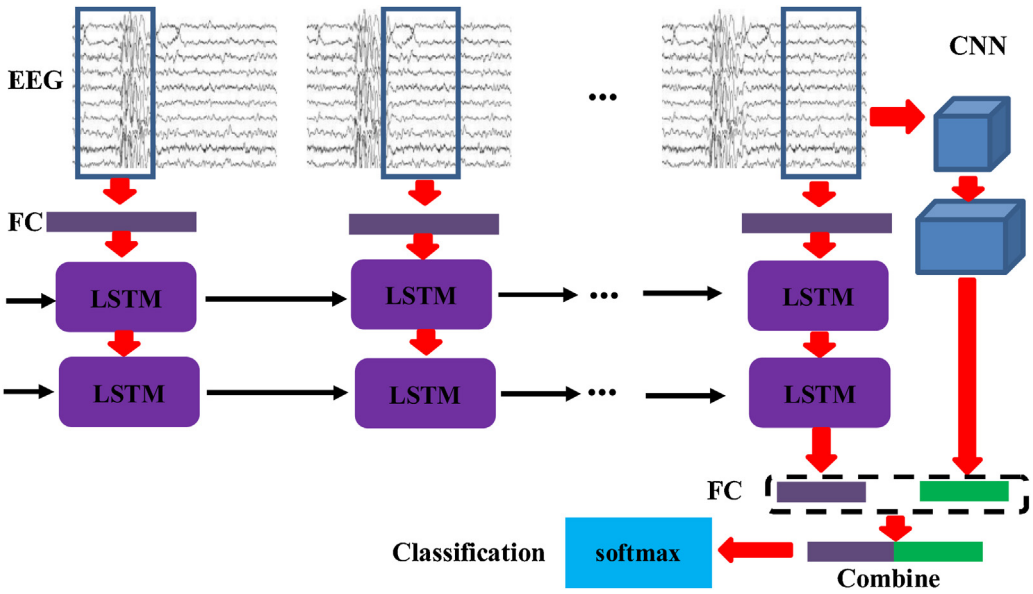

Fig. 6. Parallel convolutional recurrent neural network architecture.

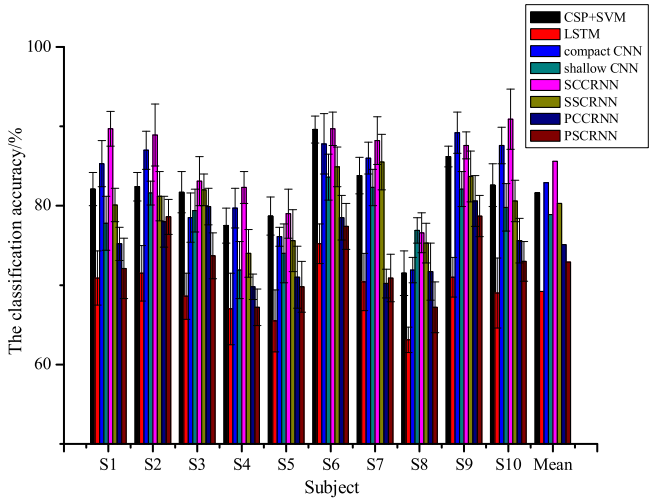

Fig. 7. After 10×10 cross validation, the results of eight methods are calculated for our own data set.

ferent results. The average accuracy of compact CNN is better than shallow CNN ($p < 0.05$). Series and parallel structures with the compact CNN also have better results than those with the shallow CNN (the series structure: $p < 0.01$, and the parallel structure: $p < 0.05$). The classification effects between series and parallel structures are also compared, and the results of two series structures are both better than two parallel structures. The reason is that LSTM does not work well with classifying the EEG signals directly, and its result is only 69.2 %. When the features extracted by LSTM are directly combined with the eigenvalues of CNN, the overall recognition rate will be reduced. The classification results of two types of eigenvalues differ too much, and the eigenvalues with lower results are easily regarded as noise by the classifier. Finally, the final classification results are reduced.

Comparing two types of CNN, the average result of shallow CNN is lower than compact CNN. The reason is that the quantity of training data is small, and it is not easy to reach the optimal solution for shallow CNN. In order to observe the training process of two models of subject S1, the accuracies of training and verification processes are shown in Fig. 8, respectively. After several epochs

