Biomedical Signal Processing and Control 63 (2021) 102144

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Hybrid deep neural network using transfer learning for EEG motor imagery decoding

Ruilong Zhang, Qun Zong, Liqian Dou *, Xinyi Zhao, Yifan Tang, Zhiyu Li

School of Electrical and Information Engineering, Tianjin University, Tianjin, China

ARTICLE INFO

Keywords:
Motor imagery
Brain–computer interfaces
Deep neural network
End-to-end learning
Transfer learning

ABSTRACT

A major challenge in motor imagery (MI) of electroencephalogram (EEG) based brain–computer interfaces (BCIs) is the individual differences for different people. That the classification model should be retrained from scratch for a new subject often leads to unnecessary time consumption. In this paper, a "brain-ID" framework based on the hybrid deep neural network with transfer learning (HDNN-TL) is proposed to deal with individual differences of 4-class MI task. An end-to-end HDNN is developed to learn the common features of MI signal. HDNN consists of convolutional neural network (CNN) and Long Short-Term Memory (LSTM) which are utilized to decode the spatial and temporal features of the MI signal simultaneously. To deal with the EEG individual differences problem, transfer learning technique is implemented to fine-tune the followed fully connected (FC) layer to accommodate new subject with fewer training data. The classification performance on BCI competition IV dataset 2a by the proposed HDNN-TL in terms of kappa value is 0.8. We compared HDNN-TL, HDNN and other state-of-art methods and the experimental results demonstrate that the proposed method can get a satisfying result for new subjects with less time and fewer training data in MI task.

1. Introduction

Brain–computer interface systems (BCIs) create a new human–computer communication for people, which can translate the brain ideas into actual commands to control the external devices [1–3]. Motor imagery (MI) is one of the most classical BCI paradigms. Through MI or some movement intentions, brain activities can be translated into control signals [4]. When people imagine any part of their body, there will be a desynchronization of neural activities in the primary motor cortex of the brain. This phenomenon is called event-related (de)synchronization (ERD/ERS) [5,6]. Specifically, the mu (8–14 Hz) and beta waves (14–30 Hz) are the main spectra of ERS and ERD which are affected by MI task in the EEG [7]. The goal of MI task is to classify human brain imagery activities according to the corresponding changing.

In the current MI-based BCI research, there are three major challenges:

a. Low signal to noise ratio (SNR). Many BCIs are based on the EEG recorded via electrodes placed on the scalp noninvasively, to reduce the trauma to human. This is one of the main reasons for low SNR.

b. Inherent non-stationarity in the recorded signals. Slightly changing of external environment or internal body states, such as noise, attention, fatigue can result in unpredictable effects on EEG signal. [8].

c. Individual differences of different subjects. EEG signals of an individual are just as unique as fingerprints [9]. The uniqueness of EEG signals is particularly strong when a person is imagining an imaginary motion.

Many state-of-art methods are proposed for the first two challenges. For example, common spatial pattern (CSP) is a classical signal processing method to address the first issue, which can extract the useful feature signal from the original EEG signal [10,11]. In recent years, dozens of methods like extension CSP methods with traditional classifier [12–16] or neural network classifiers [17–22] are adopted to promote the capacity of extracting useful information from recorded EEG signal in order to improved classification accuracy of MI task. At present, deep learning method is an excellent classifier for dealing with complex and mass data. Compared with the traditional classification method, deep learning method can describe the nonlinear features without human assistance. This makes the deep learning method a significant choice for processing MI signal based on BCI. Zhang et al. [17] propose a shared classification model based on deep CNN and LSTM to extract spatial and temporal features of MI task. A deep belief network combined with fast Fourier transformation was applied for two-class MI classification in [18]. A new CNN architecture is proposed to introduce the temporal representation of MI data in [19]. And many

* Corresponding author.
E-mail address: douliqian@tju.edu.cn (L. Dou).

https://doi.org/10.1016/j.bspc.2020.102144
Received 24 June 2020; Received in revised form 28 July 2020; Accepted 7 August 2020
Available online 25 August 2020

other methods based on deep learning framework are also used in MI classification and got an excellent classification accuracy.

Meanwhile, many algorithms have been proposed to focus on the nonstationarity effects in EEG signal, which can be divided into two main groups, adaptive parameter model to adapt the non-stationarity [23–25] and robust model to against the non-stationarity [26,27]. However, there are few researches studying focus on how to deal with individual differences of different subjects. This kind of issue is often solved by using specific classification model for corresponding subject.

Machine learning method working well needs abundant data to learn the data distribution [28]. In BCI application, it is time-wasting and troublesome to re-collect the needed training data and rebuild the models. Therefore, the goal of our research is to deal with the individual differences problem with less training sample to improve classification accuracy. Transfer learning is a method in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem [29]. Currently, transfer learning is widely used in image processing [30], natural language processing [31], including BCI research [32,33]. However, few research utilize the transfer learning to address the challenge of individual differences in MI classification task.

Motivated by the above observations, we want to build a "brain-ID" framework which can learn the common features of MI task and fine-tune the partial inner parameters for individual subject. To address the first problem, one-versus-rest filter bank common spatial pattern (OVR-FBCSP) is adopted to preprocess and pre-extract the features of four-class MI signals. Then an end-to-end hybrid deep neural network (HDNN) which consists by CNN, LSTM and FC network to learn the common spatial and temporal features of MI task simultaneously. Furthermore, to solve the issue of individual differences in EEG, transfer learning technique is implemented in the FC network to fine-tune the inner parameters for new subjects. Finally, comparative experiment has been carried out to verify the effectiveness of proposed method. Furthermore, proposed framework has also been evaluated with small training data samples. The results suggest that the end-to-end HDNN-TL framework is a preeminent way to deal with individual differences problem with less training sample in MI task.

