IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 21, 2018, accepted January 13, 2019, date of publication February 1, 2019, date of current version February 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2895133

A Novel Deep Learning Approach With Data Augmentation to Classify Motor Imagery Signals

ZHIWEN ZHANG[1], FENG DUAN[1], JORDI SOLÉ-CASALS[2,3], JOSEP DINARÈS-FERRAN[2], ANDRZEJ CICHOCKI[4,5,6,7], (Fellow, IEEE), ZHENGLU YANG[8], AND ZHE SUN[9]

[1]Department of Artificial Intelligence, Nankai University, Tianjin 300350, China
[2]Department of Engineering, University of Vic–Central University of Catalonia, 08500 Barcelona, Spain
[3]Department of Psychiatry, University of Cambridge, Cambridge CB2 3EB, U.K.
[4]Skolkowo Institute of Science and Technology, 121205 Moscow, Russia
[5]College of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China
[6]Department of Informatics, Nicolaus Copernicus University, 87-100 Toru , Poland
[7]Systems Research Institute, Polish of Academy of Science, 01-447 Warsaw, Poland
[8]Department of Computer Science, Nankai University, Tianjin 300350, China
[9]Computational Engineering Applications Unit, Head Office for Information Systems and Cybersecurity, RIKEN, Saitama 351-0198, Japan

Corresponding authors: Feng Duan (duanf@nankai.edu.cn) and Zhe Sun (zhe.sun.vk@riken.jp)

This work was supported in part by the National Natural Science Foundation of China under Grant 61673224, Grant U1636116, and Grant 11431006, in part by the Tianjin Science Fund for Distinguished Young Scholars under Grant 18JCJQJC46100, in part by the Ministry of Education and Science of the Russian Federation under Grant 14.756.31.0001, and in part by the Polish National Science Centre under Grant UMO-2016/20/W/NZ4/00354.

ABSTRACT Brain–computer interface provides a new communication bridge between the human mind and devices, depending largely on the accurate classification and identification of non-invasive EEG signals. Recently, the deep learning approaches have been widely used in many fields to extract features and classify various types of data successfully. However, the deep learning approach requires massive data to train its neural networks, and the amount of data impacts greatly on the quality of the classifiers. This paper proposes a novel approach that combines deep learning and data augmentation for EEG classification. We applied the empirical mode decomposition on the EEG frames and mixed their intrinsic mode functions to create new artificial EEG frames, followed by transforming all EEG data into tensors as inputs of the neural network by complex Morlet wavelets. We proposed two neural networks—convolutional neural network and wavelet neural network—to train the weights and classify two classes of motor imagery signals. The wavelet neural network is a new type of neural network using wavelets to replace the convolutional layers. The experimental results show that the artificial EEG frames substantially improve the training of neural networks, and both two networks yield relatively higher classification accuracies compared to prevailing approaches. Meanwhile, we also verified the performance of our new proposed wavelet neural network model in the classification of steady-state visual evoked potentials.

INDEX TERMS Motor imagery classification, deep learning, convolutional neural network, wavelet neural network, empirical mode decomposition, artificial EEG frames.

I. INTRODUCTION

Brain-computer interface (BCI) is a system designed to translate users' brain intentions into commands or machine codes. It creates a direct connection between an individual's intentions and the assistive devices [1]–[5]. Most BCI studies are based on electroencephalograms (EEG), since non-invasive EEG-based BCI provides brain signals with relative ease. There are three types of neuropsychological signatures commonly used in BCI research, steady-state visual evoked potentials (SSVEPs) [6], event-related potentials (ERPs) [7] and motor imagery [8]. Compared with the SSVEPs and ERPs measured through visual or auditory paradigms [9], motor imagery signals caused by the imagination of limb movement can provide human movement intentions without the need of external stimuli [10]–[12].

Many researchers have investigated different feature extraction and classification methods for motor imagery task recognition. A popular classification method for this BCI paradigm is the common spatial pattern (CSP) method with suitable preprocessing [13]. It is more effective than traditional time-frequency domain feature extraction method by extracting the differences in the spatial features of the