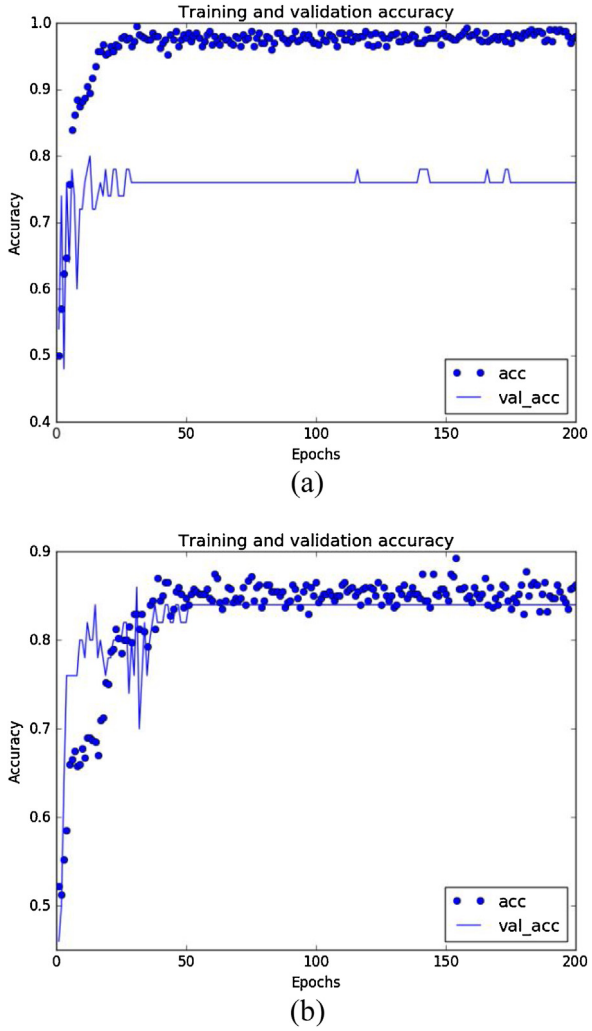

Fig. 8. The training and verification processes of subject S1: (a) shallow CNN; (b) compact CNN.

of training, the training accuracies (acc of Fig. 8) of shallow CNN quickly approach 1. However, the verification accuracies (val_acc of Fig. 8) only reach 0.76 and then stop increasing. There is an obvious phenomenon of over-fitting. The model over-optimizes the training data, and it cannot generalize the data beyond the training set. The trained model of compact CNN has better generalization performance. This is because the structures of shallow CNN and compact CNN are different. Taking our own dataset as an example, the numbers of trainable parameters are 58724 and 1884 for shallow CNN and compact CNN, respectively. In addition to having far fewer training parameters than shallow CNN, the training speed of compact CNN is also faster.

The deep learning models are often referred as black boxes. The representations learned by models are difficult to extract, and the way that they behave is hard to understand. Fortunately, compared with other deep learning models, the representations learned by CNN are more suitable for visualization. To verify the robustness of models, the interpretability of model features is an important manifestation. As a result, various methods have been proposed to interpret and understand CNN [37]. The representations are interpreted by visualizing the filtering layers of CNN in this paper. The compact CNN with better training results is selected for visualization. A random signal (the size of the signal is 35×200) with noise is transformed into a tensor, and the tensor is this shape (1, 35, 200, 1). Then, it is input into the model. Stochastic gradient descent is used to adjust the values of the input signal. A loss function is constructed to maximize the response of the filter in the convolution layer. After passing through each filter, a floating point tensor is obtained with the shape (1, 35, 200, 1). To display the tensor more clearly, the data is converted to the integer within the interval [0, 255]. Visual results of the convolution layer and depthwise convolution layer of compact CNN are obtained for subject S1, and they are shown in Fig. 9 and Fig. 10, respectively.

There are eight convolutional filter patterns in the compact CNN, and they are named conv_1, conv_2, . . ., conv_8 in Fig. 9, respectively. The EEG signals are band-pass filtered by convolutional filters. Each filter has a different filter range, so each subgraph of Fig. 9 has different color along the horizontal axis. For each subgraph, the color along the vertical axis is basically the same. This means the signals of all channels are filtered according to the same filtering range. The depth parameter of the model is 2, so two spatial filters are followed each frequency filter. Therefore, there are 16 filter patterns in the depthwise convolution layer, and they are named depthwise_1, depthwise_2, . . ., depthwise_16 in Fig. 10, respectively. After passing through one of the band-pass filters, the signals are further filtered by two spatial filters. Along the vertical axes of these subgraphs, the values are changed. Large values indicate that the signals of this channel are enhanced, and small values indicate that the signals are weakened. Finally, through these two convolutional layers, frequency-spatial features of the EEG signals are extracted.

In order to fully verify our proposed methods, the results of 2008 BCI competition IV-2a data set are also calculated. For the data set, there are four kinds of motor imagery. The results of six pairs (left hand vs right hand (L vs R), left hand vs feet (L vs F), left hand vs tongue (L vs T), right hand vs feet (R vs F), right hand vs tongue (R vs T), and feet vs tongue (F vs T)) are respectively calculated by the above eight algorithms at first. The results are shown in Fig. 11, and the best algorithm will be selected for further analysis.

According to the classification of six pairs, the results of SCCRNN are also the highest. This result is consistent with our own data set. SCCRNN has a certain adaptability to perform excellent results for different data sets. Except for several inconsistencies, the general situation of the other seven algorithms is basically the same as our own data set. Therefore, the results of the eight algorithms are no longer compared. Taking SCCRNN as an example, the results of six pairs of motor imagery will be mainly analyzed. Among the four types of imagination, there is the greatest difference between imagining the left hand and tongue movement (91.65 %). The results of right hand vs feet and right hand vs tongue are slightly lower, and they are 90 % and 89.8 %, respectively. The result of left hand vs right hand is the lowest (85.9 %), which is even lower than the result of tongue vs feet (87.6 %). For the four categories of motion imagination, left and right hands motor imagery is the most studied. However, the classification accuracy between them is not necessarily the highest. The above results are all greater than the random value (50 %), so motor imagery has application value for BCIs.