The remainder of this paper is organized as follows. Section 2 describes the HDNN-TL structure and its training processing. The applied datasets and the performed experiments are explained in Section 3. Finally, Section 4 concludes this paper.

2. Methodology

The MI feature extraction process is based on OVR-FBCSP and HDNN, and feature–subject correlation process is based on FC with transfer learning technique. The architecture of the HDNN-TL is shown in Fig. 1. In the Section 2-A the OVR-FBCSP algorithm will be reviewed. In rest part of Section 2, the HDNN-TL method will be discussed in details.

2.1. Preprocess and OVR-FBCSP

FBCSP is an extension of CSP algorithm and won the championship of 2008 BCI Competition IV-2a [11]. OVR-FBCSP is a form of many FBCSP algorithms which can deal with multi-class MI task [34]. The procedure for OVR-FBCSP lies as follows.

The recorded EEG signals from BCI device are filtered by using a filter bank with nine subbandpass filters, which are type II Chebyshev filters, starting at 4 Hz and with 4 Hz subbandwidth (4–8 Hz, 8–12 Hz, …). The 4-class OVR-FBCSP, by combining four one-versus-rest (OVR) CSP filters, is used to compute each output of the filter bank. The spatially transformed signal $\boldsymbol{Z}$ can be obtained via,

$$\boldsymbol{Z} = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}, \tag{1}$$

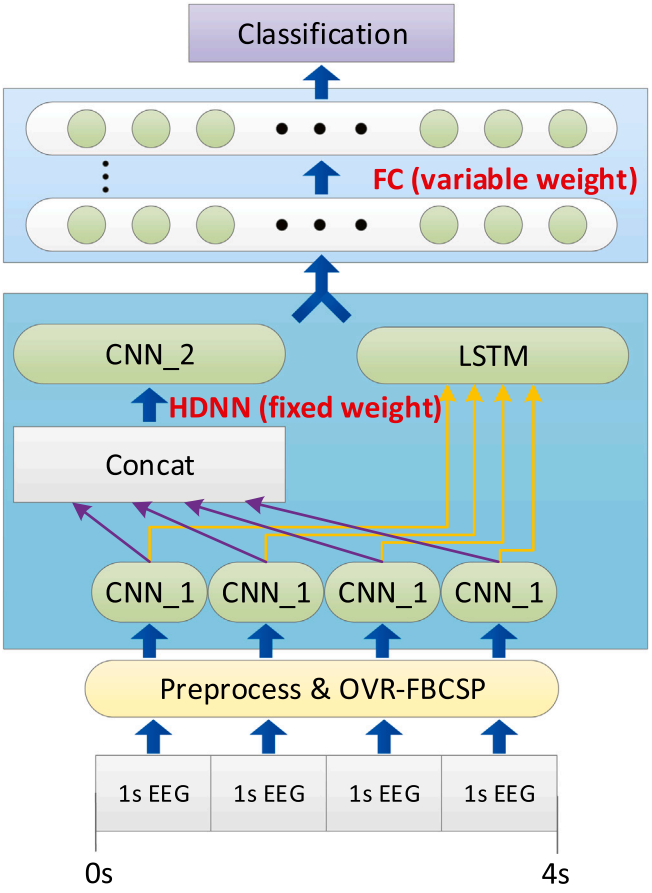

Fig. 1. The architecture of the end-to-end HDNN-TL. It is consisted by CNN, LSTM and FC neural networks. For the new subject, the parameters of HDNN are fixed and only the parameters of FC are fine-tune trained to classify the MI tasks.

where $\boldsymbol{X}$ denotes the original signal, the projection matrix $\boldsymbol{W}$ denotes the OVR-FBCSP spatial filter which consist of 4 OVR CSP filters and can be obtained by,

$$\boldsymbol{C}\boldsymbol{W} = (\sum_{1}^{4}\boldsymbol{C})\boldsymbol{W}\boldsymbol{E}, \tag{2}$$

where $\boldsymbol{C}$ denotes the covariance matrix, $\boldsymbol{E}$ denotes the diagonal matrix which contains the $\boldsymbol{C}$ eigenvalues. For each time step, 4-class MI CSP features can be given by

$$\boldsymbol{f} = \log\frac{\mathrm{diag}(\hat{\boldsymbol{W}}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\hat{\boldsymbol{W}})}{\mathrm{tr}(\hat{\boldsymbol{W}}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\hat{\boldsymbol{W}})}, \tag{3}$$

where diag() and tr() denote the diagonal elements and the trace of the matrix respectively, $\hat{\boldsymbol{W}}$ denotes the combining spatial filters which selects the first and last columns of each $\boldsymbol{W}$.

2.2. CNN

The proposed HDNN includes two sub-CNNs, CNN_1 and CNN_2. CNN_1 is a monolayer including one convolutional layer and a FC layer, which is presented in Fig. 2. It is worth noting that the parameters of 4 CNN_1 are the same. CNN_2 includes three hidden layers, as presented in Fig. 3. At each convolution layer, the 3 × 3 size convolution kernel and the activation function rectified linear unit (RELU) are used to extract spatial features of MI signal,

$$o_c = \max\left(0, \mathrm{conv}\left(W_c, x_c\right) + b_c,\right) \tag{4}$$

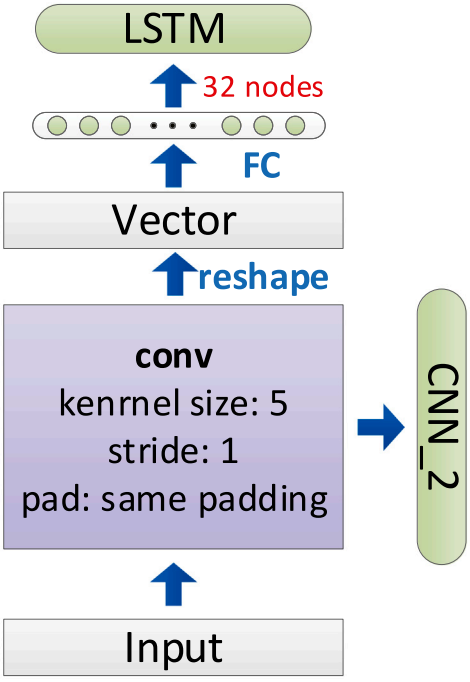

Fig. 2. Proposed CNN_1 model. It provides different forms of output for followed CNN_2 and LSTM.