two types of signals. Other well-known feature extraction and dimension reduction methods like independent component analysis (ICA) [14] and principal component analysis (PCA) [15] are used frequently to improve the classification accuracy [16]. Some motor imagery feature learning methods are also used for improving the motor imagery classification, such as the sparse Bayesian extreme learning machine [17], sparse group representation model [18] and jaya based adaptive neuro-fuzzy classifier (NFC) [19]. Moreover, tensors emerged as promising tools for the exploratory analysis of multidimensional data. Phan proposed the tensor decomposition method for the motor imagery feature extraction and classification, which dramatically improved the classification accuracy [20]. In the classification session, many traditional algorithms such as support vector machine (SVM) [21], linear discriminant analysis (LDA) [22] and Bayesian classifier [23], have been employed in different studies [24], [25]. In recent years, deep learning has achieved great progress in neuroscience, text mining, and pattern recognition. Some researchers proposed deep learning approaches for motor imagery classification [26]–[28]. However, BCIs based on a motor imagery paradigm typically require a training period to adapt the system to each user's brain, and the long training period may cause the subjects' fatigue and EEG signal drift. Furthermore, the classifiers are created with the acquired EEG by feature extraction. The property of the classifier relies on amount of training data. More data can improve the classification accuracy and be substantial important for the deep learning method. For the small EEG dataset, it is difficult to use deep learning method to achieve better results. Therefore, we need to use effective tools to analyze brain signals' features and create artificial EEG frames for deep learning classification. Common analysis tools like fast Fourier transform (FFT) and wavelets can not be adequate to extract feature and create new EEG frames in this scenario, because EEG signals are non-linear and non-stationary. Previous works have created artificial EEG frames by using some stationary approaches that use Gaussian noise as a source into an FFT-based system [29], but those approaches obviously miss the temporal features of the natural EEG signals. On the other hand, in some studies the artificial EEG is created by mixing different parts of different temporal EEG signals [30], but it keeps the temporal features of the signal but fails to process the frequency features. The empirical mode decomposition algorithm [31] fits with non-stationary signals that change in the frequency structure within a short period of time [32]–[34]. Therefore, empirical mode decomposition is proposed to generate artificial EEG frames, and it is proved effective in the use of new EEG frames for linear classifier [35].

In our approach, we applied empirical mode decomposition method on the EEG frames to generate new artificial frames to improve classification accuracy for the small training dataset. To verify the feasibility of our method, we tried different increasing ratios of artificial EEG frames to help the network training. Then, we transformed EEG data into tensors as inputs of the neural network by complex Morlet wavelets. Finally, we used a two-layer convolutional neural network (CNN) [36] to classify motor imagery feature tensors. In addition, based on the CNN model we proposed the wavelets neural network (WNN) model in which convolutional layers are replaced.

The rest of the paper is organized as follows: Methods for deep learning model and artificial EEG frames generation are explained in Section II. EEG recording experiments and preprocessing as well as their results are presented and discussed in Section III. The discussion and conclusion are given in the Section IV and Section V, respectively.

II. METHOD

In this work, we presented two deep learning methods—CNN and WNN—for EEG classification. We first introduced the two-layer CNN model for 4-dimensional tensors. Then, based on CNN model, we proposed the WNN model that used wavelets to replace convolutional layers. In order to take advantage of the small dataset that is difficult to train the network, we proposed the empirical mode decomposition method to generate the artificial EEG frames for improving the classification results. Five healthy subjects participated in the experiment to validate our methods, and all of them were required to read and sign an informed consent form approved by the Research Ethics Committee of Nankai University before the EEG experiment.

A. CONVOLUTIONAL NEURAL NETWORK

CNN is a multi-layer neural network with several convolution-pooling layer pairs and fully connected layers at the output, which is always designed to recognize shapes in images. Input image is convolved with several 2-dimensional filters in the convolutional layer and subsampled to a smaller size in the pooling layer. Network weights and filters in the convolutional layer are learned through back-propagation algorithm in order to decrease the classification errors.

As for the EEG data, we first transformed them into integrated time, frequency, and electrode location information tensor as inputs by complex Morlet wavelets with bandwidth parameters $f_b = 1$ Hz. As a result, our training data input is a 4-dimensional tensor of N sub-tensors for two classes (N denotes the number of two classes motor imagery training set): N *training samples* $\times$ 23 *frequncy bins* $\times$ 49 *time frames* $\times$ 14 *channels*. And then we normalized the 4-dimensional input tensors and used these tensors to train the network. The proposed CNN structure is summarized in Fig. 1. In the two-layer CNN model, we first set the convolution-2D layer, which includes 200 filters with the size of 3×1. Then a max-pooling2D layer was used for downsampling, which has the pool size of 4×4. Sequentially, we set a convolution-2D layer and a downsampling layer again, where the parameters of the convolution kernel are 200 filters with the size of 2×2. Finally, after flattening the tensors, two fully connected layers were set for classification.