Because SCCRNN has the highest classification accuracy among the eight algorithms, it is selected as the next method to calculate the four classifications. Unlike the above results from cross validation, the training sets (the first session) of the 2008 BCI competition IV-2a EEG data set are used to train the SCCRNN. After training, the evaluation sets (the second session) are used to test the model. In addition to accuracy, Cohen's Kappa value is used more for multi-classification [38]. This is because the accuracy is not equal to compare different categories. As a result, the Kappa value is selected to evaluate our method for the four classification algorithm in this paper. The Kappa value is calculated by $(p_o\text{-}p_e)/(1\text{-}p_e)$ [39]. p_o is the classification accuracy and p_e is a random probability (For example, the random probability of two classification is 0.5, and four classification is 0.25). To show whether our proposed method is good or not, it is compared with the methods from other literatures. The results of the comparison are shown in Table 3. To make the results clearer, the maximum Kappa value per row is marked bold. The standard deviations (Std) of the Kappa values of nine subjects are also calculated for all methods.

Compared with the other six methods, the average Kappa value obtained by our proposed method is the highest (0.64). Two subjects (A4 and A6) get the best results after using our method. The Kappa values of five subjects exceed or reach 0.7. A hybrid learning method based on two classifiers is proposed by Raza [40]. With the method, other three subjects obtain the highest Kappa value in Table 3. However, its standard deviation is the largest, which means that his method is the most volatile for different subjects. After being processed by his method, subject A2 has the smallest Kappa value in Table 3. Particularly, the average value of all subjects is much lower than our method. Olivas-Padilla proposes a novel method: the features are extracted by discriminative filter bank common spatial pattern at fisrt, and then CNN is used to further extract and classify the features [43]. Compared with using CNN to directly extract the features of EEG, the results obtained by this method are more stable and excellent (the Kappa value is 0.61, and it is the second highest value in Table 3). The standard deviation of the method of Olivas-Padilla is the lowest. Because it does not extract the time features of EEG, the average value is also lower than our method ($p < 0.05$). Another 0.61 is obtained by the method proposed by Sharbaf, which is lower than our method ($p < 0.05$). FBCSP is proposed by Ang, and it is developed from CSP. FBCSP focuses on subdividing the frequency components to get higher accuracy [38]. However, FBSCP does not take account of the temporal features of EEG, so its Kappa value is the second-lowest of the seven methods. After comparing with conventional methods and deep learning algorithms, the results obtained by our method is the best. A novel series structure is proposed: the frequency and spatial features of EEG signals are extracted by compact CNN at first, and then the temporal features are extracted by LSTM. Therefore, the frequency, spatial and temporal features of EEG can be fully mined, and the classification results are also improved.

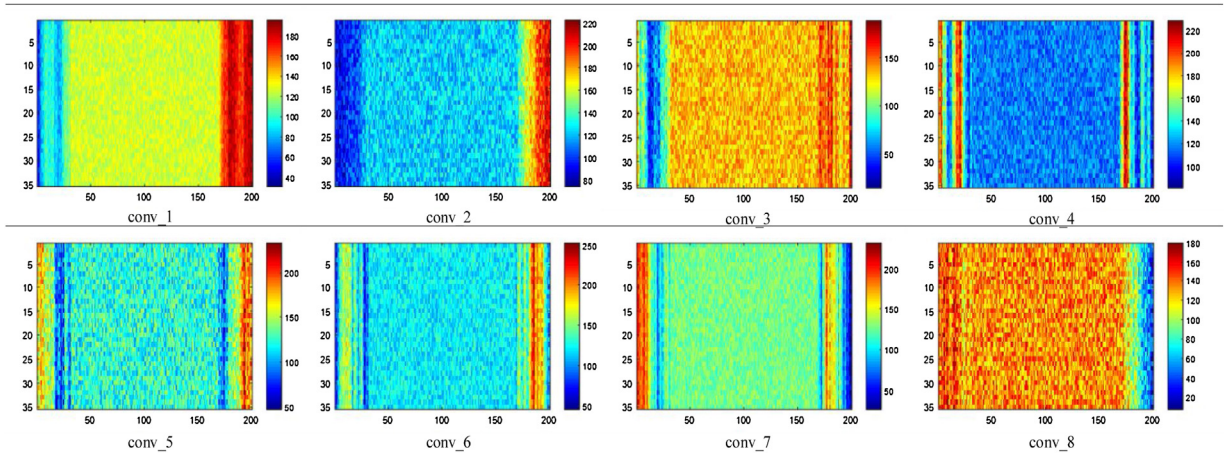

Fig. 9. Visualization of filter patterns of the convolution layer of compact CNN for subject S1.