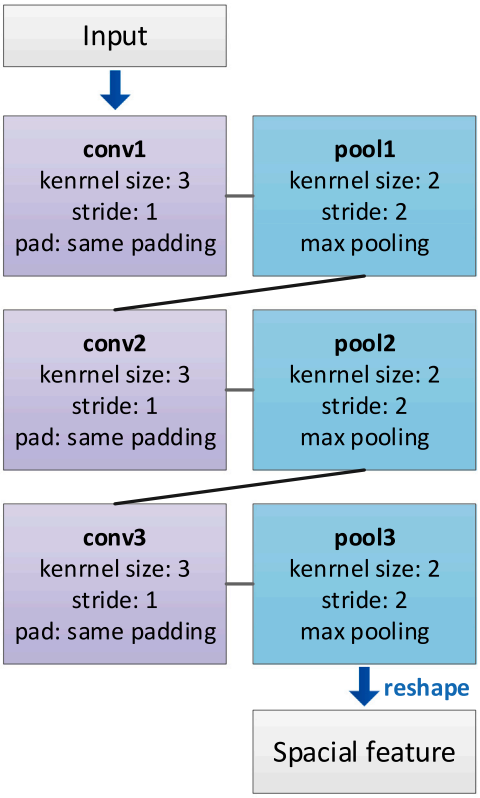

Fig. 3. Proposed CNN_2 model. It contains 3 convolutional and max pooling layers respectively. The details of each layer are illustrated in the figure.

where conv denotes the convolutional operator, x_c, W_c and b_c are the input signal, weight and bias of convolution layer respectively.

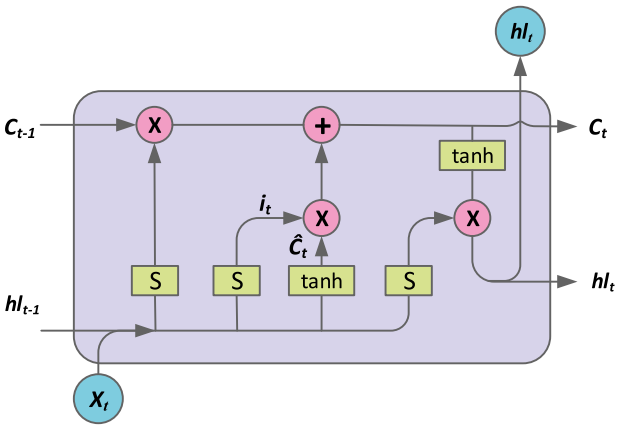

Fig. 4. The framework of LSTM cell, "S" denotes sigmoid activation function, "tanh" denotes hyperbolic tangent activation function, "+ " is plus and "×" is multiplication. The "C_t" presents the state of LSTM cell at current moment [36].

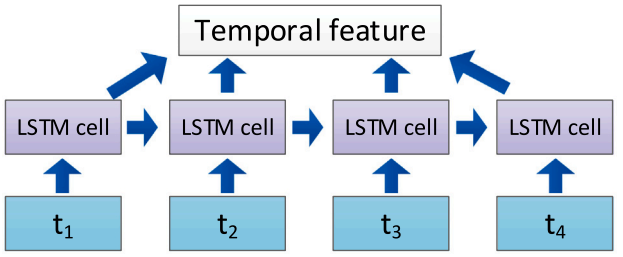

Fig. 5. Proposed Long Short-Term Memory (LSTM) network, which consists of 4 LSTM cell. The output of CNN_1 as the input signal feeds into LSTM cell sequentially at every time step. The parameters of 4 LSTM cell are the same and the framework of LSTM cell is shown in Fig. 4.

A max-pooling layer with kernel size of 2 × 2 is applied to reduce the size of the feature matrix followed each convolutional layer. Furthermore, zero-padding technique is exploited into convolution layer to ensure the output sizes is consistent with input size and avoid losing edge information of spatial feature map.

2.3. LSTM

The extraction of temporal features of EEG signals is as important as that of spatial features. LSTM, an extension neural network of recurrent neural network (RNN), is an excellent way to reveal the internal temporal correlation of time series signals [35]. Therefore, we adopt a LSTM neural network paralleled with CNN to extract temporal features. LSTM cell is a basic unit of LSTM to process the input signal of every time step, as shown in Fig. 4. The output of every time step CNN_1 as the input signal feeds into LSTM cell sequentially to get the temporal feature, as presented in Fig. 5.

Compared with the original RNN, three gates, which are forget gate, external input gate and output gate, are existed in LSTM to deal with exploding gradient problem of RNN. The forget gate is used for discarding useless information of the prior LSTM cell,

$$f_t = \sigma(\boldsymbol{W}^f \cdot [\boldsymbol{hl}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}^f), \tag{5}$$

where $\boldsymbol{hl}_{t-1}$ denotes the prior LSTM cell output, $\boldsymbol{x}_t$ is the current input of the LSTM cell. $\boldsymbol{W}^f$ and $\boldsymbol{b}^f$ represent the weight and bias, respectively. The degree of forgotten information is calculated by a sigmoid function σ to limit the result f_t of operator between 0 and 1.