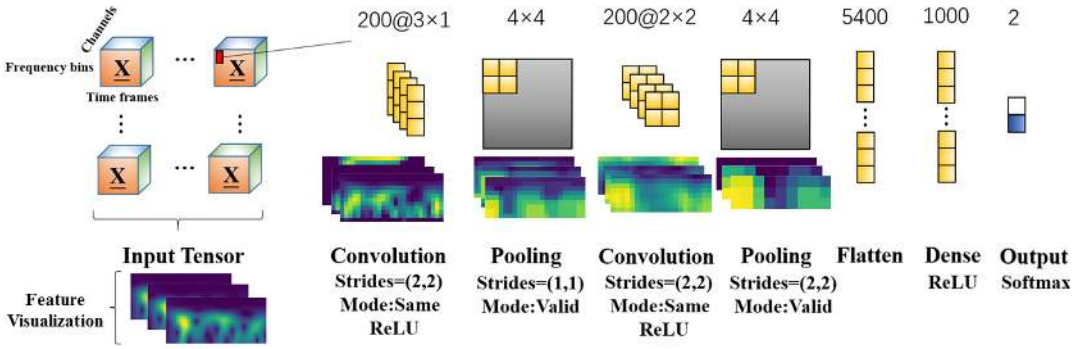


FIGURE 1. The two-layer CNN model for motor imagery classification. The input tensors are N *training samples* × 23 *frequncy bins* × 49 *time frames* × 14 *channels*. And we showed an input tensor's feature representations of first three channels learned at every hidden layer as the feature visualization.

At the convolutional layer, the input tensor is convolved with trainable filters and put through the output function f to form the output feature map. The N_Fth feature map at the convolution-2D layer is obtained as

$$h_{ij}^{N_F} = f(a) = f((W^{N_F} * x)_{ij} + b_{N_F}) \tag{1}$$

where x is the input matrix of the 4-dimensional tensor, W^{N_F} is the weight matrix for the convolution kernel N_F and b_{N_F} is the bias value. Here, the output function of the convolution-2D layer is selected as the rectified linear unit (ReLU) function. The output function is defined as:

$$Relu(a) = ln(1 + e^a) \tag{2}$$

Following the double convolution and pooling, two fully connected layers have two outputs representing the imagery of left and right hand movement. In this way, the labeled training set is imported to the network and can be computed the error between the network output and the desired output. And we used gradient method to minimize the errors by changing the parameters of neural network.

B. WAVELET NEURAL NETWORK

For the feature extraction of CNN model, considering the convolutional layer may not accomplish much for the time-frequency domain information, we used wavelets to replace the convolutional layers. And we changed the input of the network into 5-dimensional tensors for the EEG data, which includes six dictionaries consisted of six sub 4-dimensional tensors made by different bandwidth parameters of complex Morlet wavelets. The proposed WNN model is summarized in Tab I. The size of 5-dimensional input tensor is N *training samples* × 23 *frequncy bins* × 49 *time frames* × 14 *channels* × 6 *dictionaries*. In the WNN model, We first set the dense layer for choosing the suitable wavelets data for subsequent process with the output dimension 2. Specially, the output dimension 2 is the optimal output dimension among 1 to 6, which achieves the highest accuracy. Sequently, we set the max-pooling3D layer with the pool size 4 × 5 × 4 for downsampling. Finally, classification structure has the same structure with the CNN model, including the flatten layer and two fully connected layers for motor imagery classification.

C. TRAINING OF TWO NETWORKS

The implementations were all written in *Keras* [37] with a *Tensorflow* backend [38]. The stochastic gradient descent optimizer was used for the optimization with the following parameters $lr = 0.001$ (learning rate), $momentum = 0.9$ and $decay = 1e^{-6}$. Both two networks were trained by using batch training method for 300 epochs. The batch size varies from 16, 32 and 64, determined by the magnification time of the EEG data augmentation. Consequently, every epoch is guaranteed to have approximative six iterations. In order to solve the overfitting of the models, we applied the regularization for both two networks with the parameters $l2 = 0.004$. And we applied the normalization of the input data for both two networks to improve the accuracy of the model.

D. EMPIRICAL MODE DECOMPOSITION

The empirical mode decomposition method is based on the algorithm that allows users to conduct a data-driven analysis for nonlinear and non-stationary signals. The algorithm decomposes the original signals into a finite number of functions called intrinsic mode functions (IMFs) [31], each of which represents a non-linear oscillation of the signal. This technique fits very well for the non-linear and non-stationary signals, such as EEG signals. Once the signal has been decomposed, we can recover it by adding all the IMFs and the residue without loss.

The main idea, originally presented in [35], is to use empirical mode decomposition to generate new frames by swapping IMFs of the decompositions. This will create new EEG frames similar to the original ones but not equal. Moreover, the intrinsic characteristics of each class will be preserved because we will only mix IMFs of the same class when generating new frames of this class. We thereby generated the appropriate ratio of the artificial EEG frames, with the aim of decreasing overfitting problems in a deep learning system, and eventually improved classification results.

TABLE 1. Detailed architecture for the feature extraction of the WNN.