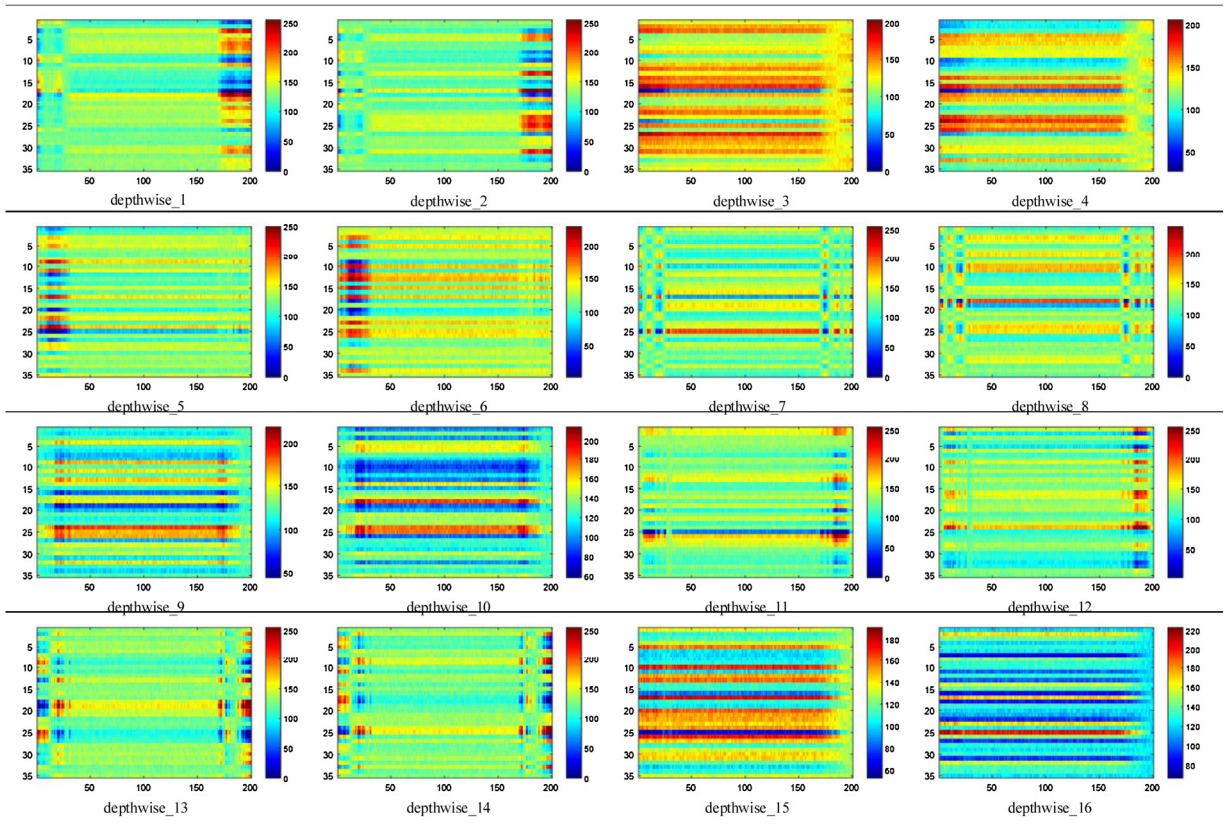

Fig. 10. Visualization of filter patterns of the depthwise convolution layer of compact CNN for subject S1.

Table 3
Our proposed method in comparison to past studies by calculating the Kappa value, and the maximum Kappa value per row is marked bold.

Subject	Ang et al. [38]	Xie et al. [39]	Zeng et al. [40]	Raza et al. [41]	Sharbaf et al. [42]	Olivas-Padilla et al. [43]	Our proposed method
A1	0.68	0.77	0.74	**0.88**	0.75	0.68	0.77
A2	**0.42**	0.33	0.27	0.22	0.31	0.36	0.38
A3	0.75	0.77	0.77	**0.88**	0.82	0.69	0.75
A4	0.48	0.51	0.43	0.39	0.56	0.62	**0.65**
A5	0.4	0.35	0.29	0.53	0.47	**0.6**	0.54
A6	0.27	0.36	0.27	0.33	0.38	0.45	**0.47**
A7	**0.77**	0.71	0.73	0.38	0.75	0.71	0.76
A8	0.76	0.72	0.77	**0.85**	0.74	0.72	0.78
A9	0.61	**0.83**	0.8	0.81	0.67	0.66	0.7
Mean	0.57	0.59	0.56	0.58	0.61	0.61	**0.64**
Std	0.17	0.19	0.23	0.25	0.17	0.11	0.14