The input gate is used for replacing the forgotten information with useful information of current input. Meanwhile, the state of LSTM cell

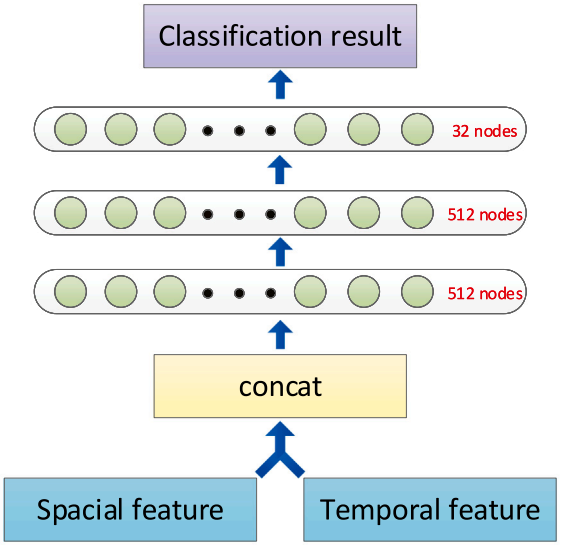

Fig. 6. The structure of full connect (FC) neural network for transfer learning. It contains 3 hidden layers (512 nodes, 512 nodes, 32 nodes in each hidden layer) and dropout technique is used in second layer. When we train a classification model for new subject, we only need to fine-tune the parameters of FC.

is updated by residual and new useful information,

$$i_t = \sigma(\boldsymbol{W}^i \cdot [\boldsymbol{hl}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}^i), \tag{6}$$

$$\hat{\boldsymbol{C}}_t = \tanh(\boldsymbol{W}^c \cdot [\boldsymbol{hl}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}^c), \tag{7}$$

$$\boldsymbol{C}_t = f_t \times \boldsymbol{C}_{t-1} + i_t \times \hat{\boldsymbol{C}}_t. \tag{8}$$

where $\boldsymbol{W}^i$, $\boldsymbol{W}^c$, $\boldsymbol{b}^i$ and $\boldsymbol{b}^c$ are the weight and the bias of input gate respectively, $\boldsymbol{C}_t$ denotes the current state of LSTM cell which has been updated.

Eventually, the output of current LSTM cell can be obtained by,

$$o_t = \sigma(\boldsymbol{W}^o \cdot [\boldsymbol{hl}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}^o), \tag{9}$$

$$\boldsymbol{hl}_t = o_t \times \tanh(\boldsymbol{C}_t). \tag{10}$$

where $\boldsymbol{W}^o$ and $\boldsymbol{b}^o$ is weight and bias respectively, $\boldsymbol{hl}_t$ denotes the current temporal feature of MI signal.

2.4. FC with transfer learning

After CNN and LSTM extract the spatial and temporal features of MI signal, FC is proposed to analyze the obtained features by synthesis and give us a classification result. The structure of FC is shown in Fig. 6. FC is adopted by 3-layer feedforward neural network (FNN). RELU is utilized as activation function in each hidden layer and the softmax function is selected to represent exponential probability distribution between different classes in output layer,

$$y_{p,m} = \frac{e^{y_m}}{\sum_m^T e^{y_m}}, \tag{11}$$

where m is the index of each class, T denotes the total number of classes. Furthermore, dropout technology is used in second hidden layer to reduce the network overfitting.

For transfer learning technique, we follow the method of [37,38] where the parameters of HDNN are fine-tuned as a constant. The FC parameters are random initialized and freshly trained, in order to accommodate the new subject MI feature in our application. The learning rate of FC is kept with default learning rate. Transfer learning in deep learning representation, as empirically is verified in many previous literature, including many application in medical image, such as [39–42]. More thorough theoretical studies on imaging statistics with transfer learning will be needed for future studies.

2.5. Training process

In this paper, one-hot code is utilized to present the result of MI classification. The cross-entropy function is used as the loss function for measuring the difference between two probability distributions of the prediction value $\boldsymbol{y}_p$ and actual label value $\boldsymbol{y}_l$, $\boldsymbol{y}_l$

$$L(\boldsymbol{y}_p, \boldsymbol{y}_l) = -\sum_m y_{p,m} \log y_{l,m}. \tag{12}$$

To reduce the difference between the two probability distributions, adaptive moment estimation (ADAM) [43] approach is used as the optimization method for neural network training.

2.5.1. Pre-training

When learned for common spatial and temporal features of MI, all parameters of the HDNN model are initialized with random Truncated Normal distribution (mean = 0, std = 0.1) from scratch, and trained for 1000 epochs with the mini-batch size of 24 samples. Some other hyperparameters are learning rate: 0.001, LSTM time step: 4 and LSTM cell size: 32. We use the Tensorflow framework and NVidia GTX 1070 GPU to train the neural network.

2.5.2. Fine-tune training

In fine-tune training step, the parameters of HDNN as a constant are absent from training process. The FC parameters are random initialized and freshly trained, and the learning rate is also default learning rate: 0.001.

3. Evaluation and discussion

In this section, we have evaluated our proposed method and compared the performances of HDNN-TL, HDNN and other state-of-art methods on BCI competition IV dataset 2a.

3.1. Dataset

In this paper, 2008 BCI Competition IV dataset 2a public EEG dataset [44] is adopted for evaluating the proposed HDNN and HDNN-TL. The dataset is a 4-class MI task (left hand, right hand, feet, and tongue) recorded by 9 healthy subjects from 22 scalp electrodes with a 250-Hz sampling rate. Each session has 72 trials per class corresponding 288 total samples. The timing scheme consists of 2 s fixation, 1.25 s cue time and followed 4 s MI process. Similar to previous research of MI classification, the performance of proposed method is measured in terms of accuracy and Cohen kappa.