	Layer	Layer Type	size	stride	Output dimension	Activation	Mode
Features Extraction	Input	1			(N,23,49,14,6)		
	Dense	2			(N,23,49,14,2)	ReLU	
	Maxpooling 3D	3	(4,5,4)	(2,2,2)	(N,10,23,6,2)	ReLU	valid
	Flatten	4			(N,10×23×6×2)		
	Dense	5			(N,200)	ReLU	
Classifier	Dense	6			(N,2)	Softmax	

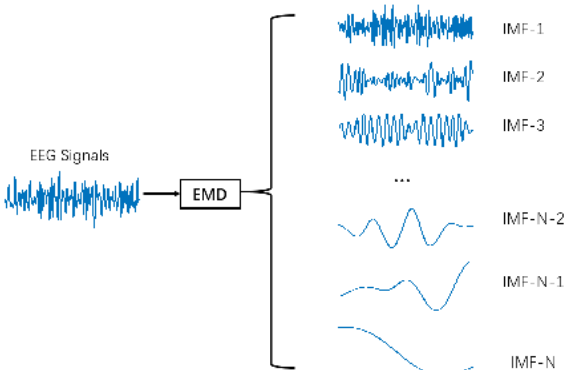


FIGURE 2. Empirical mode decomposition (EMD) of an EEG signal, of which decomposed EEG signals had less than 12 IMFS.

For these intrinsic mode functions by empirical mode decomposition, each signal fulfills two conditions: (1) The number of maxima is the same as the number of zero-crossing, or differs by at most one. (2) The mean value between the envelope of the local maxima and the envelope of the local minima is zero. The process to obtain the IMFs from a signal $x(t)$ is shown as Algorithm. 1.

Once $r_i(t)$ fulfills terminating condition, the signal can be recovered by adding its all IMFs and the final residue $r_n(t)$, where n denotes the number of IMFs by empirical mode decomposition. The structure of the decomposed signal decides the number of IMFs. For our EEG signals, we found that the decomposed signals has at most 12 IMFs. The decomposition process of this method is depicted in Fig. 2.

$$x(t) = \sum_{k=1}^{n} IMF_k(t) + r_n(t) \tag{3}$$

E. THE STRATEGY OF NEW ARTIFICIAL EEG FRAMES

We obtained some IMFs of every decomposed EEG signals by empirical mode decomposition. So we can generate a new EEG frame by combining IMFs from different EEG signals. Most of all, they will exhibit similar characteristics in time and frequency with the signals that contributes with their IMFs, due to the fact that each IMF represents a specific non-linear oscillation. We need create new frames that have same channel number with the EEG signals of our motor imagery paradigm, meaning that any new frames are 14-channel artificial signals.

Here, the strategy of artificial EEG frames has been proved in the linear classification [33]. Initially the real EEG frames are collected in the EEG experiments, then the new EEG collection containing artificial frames is generated as follows:

1. Define the number of artificial frames to be created for deep learning approach, meeting the requirement that each class (left hand class and right hand class) has the same number of artificial EEG frames.

2. Randomly select the frames that contribute with their IMFs to generate the artificial EEG frames. To generate an artificial EEG frame of a specific class, selected EEG frames are split into two sets of EEG frames according to their class.

3. Randomly select EEG frames from the set of frames belonging to the same class. The first selected EEG frame contributes with all its first IMFs (14 IMFs, one IMF per channel), the second one with its second IMFs, and successively until the nth frame, which contributes with its nth IMFs.

4. Generate a new artificial 14-channel EEG frame by adding up all the IMFs corresponding to the same channel.

As explained above, different EEG signals might generate different numbers of IMFs, so we need to predefine the number of artificial EEG frames which contribute with their

Algorithm 1 The Algorithmic Process of Empirical Mode Decomposition

Input: $x(t)$
Output: $imf(t)$
initial $r_0(t) = x(t)$;
repeat(i=1,2...)
  repeat(j=1,2...)
    $h_{j-1}(t) = r_{i-1}(t)$;
    Detect the local maxima and the local minima of $h_{j-1}(t)$;
    Interpolate all local maxima to generate the upper envelope; and interpolate all local minima to generate the lower envelope, respectively;
    Obtain the local mean $m_{j-1}(t)$ by averaging the upper and lower envelopes;
    $h_j(t) = h_{j-1}(t) - m_{j-1}(t)$;
  until $h_j(t)$ satisfies the IMF's conditions
  $imf_i(t) = h_j(t)$;
  $r_i(t) = r_{i-1}(t) - imf_i(t)$;
until If $r_i(t)$ is a monotonic function or does not have enough extrema to calculate the upper and lower envelopes
return imf(t);

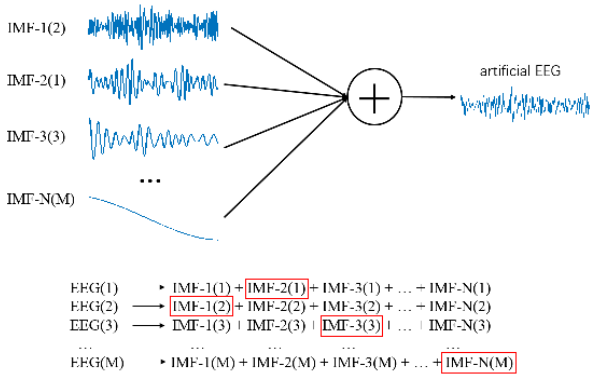


FIGURE 3. Mixing modes strategy for creating artificial EEG frames. In this example we can see that the new artificial frame is created by summing together the IMF-1 from original frame 2, the IMF-2 from original frame 1, the IMF-3 from original frame 3... and the IMF-N from original frame M.