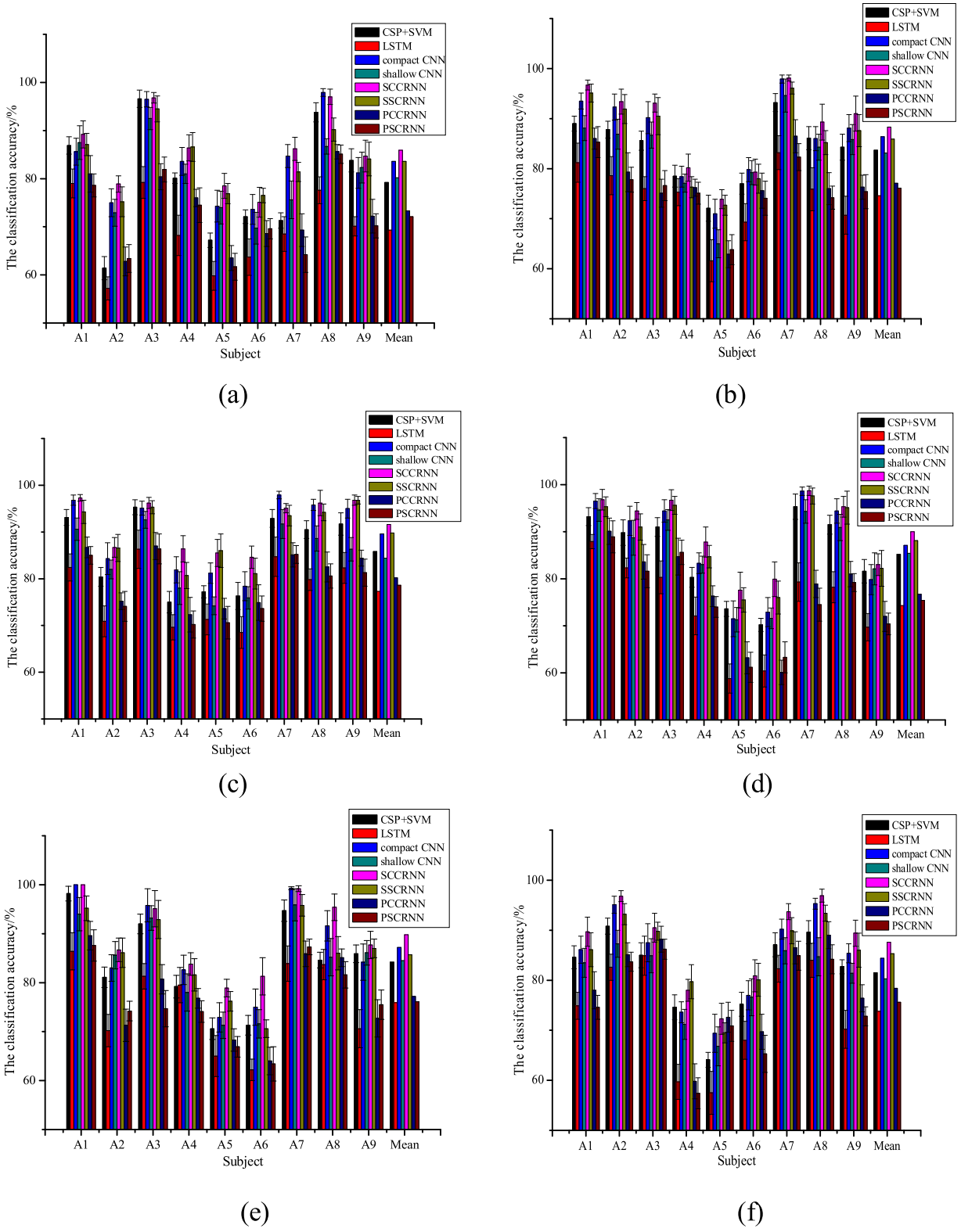

Fig. 11. After 10 × 10 cross validation, the results of eight algorithms are calculated for the 2008 BCI competition IV-2a data set: (a) L vs R; (b) L vs F; (c) L vs T; (d) R vs F; (e) R vs T; (f) F vs T.

4. Discussion

The efficiency of BCIs is often determined by the classification accuracy of the EEG signals. Compared with other CNN-LSTM models, we propose three novelty parts in the DNN models to improve the classification accuracy. Firstly, novel DNN architectures are proposed to jointly extract the temporal-spatial-frequency features. CNN-LSTM models proposed by Zhang only extract the spatial and temporal features of EEG, and they ignored the frequency features [31]. Secondly, series structures are found to be superior to parallel structures. This is due to different extraction sequences. Thirdly, our method makes the models as simple as possible to accommodate the wearable devices in the future. CNN-LSTM models proposed by Bashivan and Schwemmer deal with images [28,30], and our models directly deal with EEG signals. With the addition of image processing, the complexity of the models can be increased. The three features can be extracted by Xie's model [29]. Its structure is similar to the series structure with shallow CNN, whose classification accuracy is not as good as the series structure with compact CNN. In addition, the number of parameters of shallow CNN is much more than that of compact CNN, and this is not conducive to the datasets with less data. Our own data set and the 2008 BCI competition IV-2a EEG data set are both used to validate the proposed models. Their results are compared with the results of CSP + SVM, LSTM, compact CNN and shallow CNN. For both sets of data, the highest results are obtained by SCCRNN. SCCRNN is a series structure with compact CNN and LSTM. At the same time, we compare SCCRNN with the methods from other literatures, and its result is also the best.