3.2. HDNN evaluation

In BCI Competition IV dataset 2a, each subject is recorded with two sessions, among which the first one is datasets "T" for training the classification algorithm and the other one is datasets "E" for evaluating the trained classification algorithm. In the training process, the HDNN was trained in 500 iterations by datasets "T" (an iteration corresponding a mini-batch samples for training), and datasets "E" was used for evaluating the training effect after every training iteration. The relationship between training error rate and evaluating accuracy were overlaid in Fig. 7. The mean total training time of nine subjects was about 15 s. As shown in Fig. 7, the training error was less than 0.1 and evaluating accuracy curve converged obviously to around 0.8 after 300 iterations. To compare HDNN with other state-of-art method, we provided kappa values for HDNN, as presented in Table 1.

Table 1
Mean kappa values of the HDNN, HDNN-TL and competing methods on the BCI Competition IV Dataset 2a.

Methods	Subjects									Average
	A01	A02	A03	A04	A05	A06	A07	A08	A09	
Ang et al. (FBCSP). [11]	0.68	0.42	0.75	0.48	0.40	0.27	0.77	0.75	0.61	0.57
Kam et al. [45]	0.74	0.35	0.76	0.53	0.38	0.31	0.84	0.74	0.74	0.60
LDA	0.76	0.41	0.83	0.56	0.35	0.26	0.79	0.80	0.72	0.60
Blumberg et al. [46](EM-LDA)	0.59	0.41	0.82	0.57	0.38	0.29	0.79	0.80	0.72	0.60
Vidaurre et al. [13](PMean)	0.76	0.38	0.87	0.60	0.46	0.34	0.77	0.76	0.74	0.63
Luis et al. [47]	0.83	0.51	0.88	**0.68**	0.56	0.35	0.90	0.84	0.75	0.70
Rebeca et al. [48]	0.84	0.55	**0.90**	0.71	0.66	0.44	0.94	0.85	0.76	0.74
Sakhavi et al. [19]	0.88	**0.65**	**0.90**	0.66	0.62	0.45	0.89	0.83	0.79	0.74
Ai et al. [49]	0.77	0.54	0.84	70	0.63	0.61	0.77	0.84	0.86	0.73
zhang et al. [17](shared network)	0.87	0.59	**0.90**	0.76	**0.82**	0.66	**0.95**	0.86	0.89	0.80
HDNN	0.82	0.61	0.85	0.64	0.78	0.73	0.84	0.85	0.87	0.78
HDNN-TL	**0.92**	0.63	0.86	0.67	0.81	**0.75**	0.86	**0.87**	**0.91**	**0.81**

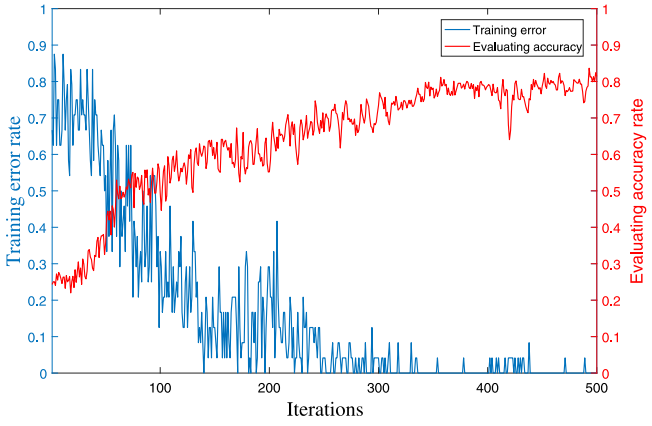

Fig. 7. The relationship between training error, evaluating accuracy and iterations. We iterated training the HDNN in 500 times. For each iteration, training batch is 24 sets of EEG training samples. We also used corresponding evaluation dataset to evaluate the network for every iteration training. Total training time is about 16 s.

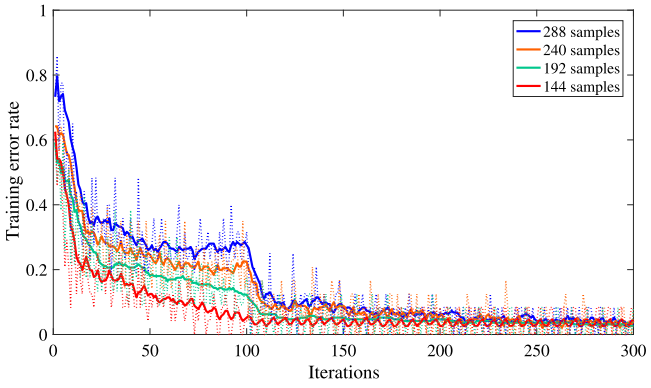

Fig. 8. The relationship of training errors and iterations with different size of training samples. The training error converges faster with a small training sample set.

3.3. HDNN-TL evaluation

Training HDNN-TL needs two step: pre-training and fine-tune training. We merged datasets "T" from all the subjects except the subject, who we want to evaluate, to pre-train the HDNN-TL, and we used the datasets "T" of the absent subject in pre-training process to fine-tune FC parameters. For example, if we want to evaluate the HDNN-TL for subject 8, we merge datasets "T" from subject 1∼7 and 9 to pre-train HDNN-TL, then datasets "T" from subject 8 is used to fine-tune FC layers of HDNN-TL. The comparison of HDNN-TL, HDNN and other art-of-state methods evaluated on BCI Competition IV Dataset 2a was shown in Table 1, including the champion of this Competition [11]. Moreover, LDA results were also provided as the baseline. The highest kappa value was highlighted in boldface.