IMFs. We set the number of artificial EEG frames as 15, and additional zero value IMFs need to be added in order to reach 15 IMFs for every decomposed signal. A graphical example of this procedure is depicted in Fig. 3.

III. EXPERIMENT AND RESULT

A. EEG RECORDING AND PREPROCESSING STEPS

To evaluate our methods, we used our experimental motor imagery EEG data and dataset III from BCI Competition II. In the EEG experiment, we selected 14 Ag/AgCl electrodes that were relevant to the motor cortex of brain region following the Brodmann brain function partition and international 10/20 electrode lead system [39]–[41]. Among the 14 electrodes, two (Fz and Cz) were placed in the central brain region, six (T7, P3, P7, CP3, FC3, and C3) were placed in the left brain region, and six (T8, P4, P8, CP4, FC4, and C4) were placed in the right brain region. The electrodes in the left and right brain regions are symmetric (see Fig. 4). Monopolar derivations were used throughout the recordings. All EEG data are acquired by using a g.tec device (g.tec medical engineering GmbH, Schiedlebrg, Austria). The left mastoid and forehead served as the reference and ground, respectively. And the EEG signals were sampled at 256 Hz, then a 50 Hz notch filter was enabled to suppress the power line interference.

In the experiment, a subject sat on a relaxing chair, and placed both arms on his legs in a relaxed position. Our experimental paradigm consisted of two imagery tasks, that are imagery movements with the left hand and right hand [42], [43]. For the experimental flow, at the beginning of a trial ($t = 0s$), a fixation cross "+" was displayed on a black screen and a short acoustic warning tone was played. After two seconds ($t = 2s$), a text prompt for either the left hand or the right hand was displayed in the center of the screen and remained on the screen for two seconds. This prompted the subject to perform the desired motor imagery task. The subject was asked to keep performing the motor imagery task till the fixation cross "+" disappeared from the

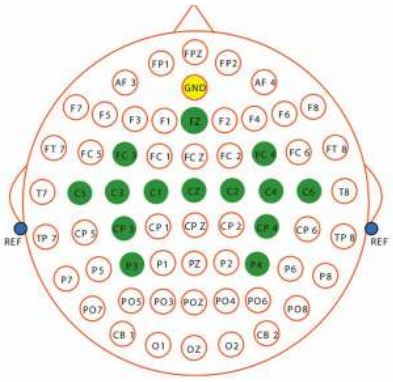


FIGURE 4. The positions of the EEG electrodes for motor imagery experiment.

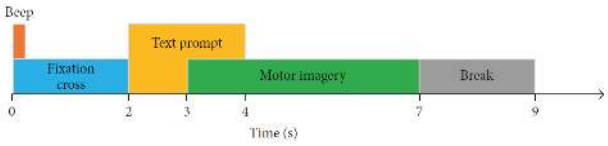


FIGURE 5. The paradigm of the motor imagery signal recording.

screen at $t = 7s$. A short break followed, with a blank screen lasting for two seconds. The paradigm is illustrated in Fig. 5.

Five healthy subjects, that three men (Subject 1, 2, and 5) were 30, 25, and 23 years of age, and two women (Subject 3 and 5) were 21 and 23 years of age, participated in the experiment and performed the two motor imagery tasks. Subject 3 was left-handed, and the other subjects were right-handed. Each session consisted of 60 trials separated by a short break (lasting a couple of minutes), in which each imagery state was performed 30 trials. In total, two sessions of 120 trials were performed per subject.

As for the acquisition of the EEG signals in the BCI Competition II dataset III, at $t = 2s$ an acoustic stimulus indicating the beginning of the trial was used and a cross "+" was displayed for one second. Then, at $t = 3s$, the subject was asked to perform the related motor imagery task by displaying an arrow (left or right). Visual feedback was provided during the motor imagery task. Here we extracted the time interval between 0.5 and 2.5s after the arrow was displayed. The dataset has 280 trials for only one subject (two classes of motor imagery, 140 trials for each class), and EEG signals was sampled at 128 Hz with only three channels—C3, C4 and Cz.