Via multiple channels, the EEG signals are electrical signals that can respond to change in the brain over time. Therefore, the EEG signals have useful features in temporal, spatial and frequency domains. The features of these three domains should be simultaneously extracted and analyzed. This process can not only comprehensively analyze the EEG signals, but also improve the classification accuracy of metal tasks. As a traditional method, temporal-spatial-frequency features extraction usually develops on the basis of spatial features (such as CSP). The multichannel EEG signals can be decomposed into multiple time and frequency segments by temporal windows and band-pass filters, respectively. Then, the spatial features of each time-frequency segment are extracted by CSP. Finally, temporal-spatial-frequency features are selected and classified by generalized sparse linear discriminant analysis method [17]. The feature extraction and feature classification are separated in the conventional processing method, and it may not necessarily achieve the optimal results. However, these two steps are closely integrated and jointly optimized in deep learning. During the training of the model, the effect of feature extraction can be directly adjusted according to the classification results. The deep learning networks with series and parallel structures are proposed in this paper. Among them, the frequency and spatial features can be extracted by CNN, and the temporal features can be extracted by LSTM. After extracting all three kinds of features, the networks obtain significantly better results than those extracting only one or two kinds of features. Furthermore, different extraction order of three kinds of features leads to different results. For two data sets, the results of series structures are better than those of parallel structures. It is easy to over-fit, when the temporal features of EEG are directly extracted by LSTM. The classification results are also not good. In order to improve the classification accuracy of LSTM, the spatial-frequency features can be extracted by FBCSP at first [44].

The result of compact CNN is superior to that of shallow CNN. For a single channel, EEG signals belong to one-dimensional time signals. For ease of software implementation, the two-dimensional convolution is used in the standard convolution layer. This operation also keeps each channel independent. Different from the structure of shallow CNN, convolution kernels of compact CNN are used to separate the correlation of signal in plane space and different channels. This method not only simplifies the structure of convolution kernels but also maximizes the characteristic information. This is one benefit of the depthwise and separable convolutions. The previous feature maps are not all connected in the compact CNN, which can reduce the training parameters. The reduction of parameters is more conducive to the classification of less data. Compared with other CNN architectures (such as DeepConvNet and ShallowConvNet) [36], the number of trainable parameters of compact CNN has been decreased at least 31 times (For compact CNN and shallow CNN, the numbers of trainable parameters are 1884 and 58724, respectively). In order to compare with the compact CNN, the results of shallow CNN are also calculated. In the process of training, shallow CNN needs to fit more parameters. From the training and verification processes of shallow CNN, the phenomenon of overfitting still appears. Batch Normalization, average pooling layer and dropout technique have been added, but the effect is not obvious. As the amount of data is limited, the training of DeepConvNet is too easy to over-fit. Therefore, the results of DeepConvNet are not shown.

Although deep learning can be better than conventional algorithms in some cases, it is not absolutely perfect. The major disadvantage of deep learning is that it consumes a lot of computer hardware resources, and GPUs are often needed to run deep learning. For the BCIs, the algorithms have practical value with the ability to process data online. Particularly, low latency can greatly improve user experience. Therefore, in addition to improving the classification accuracy of deep learning, reducing the running time of the model is also an important research direction. To improve the training speed, the easiest way is to reduce the number of model parameters. Taking CNN as an example, compact CNN has fewer trainable parameters than shallow CNN, and the training of compact CNN is also faster. The trained models can be directly used to classify the EEG signals, which is much faster than the training process. However, in the follow-up experiments, whether the delay of models can be suitable for the online BCIs is needed to further verify.

With the small data sets of EEG signals, the training of the standard DNN may be easy to overfit. This leads to poor generalization performance of the trained model. To solve this problem, the amount of data should be increased as much as possible. Prolonged training may cause subjects fatigue, and it results in the decrease of signal to noise ratio of the EEG signals. Therefore, it is unrealistic to increase the amount of data by overextending the experimental time. To increase the amount of data in this paper, the data of imaginary period is intercepted by a sliding time window. In addition to this method, new data can be generated by adding noise to the original data [45]. Besides using your own data to augment the data set, other subjects' data can be used to perform inter-subject transfer learning techniques [46]. To improve classification results, the data augmentation techniques will be enriched in our future research.

5. Conclusions

Without additional feature extraction, the classification results can be gotten from end-to-end learning by DNN. During the training of the model, feature extraction and classification can achieve the best matching. For the traditional processing algorithms of EEG signals, the feature extraction and classification algorithms are set separately. Therefore, DNN has more potential to improve the classification accuracy. In this paper, the novel series and parallel structures with CNN and LSTM are proposed to extract frequency,

spatial and temporal features of the EEG signals, respectively. With better generalization performance, the series architecture can get better results than traditional methods and other structures of DNN. In our future work, to test the practicability of DNN, we will use it as a processing method to control the BCIs online. Whether its delay can meet the needs of human-computer interaction will be also analyzed.

CRediT authorship contribution statement

Li Wang: Conceptualization, Methodology, Funding acquisition, Resources, Writing - original draft. **Weijian Huang:** Data curation, Formal analysis, Investigation, Visualization. **Zhao Yang:** Project administration, Software, Supervision. **Chun Zhang:** Validation, Writing - review & editing.