Although the classification performance is difference for each subject, HDNN and HDNN-TL have a huge improvement compared with previously traditional methods in general. The corresponding kappa values are 0.57 for FBCSP and 0.60 for LDA, whereas they are for HDNN-TL(0.81) and for HDNN(0.78) which demonstrate and promotion 42%, 37%(FBCSP) and 35%, 30%(LDA), respectively, with respect on kappa values. Compared with HDNN, the improvement of HDNN-TL is about 4%. The reason of the improvement maybe that HDNN-TL can extract and learn more common features of MI signal. Compared with some current-state of the art methods [19,48,49], not only HDNN but also HDNN-TL has a improvement about 5 percent in Mean kappa values. Compared with the shared network in [17], HDNN is less impressive. The possible reason behind this is that all the subjects datasets are used to train the shared network and the MI features is learned more comprehensively. However, the defect of shared network in [17] is that the training datasets need to contain a certain number samples of the subjects expected to be classified. For that reason, the network need to be retrained when comes a new subject. As shown in Table 1, there is nothing to choose between proposed HDNN-TL and the shared network [17] in 4-class MI classification. However, the remarkable advantage of proposed HDNN-TL is the network can learn the individual MI features faster and only need fewer training samples for a new subject.

Fig. 8 shows the training error convergence of HDNN-TL in fine-tune training process of subject 9 with different size of training samples. Just as we expected, the training error converges faster within a smaller training samples. Fig. 9 shows the evaluation accuracy for each subject with different numbers of training samples. Although larger dataset often leads to better classification results, a satisfactory classification result can also be gotten by a small one, even better than traditional methods. The advantages of training network by a small sample are faster network convergence and fewer training data needed to be collected from the new subject.

4. Conclusion

The emergence of DL technique has greatly enhanced classification tasks in several fields, such as natural language and image processing. In recent years, deep learning methods have also used widely in BCI applications. Huge amount of multi-channel EEG time series can be fed into deep neural networks to get a satisfactory result. The challenges of improving the classification accuracy due to (1) how to learn the features presentation of MI task sufficiently and (2) how to deal with individual differences in EEG signal of different subjects. To address the above challenges, we proposed "brain-ID" framework based on deep transfer learning methods for the classification of MI tasks.

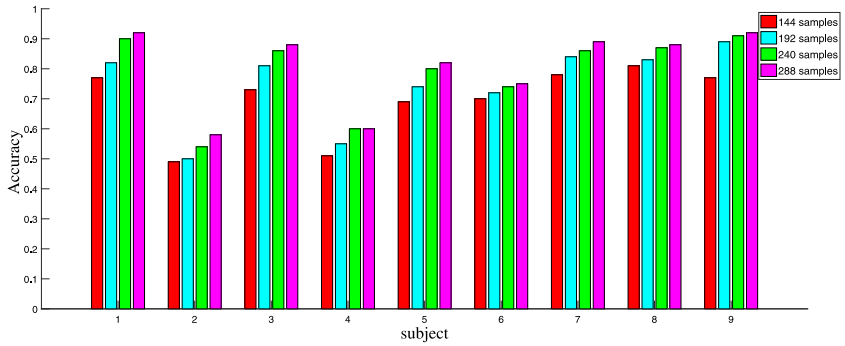

Fig. 9. The comparison of evaluating accuracy different size of training samples. Although a better classification result can be gotten by a larger dataset, the classification result is also acceptable by a small dataset.

The proposed end-to-end HDNN avoids the loss of features by learning the spatial and temporal features simultaneously from MI signal (addressing problem 1). The combination of CNN and LSTM builds a parallel network that has a significantly higher accuracy than traditional feature extraction and classification technique. Furthermore, OVR-FBCSP method is adopted to make the MI features more prominent. Exactly as the conclusion in [19], the CSP method is not affected by the network optimization and in turn, the network is forced to work with an input it has no control over.

Unlike other methods, the end-to-end HDNN-TL framework can fine-tune the parameters to adjust to a new subject quickly by using transfer learning technique (addressing problem 2). In the traditional classification technique, it is a waste of time and troublesome to train a classification model repeatedly when it comes to a new subject, due to *a*. It needs sufficient MI data to make the classification model accurately. It is obviously that collecting the data needs a lot of time and labor resources; *b*. It also takes a lot of time to train the classification model from scratch. To address the above problems, we proposed HDNN to learn the common feature of MI task and FC to map the common features with a new subject. Therefore, a small sample size can achieve a acceptable classification accuracy and the training time decreases with fewer trainable parameters.

Furthermore, to settle over-fitting problem, the regularization technique, such as dropout, is utilized. Another advantage of this work is the zero-padding strategy which is used in CNN framework to avoid losing the edge information of EEG spatial features.

One limitation of this study is that transfer learning method we used still requires a small number of samples with labels to training the networks. According to the latest transfer learning researches, they can exploit rich labeled data from relevant domains to help the learning in the target task with unsupervised domain adaptation [50,51]. This might be a possible way for the improved performance. In the future, we will attempt to explore an unsupervised domain adaptation transfer learning method for MI classification.

Overall, this study indicates that "brain-ID" framework by means of CNN, LSTM and transfer learning is a promising classification technique for MI task which outperforms other technique such as LDA, SVM, and other traditional classifier. The presented work can be applied to imagery-based BCI systems and extended to other types of EEG-based BCIs.

CRediT authorship contribution statement

Ruilong Zhang: Writing - original draft, Conceptualization, Methodology, Software. **Qun Zong:** Project administration, Supervision. **Liqian Dou:** Writing - review & editing, Supervision. **Xinyi Zhao:** Writing - review & editing, Conceptualization. **Yifan Tang:** Software, Validation. **Zhiyu Li:** Investigation, Software.

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.

Acknowledgments

This research was supported in Fund of Science and Technology on Space Intelligent Control Laboratory (6142208180202); the Ministry of Education Equipment Development Fund (6141A0202304, 6141A02033311) in part by National key research and development Program of China under Grant(2018AAA0102401).

References

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophysiol. 113 (6) (2002) 767–791, http://dx.doi.org/10.1016/S1388-2457(02)00057-3.