All the EEG signals are preprocessed by a bandpass filter with cutoff frequencies of 8 Hz and 30 Hz. The key point of the enhancement methods for deep learning is that the trials are augmented to become 3-dimensional or 4-dimensional tensors by using the complex Morlet wavelets. Transformation of EEG signals into the time frequency domain is a standard technique to augment dimensionality. In previous research, some researchers used short time Fourier transform (STFT) to get time-frequency domain information for neural network. However, complex Morlet wavelets transformation of the EEG signals has been proved

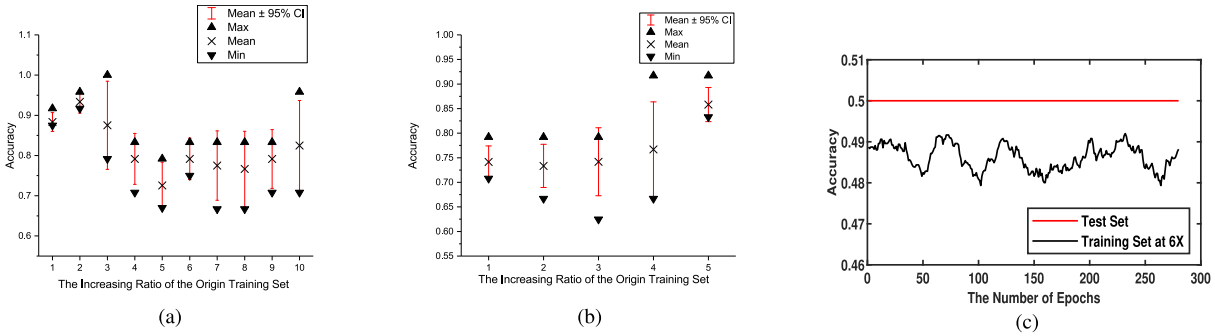


FIGURE 6. The detection of optimal increasing ratio of both two networks for sub1 are depicted in Fig. 6(a) and Fig. 6(b). Every magnification is tested 5 times by generating different artificial EEG frames, and we draw the box chart for the results of the different magnifications. In Fig. 6(c), the WNN arises the exception of no convergence at 6 times magnification (6X). (a) The detection of CNN. (b) The detection of WNN. (c) The exception at 6 times magnification.

effective in recent motor imagery studies, including tensor decomposition and wavelet-based combined feature vectors method [44]. Therefore, we transformed the EEG signals into the time-frequency domain by using the complex Morlet wavelets in two ways. For 3-dimensional tensors used by CNN, the parameters for Morlet wavelets are bandwidth parameter $f_b = 1$ Hz, and the wavelet center frequency $f_c = 1$ Hz. And for the 4-dimensional tensor used by WNN, the bandwidth parameter f_b ranges from 1 Hz to 6 Hz, and the rest of parameters including the wavelet center frequency, frequency bins and time frames are same as the 3-dimensional tensor, meaning that a trial becomes a 4-dimensional tensors with 6 dictionaries.

B. RESULT AND DISCUSSION

In this study, the experiments were conducted in Anoconda environment on an Intel 4.00 GHz Core i7 PC with 16GB of RAM. The code was written in *Keras* with a *Tensorflow* backend.

First of all, we validated our algorithms on our experimental dataset. Besides across subjects, BCI performance is inconsistent within subjects and fluctuates greatly over time. We tried to remove the effect of within subject variation on our results by using cross validation [45]. The performance of proposed networks for the experimental data was evaluated by comparing with tensor decomposition methods using 5-fold cross-validation. In this way, for our small dataset, 80% of 120 trials were selected randomly as the training set and the remaining 20% were selected as the test set. Due to the fact that the training set is rather small for deep learning approach, we need to generate new EEG frames to improve the neural network training. Firstly, we need to confirm the optimal increasing ratio of the origin training set. Every magnification is tested 5 times by generating different artificial EEG frames. For CNN model, as is depicted in Fig. 6(a), the detection results of optimal increasing ratio of the origin training set for Subject 1, we found that the magnifying two times the original training sets had highest mean value and better stability. So we applied the 2 times magnification for all the subjects to analyze the accuracy in CNN model. And as for the WNN, the tested highest magnification was set to 5 because at higher magnification the network sometimes had no convergence, as is depicted as Fig. 6(c). We speculated what the reason of this phenomenon is that the artificial EEG frames influences the performance of WNN model. In the subsequent discussion, we applied the actual dataset without artificial frames to verify our speculation. Here we applied the 5 times magnification as the optimal ratio of the WNN, as is depicted as Fig. 6(b).

For both two networks, in each cross validation we used the optimal increasing ratio to generate EEG frames 5 times, respectively. And we picked the highest accuracy as the results of this cross validation. We verified our method in contrast with tensor decomposition. In this method, the 4-dimensional tensors, which was the same as the inputs of CNN model, were solved by the TUCKER-3 decomposition. Then we selected a good set of features from them by a descending order of Fisher scores. Finally we trained an SVM classifier using the Gaussian Radial Basis Function kernel [46], a K-Nearest Neighbor (KNN) classifier using Euclidean distance and a LDA classifier to compare with the deep learning approach.