Acknowledgments

This work was supported by the Science and Technology Planning Project of Guangzhou Municipal Government (201904010466, 201605030014), the Scientific Research Project of Municipal Colleges and Universities of Guangzhou (1201630210).

Declaration of Competing Interest

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

References

[1] H. Yuan, B. He, Brain-computer interfaces using sensorimotor rhythms: current state and future perspectives, IEEE Trans. Biomed. Eng. 61 (5) (2014) 1425–1435.
[2] W. He, Y. Zhao, H. Tang, C. Sun, W. Fu, A wireless BCI and BMI system for wearable robots, IEEE Trans. Syst. Man Cybern.: Syst. 46 (7) (2016) 936–946.
[3] E.A. Aydin, O.F. Bay, I. Guler, Implementation of an embedded web server application for wireless control of brain computer interface based home environments, J. Med. Syst. 40 (1) (2016) 1–27.
[4] A.R. Rabie, V.V. Athanasios, Brain computer interface: control signals review, Neurocomputing 223 (2017) 26–44.
[5] R. Abiri, S. Borhani, E.W. Sellers, Y. Jiang, X.P. Zhao, A comprehensive review of EEG-based brain-computer interface paradigms, J. Neural Eng. 16 (1) (2019), 011001.
[6] G. Pfurtscheller, C. Neuper, D. Flotzinger, M. Pregenzerb, EEG-based discrimination between imagination of right and left hand movement, Electroencephalogr. Clin. Neurophysiol. 103 (6) (1997) 642–651.
[7] N. Padfield, J. Zabalza, H.M. Zhao, V. Masero, J.C. Ren, EEG-based brain-computer interfaces using motor-imagery: techniques and challenges, Sensors 19 (6) (2019) 1423.
[8] M. Naeem, C. Brunner, R. Leeb, B. Graimann, G. Pfurtscheller, Seperability of four-class motor imagery data using independent components analysis, J. Neural Eng. 3 (3) (2006) 208–216.
[9] M. Bittencourt-Villalpando, N.M. Maurits, Stimuli and feature extraction algorithms for brain-computer interfaces: a systematic comparison, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (9) (2018) 1669–1679.
[10] G.K. Anumanchipalli, J. Chartier, E.F. Chang, Speech synthesis from neural decoding of spoken sentences, Nature 568 (2019) 493–498.
[11] C.S. DaSalla, H. Kambara, M. Sato, Y. Koike, Single-trial classification of vowel speech imagery using common spatial patterns, Neural Netw. 22 (9) (2009) 1334–1339.
[12] C.H. Nguyen, G.K. Karavas, P. Artemiadis, Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features, J. Neural Eng. 15 (1) (2018), 016002.
[13] L. Wang, X. Zhang, X.F. Zhong, Y. Zhang, Analysis and classification of speech imagery EEG for BCI, Biomed. Signal Process. Control 8 (6) (2013) 901–908.
[14] L. Wang, X. Zhang, X.F. Zhong, Z.W. Fan, Improvement of mental tasks with relevant speech imagery for brain-computer interfaces, Measurement 91 (2016) 201–209.
[15] J.S. Kirar, R.K. Agrawal, Relevant feature selection from a combination of spectral-temporal and spatial features for classification of motor imagery EEG, J. Med. Syst. 42 (5) (2018) 78.
[16] M. Miao, H. Zeng, A. Wang, C. Zhao, F. Liu, Discriminative spatial-frequency-temporal feature extraction and classifcation of motor imagery EEG: an sparse regression and weighted nave bayesian classifer based approach, J. Neurosci. Methods 278 (2017) 13–24.
[17] V. Peterson, D. Wyser, O. Lambercy, R. Spies, R. Gassert, A penalized time-frequency band feature selection and classification procedure for improved motor intention decoding in multichannel EEG, J. Neural Eng. 16 (1) (2019), 016019.
[18] X. Li, H. Fan, H. Wang, L. Wang, Common spatial patterns combined with phase synchronization information for classification of EEG signals, Biomed. Signal Process. Control 52 (2019) 248–256.
[19] X.Y. Li, C.T. Guan, H.H. Zhang, K.K. Ang, C.C. Wang, A unified Fisher's ratio learning method for spatial filter optimization, IEEE Trans. Neural Netw. Learn. Syst. 28 (11) (2017) 2727–2737.
[20] K.K. Ang, Z.Y. Chin, H.H. Zhang, C.T. Guan, Filter bank common spatial pattern (FBCSP) in brain-computer interface, in: IEEE International Joint Conference on Neural Networks, IEEE, Hong Kong, 2008, pp. 2390–2397.
[21] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, F. Yger, A review of classification algorithms for EEG-based brain-computer interfaces: a 10-year update, J. Neural Eng. 15 (3) (2018), 031005.
[22] M. Mahmud, M.S. Kaiser, A. Hussain, S. Vassanelli, Applications of deep learning and reinforcement learning to biological data, IEEE Trans. Neural Netw. Learn. Syst. 29 (6) (2018) 2063–2079.
[23] M.F. Liu, W. Wu, Z.H. Gu, Z.L. Yu, F.F. Qi, Y.Q. Li, Deep learning based on Batch Normalization for P300 signal detection, Neurocomputing 275 (2018) 288–297.