[2] d.L.B. Van, B.D. Plass-Oude, B. Reuderink, M. Poel, A. Nijholt, How much control is enough? Influence of unreliable input on user experience, IEEE Trans. Cybern. 43 (6) (2013) 1584–1592, http://dx.doi.org/10.1109/TCYB.2013.2282279.

[3] J. Long, Y. Li, H. Wang, T. Yu, J. Pan, F. Li, A hybrid brain computer interface to control the direction and speed of a simulated or real wheelchair, IEEE Trans. Neural Syst. Rehabil. Eng. 20 (5) (2012) 720–729, http://dx.doi.org/10.1109/TNSRE.2012.2197221.

[4] M. Arvaneh, C. Guan, K.K. Ang, C. Quek, Optimizing spatial filters by minimizing within-class dissimilarities in electroencephalogram-based brain–computer interface, IEEE Trans. Neural Netw. Learn. Syst. 24 (4) (2013) 610–619, http://dx.doi.org/10.1109/TNNLS.2013.2239310.

[5] G. Pfurtscheller, Graphical display and statistical evaluation of event-related desynchronization (ERD), Electroencephalogr. Clin. Neurophysiol. 43 (5) (1977) 757–760.

[6] G. Pfurtscheller, Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest, Electroencephalogr. Clin. Neurophysiol. 83 (1) (1992) 62–69.

[7] G. Pfurtscheller, C. Neuper, D. Flotzinger, M. Pregenzer, EEG-based discrimination between imagination of right and left hand movement, Electroencephalogr. Clin. Neurophysiol. 103 (6) (1997) 642–651.

[8] T.M. Vaughan, W.J. Heetderks, L.J. Trejo, W.Z. Rymer, M. Weinrich, M.M. Moore, A. Kübler, B.H. Dobkin, N. Birbaumer, E. Donchin, Brain-computer interface technology: a review of the second international meeting, IEEE Trans. Neural Syst. Rehabil. Eng. 11 (2) (2003) 94–109.

[9] J. Klonovs, C.K. Petersen, H. Olesen, A. Hammershoj, ID proof on the go: Development of a mobile EEG-based biometric authentication system, IEEE Veh. Technol. Mag. 8 (1) (2013) 81–89, http://dx.doi.org/10.1109/MVT.2012.2234056.

[10] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K. Muller, Optimizing spatial filters for robust EEG single-trial analysis, IEEE Signal Proc. Mag. 25 (1) (2008) 41–56, http://dx.doi.org/10.1109/MSP.2008.4408441.

[11] K.K. Ang, Z.Y. Chin, C. Wang, C. Guan, H. Zhang, Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b, Front. Neurosci. 6 (2012) 39, http://dx.doi.org/10.3389/fnins.2012.00039.

[12] S. Kumar, A. Sharma, T. Tsunoda, An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information, BMC Bioinformatics 18 (Suppl 16) (2017) 545.

[13] C. Vidaurre, M. Kawanabe, P. von Bünau, B. Blankertz, K.R. Müller, Toward unsupervised adaptation of LDA for brain–computer interfaces, IEEE Trans. Biomed. Eng. 58 (3) (2011) 587–597, http://dx.doi.org/10.1109/TBME.2010.2093133.

[14] Y. Li, C. Guan, An extended EM algorithm for joint feature extraction and classification in brain-computer interfaces, Neural Comput. 18 (11) (2006) 2730–2761.

[15] S. Sun, C. Zhang, Adaptive feature extraction for EEG signal classification, Med. Biol. Eng. Comput. 44 (10) (2006) 931.

[16] W. Wu, X. Gao, B. Hong, S. Gao, Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL), IEEE Trans. Biomed. Eng. 55 (6) (2008) 1733–1743, http://dx.doi.org/10.1109/TBME.2008.919125.

[17] Z. Ruilong, Z. Qun, L. Dou, Z. Xinyi, A novel hybrid deep learning scheme for four-class motor imagery classification, J. Neural Eng. (2019) http://dx.doi.org/10.1088/1741-2552/ab3471, URL http://iopscience.iop.org/10.1088/1741-2552/ab3471.

[18] N. Lu, T. Li, X. Ren, H. Miao, A deep learning scheme for motor imagery classification based on restricted Boltzmann machines, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (6) (2017) 566–576, http://dx.doi.org/10.1109/TNSRE.2016.2601240.

[19] S. Sakhavi, C. Guan, S. Yan, Learning temporal information for brain-computer interface using convolutional neural networks, IEEE Trans. Neural Netw. Learn. Syst. 29 (11) (2018) 5619–5629, http://dx.doi.org/10.1109/TNNLS.2018.2789927.

[20] R.T. Schirrmeister, J.T. Springenberg, F. Ldj, 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.

[21] S. Sakhavi, C. Guan, S. Yan, Parallel convolutional-linear neural network for motor imagery classification, in: 2015 23rd European Signal Processing Conference (EUSIPCO), 2015, pp. 2736–2740, http://dx.doi.org/10.1109/EUSIPCO.2015.7362882.

[22] Y.R. Tabar, U. Halici, A novel deep learning approach for classification of EEG motor imagery signals, J. Neural Eng. 14 (1) (2017) 016003.

[23] K.P. Thomas, C. Guan, C.T. Lau, A.P. Vinod, K.A. Kai, Adaptive tracking of discriminative frequency components in electroencephalograms for a robust brain–computer interface, J. Neural Eng. 8 (3) (2011) 1–15.

[24] S. Lu, C. Guan, H. Zhang, Unsupervised brain computer interface based on intersubject information and online adaptation, IEEE Trans. Neural Syst. Rehabil. Eng. 17 (2) (2009) 135–145, http://dx.doi.org/10.1109/TNSRE.2009.2015197.