As depicted in Table 2, the average accuracy between all subjects for CNN and WNN is 90.0% and 85.2%, respectively. And compared with the tensor decomposition method, both two networks achieved obviously better results. Especially for Subject 4, the poor performance in the tensor decomposition method has been improved by our proposed methods. The CNN method not only has the highest accuracy, but also has small standard deviation. But the accuracies of WNN method have rather high standard deviation, which may be the influence of artificial EEG frames.

Then, we compared our proposed methods with previous research by the public BCI dataset. In the BCI Competition II dataset III, 50% trials were set as the training set, and the remaining 50% were selected as the test set. And every magnification was also tested 5 times by generating different artificial EEG frames. The performance of the artificial EEG frames is depicted as Table 3, and the highest accuracy of our algorithm is obtained as 90.1%, which is better than 89.3%

TABLE 2. Two networks with optimal artificial frames ratio compared with tensor decomposition method (5 fold cross validation).

	Accuracy %(mean±std)				
	KNN	LDA	SVM	CNN	WNN
Sub1	89.9±5.6	78.3±6.2	75±4.2	92.5±3.5	86.7±9.0
Sub2	91.7±2.9	81.7±6.3	80.8±5.6	95.8±2.9	90.8±3.5
Sub3	70.8±6.6	70.0±12.3	62.5±7.8	87.5±3.0	82.5±3.5
Sub4	68.3±7.0	67.5±7.5	66.7±5.9	83.3±2.9	78.3±5.4
Sub5	84.2±3.5	80±4.6	78.3±3.5	90.8±4.6	87.5±8.3
Average	81.0	75.5	72.3	90.0	85.2

TABLE 3. Performance of the artificial EEG frames for BCI competition dataset.

	Accuracy %(mean±std)	
	CNN	WNN
0×Dataset	77.9±0	88.0±0
1×Dataset	88.9±1.9	90.1±1
2×Dataset	85.6±2.2	86.7±2.5
3×Dataset	86.4±2.6	87.3±1.7
4×Dataset	83.6±2.9	85.0±2.6
5×Dataset	82.9±2.7	84.3±1.5

of the winner algorithm of the competition [47] and 88.2% of deep network [48].

IV. DISCUSSION

We proposed two network models for motor imagery classification, and we generated the new artificial EEG frames to help the network training by empirical mode decomposition. However, both two networks were shallow networks. It is necessary to try the deeper network for the motor imagery signal classification. And we only detected the magnification of the origin dataset, which was less than 10. We need to detect larger artificial EEG frames for network training. Moreover, the WNN model arose no convergence at the 6 times magnification of the origin training. Therefore, we need to verify the feasibility of our new proposed WNN model by actual motor imagery dataset to exclude the influence of the artificial EEG frames, and also evaluate the performance for the SSVEP dataset.

A. THE ATTEMPT OF THE DEEP RESIDUAL NETWORK

For the EEG feature tensors made by complex Morlet wavelets, we considered using the deep network to train the weights for motor imagery classification. Residual network was attempted to test deep network's property in this study. Deeper neural networks are more difficult to train, especially for relatively small EEG dataset. As a result, we applied a residual learning framework that can ease the training of networks to guarantee the convergence and rapidity by the short circuit path design. We applied the ResNet-18 model and set the 4-dimensional tensor as the network input, which was the same as the CNN model. We validated the performance of Residual network by comparing with the CNN model, as is depicted as Fig. 7(a) and Fig. 7(b), the ResNet-18 has faster convergence but the accuracy of the test set is lower than that of CNN. And the loss of the test set was abnormally rising as the iteration. Therefore, in this study, we did not apply the deep network for motor imagery classification.

B. THE INFLUENCE OF ARTIFICIAL EEG FRAMES

We applied the empirical mode decomposition algorithm for expanding the origin training set to train the network. For the CNN model, the artificial EEG frames improve the classification accuracy at the optimal ratio. Towards WNN model, although the network has the exception of no convergence at 6 times magnification, it improves the network classification result from 74.2% to 83.3% for Subject 1 at 5 times magnification. Therefore, we considered the artificial EEG frames method is valid for both two networks.

Furthermore, we also detected the larger artificial EEG frames for both two neural networks. For the CNN model, we tried the 20 times magnification of the origin training set to test the classification performance. As is depicted as Fig. 7(c), we found that it achieved better performance than the previous optimal ratio. And it proves that the over 10 times magnification of origin training set has the possibility of further improving the classification effects. But for WNN, at 10 times magnification it also has the poor performance in motor imagery classification, as is depicted as Fig. 7(d). We thought that the WNN model may need mass actual data to train its network.