[24] N. Lu, T.F. Li, X.D. Ren, H.Y. Miao, A deep learning scheme for motor imagery classification based on restricted Boltzmann machines, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (6) (2017) 556–576.
[25] Y.Q. Chu, X.G. Zhao, Y.J. Zou, W.L. Xu, J.D. Han, Y.W. Zhao, A decoding scheme for incomplete motor imagery EEG with deep belief network, Front. Neurosci. 12 (2018) 680.
[26] V.J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung, B.J. Lance, EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces, J. Neural Eng. 15 (5) (2018), 056013.
[27] P. Kaushik, A. Gupta, P.P. Roy, D.P. Dogra, EEG-based age and gender prediction using deep BLSTM-LSTM network model, IEEE Sens. J. 19 (7) (2019) 2634–2641.
[28] P. Bashivan, I. Rish, M. Yeasin, N. Codella, Learning representations from EEG with deep recurrent-convolutional neural networks, International Conference on Learning Representations (ICLR) (2016) 1–15.
[29] Z.Q. Xie, O. Schwartz, A. Prasad, Decoding of finger trajectory from ECoG using deep learning, J. Neural Eng. 15 (3) (2018), 036009.
[30] M.A. Schwemmer, N.D. Skomrock, P.B. Sederberg, J.E. Ting, G. Sharma, M.A. Bockbrader, D.A. Friedenberg, Meeting brain-computer interface user performance expectations using a deep neural network decoding framework, Nat. Med. 24 (11) (2018) 1669–1676.
[31] D.L. Zhang, L.N. Yao, K.X. Chen, S. Wang, X.J. Chang, Y.H. Liu, Making sense of spatio-temporal preserving representations for EEG-based human intention recognition, IEEE transactions on cybernetics, IEEE Trans. Cybern. (2019) 1–12.
[32] 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, Review of the BCI competition IV, Front. Neurosci. 6 (2012) 55.
[33] K. Georgiadis, N. Laskaris, S. Nikolopoulos, I. Kompatsiaris, Exploiting the heightened phase synchrony in patients with neuromuscular disease for the establishment of efficient motor imagery BCIs, J. Neuroeng. Rehabil. 15 (2018) 90.
[34] Z.J. Wang, L. Cao, Z. Zhang, X.L. Gong, Y.R. Sun, H.R. Wang, Short time Fourier transformation and deep neural networks for motor imagery brain computer interface recognition, Concurr. Comput. Pract. Exper. 30 (23) (2018) e4413.
[35] X. Zhang, G.H. Xu, X. Mou, A. Ravi, M. Li, Y.W. Wang, N. Jiang, A convolutional neural network for the detection of asynchronous steady state motion visual evoked potential, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (6) (2019) 1303–1311.
[36] R.T. Schirrmeister, J.T. Springenberg, L.D.J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, Deep learning with convolutional neural networks for EEG decoding and visualization, Hum. Brain Mapp. 38 (11) (2017) 5391–5420.
[37] G. Montavon, W. Samek, K.-R. Müller, Methods for interpreting and understanding deep neural networks, Digit. Signal Process. 73 (2018) 1–15.
[38] K.K. Ang, Z.Y. Chin, C.C. Wang, C.T. Guan, H.H. Zhang, Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b, Front. Neurosci. 6 (2012) 39.
[39] X.F. Xie, Z.L. Yu, H.P. Lu, Z.H. Gu, Y.Q. Li, Motor imagery classification based on bilinear sub-manifold learning of symmetric positive-definite matrices, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (6) (2016) 504–516.
[40] H. Zeng, A.G. Song, Optimizing single-trial EEG classification by stationary matrix logistic regression in brain-computer interface, IEEE Trans. Neural Netw. Learn. Syst. 27 (11) (2016) 2301–2313.
[41] H. Raza, H. Cecotti, G. Prasad, A combination of transductive and inductive learning for handling non-stationarities in motor imagery classification, in: IEEE International Joint Conference on Neural Networks, IEEE, Vancouver, 2016, pp. 763–770.
[42] M.E. Sharbaf, A. Fallah, S. Rashidi, Shrinkage estimator based common spatial pattern for multi-class motor imagery classification by hybrid classifier, in: 3rd International Conference on Pattern Analysis and Image Analysis, IEEE, Shahrekord, 2017.

[43] B.E. Olivas-Padilla, M.I. Chacon-Murguia, Classification of multiple motor imagery using deep convolutional neural networks and spatial filters, Appl. Soft Comput. 75 (2019) 461–472.
[44] T.J. Luo, C.L. Zhou, F. Chao, Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network, BMC Bioinformatics 19 (2018) 344.
[45] Y.J. Li, J.J. Huang, H.Y. Zhou, N. Zhong, Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks, Appl. Sci. 7 (10) (2017) 1060.
[46] F. Fahimi, Z. Zhang, W.B. Goh, T.S. Lee, K.K. Ang, C.T. Guan, Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI, J. Neural Eng. 16 (2) (2019), 026007.