[25] C. Vidaurre, C. Sannelli, K.R. Müller, B. Blankertz, Machine-learning-based coadaptive calibration for brain-computer interfaces, Neural Comput. 23 (3) (2011) 791–816.

[26] W. Samek, C. Vidaurre, K.R. Müller, M. Kawanabe, Stationary common spatial patterns for brain–computer interfacing, J. Neural Eng. 9 (2) (2012) 026013.

[27] F. Lotte, C. Guan, Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms, IEEE Trans. Biomed. Eng. 58 (2) (2011) 355–362, http://dx.doi.org/10.1109/TBME.2010.2082539.

[28] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359, http://dx.doi.org/10.1109/TKDE.2009.191.

[29] Q. Yang, W.U. Xindong, 10 challenging problems in data mining research, Int. J. Inf. Tech. Decis. 5 (04) (2006) 597–604.

[30] Y.A.L. Alsabahi, L. Fan, X. Feng, Image classification method in DR image based on transfer learning, in: 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA), 2018, pp. 1–4, http://dx.doi.org/10.1109/IPTA.2018.8608157.

[31] Y. Zhang, X. Gao, L. He, W. Lu, R. He, Objective video quality assessment combining transfer learning with CNN, IEEE Trans. Neural Netw. Learn. Syst. (2019) 1–15, http://dx.doi.org/10.1109/TNNLS.2018.2890310.

[32] M. Völker, R.T. Schirrmeister, L.D.J. Fiederer, W. Burgard, T. Ball, Deep transfer learning for error decoding from non-invasive EEG, in: 2018 6th International Conference on Brain-Computer Interface (BCI), 2018, pp. 1–6, http://dx.doi.org/10.1109/IWW-BCI.2018.8311491.

[33] D. Wu, B. Lance, V. Lawhern, Transfer learning and active transfer learning for reducing calibration data in single-trial classification of visually-evoked potentials, in: 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2014, pp. 2801–2807, http://dx.doi.org/10.1109/SMC.2014.6974353.

[34] G. Dornhege, B. Blankertz, G. Curio, K.-R. Müller, Increase information transfer rates in BCI by CSP extension to multi-class, in: Advances in Neural Information Processing Systems, 2004, pp. 733–740.

[35] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.

[36] K. Greff, R.K. Srivastava, J. Koutník, B.R. Steunebrink, J. Schmidhuber, LSTM: A search space odyssey, IEEE Trans. Neural Netw. Learn. Syst. 28 (10) (2017) 2222–2232.

[37] A.S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, CNN features off-the-shelf: An astounding baseline for recognition, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014, pp. 512–519, http://dx.doi.org/10.1109/CVPRW.2014.131.

[38] R. Girshick, J. Donahue, T. Darrell, J. Malik, Region-based convolutional networks for accurate object detection and segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 38 (1) (2016) 142–158, http://dx.doi.org/10.1109/TPAMI.2015.2437384.

[39] H. Shin, L. Lu, L. Kim, A. Seff, J. Yao, R.M. Summers, Interleaved text/image deep mining on a large-scale radiology database, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1090–1099, http://dx.doi.org/10.1109/CVPR.2015.7298712.

[40] A. Gupta, M. Ayhan, A. Maida, Natural image bases to represent neuroimaging data, in: International Conference on Machine Learning, 2013, pp. 987–994.

[41] F. Fahimi, Z. Zhang, W. Boon Goh, T.-S. Lee, K. Ang, C. Guan, Inter-subject transfer learning with end-to-end deep convolutional neural network for EEG-based BCI, J. Neural Eng. 16 (2018) http://dx.doi.org/10.1088/1741-2552/aaf3f6.

[42] C. Wei, Y. Lin, Y. Wang, T. Jung, N. Bigdely-Shamlo, C. Lin, Selective transfer learning for EEG-based drowsiness detection, in: 2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015, pp. 3229–3232, http://dx.doi.org/10.1109/SMC.2015.560.

[43] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.

[44] M. Tangermann, K.-R. Müller, A. Aertsen, B. N, C. Braun, B. C, R. Leeb, M. C, M. KJ, G. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, S. A, V. C, S. Waldert, B. B, Review of the BCI competitioncompetition IV, Front. Neurosci. 55 (2012).

[45] T.E. Kam, H.I. Suk, S.W. Lee, Non-homogeneous spatial filter optimization for electroencephalogram (EEG)-based motor imagery classification, Neurocomputing 108 (5) (2013) 58–68.

[46] J. Blumberg, J. Rickert, S. Waldert, A. Schulze-Bonhage, A. Aertsen, C. Mehring, Adaptive classification for brain computer interfaces, in: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, pp. 2536–2539, http://dx.doi.org/10.1109/IEMBS.2007.4352845.

[47] L.F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, R. Hornero, Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain-computer interfaces, Neurocomputing 159 (C) (2015) 186–196, http://dx.doi.org/10.1016/j.neucom.2015.02.005.

[48] L.F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, D. Álvarez, R. Hornero, Adaptive stacked generalization for multiclass motor imagery-based brain computer interfaces, IEEE Trans. Neural Syst. Rehabil. Eng. 23 (4) (2015) 702–712, http://dx.doi.org/10.1109/TNSRE.2015.2398573.

[49] Q. Ai, A. Chen, K. Chen, Q. Liu, T. Zhou, S. Xin, Z. Ji, Feature extraction of four-class motor imagery EEG signals based on functional brain network, J. Neural Eng. 16 (2) (2019) 026032.

[50] E. Tzeng, J. Hoffman, K. Saenko, T. Darrell, Adversarial discriminative domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7167–7176.

[51] Y. Zhang, Y. Wei, Q. Wu, P. Zhao, S. Niu, J. Huang, M. Tan, Collaborative unsupervised domain adaptation for medical image diagnosis, 2020, arXiv preprint arXiv:2007.07222.