C. THE NEW PROPOSED WAVELET NEURAL NETWORK

In the EEG experiments, the new proposed WNN model has rather high deviation, and the classification accuracy is slightly lower than the CNN model. In order to verify the network's feasibility and exclude the influence of artificial EEG frames, we used the actual motor imagery big dataset, which involves left/right hand motor imagery movements recorded from 62 channels (with sampling frequency 500 Hz) with duration of two seconds with a four seconds' break between the trials. The dataset was used in tensor decomposition method for motor imagery classification [20] and the data were collected over two sessions with a

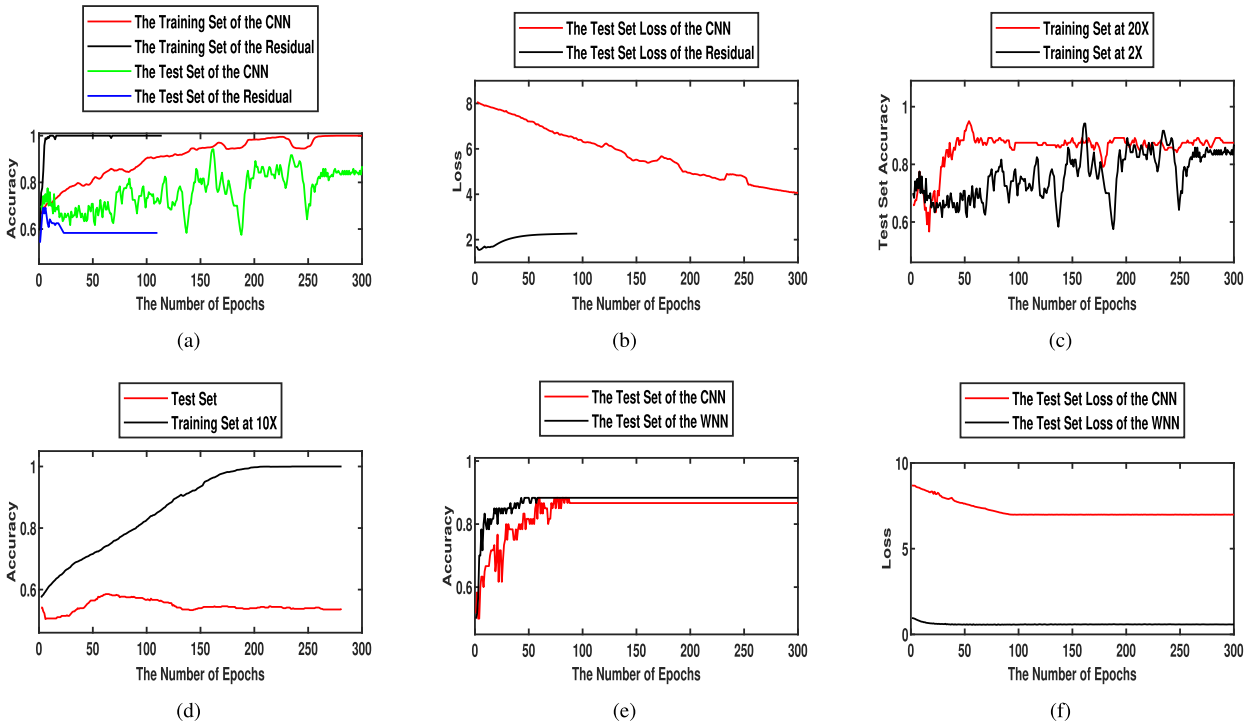


FIGURE 7. The comparisons of CNN and ResNet-18 for Subject 1 in accuracy and loss are depicted as Fig. 7(a) and Fig. 7(b), respectively. In Fig. 7(c), the origin training set at 20 times magnification (20X) for CNN has better performance comparing with the previous optimal ratio. In Fig. 7(d), the origin training set at 10 times magnification (10X) for WNN still has poor performance. And in Fig. 7(e) and Fig. 7(f), the new proposed WNN is validated by comparing with CNN.

15 minutes' break in between. The first session was used as the training set, which included 140 trials (70 trials for each class). And the second session was used as the test set, which included 60 trials (30 trials for each class). Here, the WNN had the same basic structure with the Table 1. But we changed the first dense layer output dimension from 2 to 1, because we found that the multichannel EEG dataset has enough spatial information. Therefore we can decrease the wavelets feature for classification, and it significantly reduced the parameters of the network and computation. As is depicted in Fig. 7(e) and Fig. 7(f), the WNN has better classification performance and smaller loss than the CNN. However, each iteration of the WNN model takes almost five times as long as the CNN. And we speculated that it's because the WNN lacks the consideration of parallel computing.

Also, we used the SSVEP dataset to validate our proposed method. In the SSVEP dataset, the EEG signals were sampled at 256 Hz and only had one channel-Oz for one subject. We extracted one-second EEG signals used for the classification of WNN, which is evoked by 6 Hz and 8 Hz frequency flicking. The training set had 120 trials and the testing set also had 120 trials. The accuracy of the WNN model is 98.3%, which proves that our model is valid for the classification of SSVEP.

V. CONCLUSION

Deep learning network needs to obtain massive data to maintain the classification accuracy, but long period of the EEG acquisition may cause the fatigue of the subjects and data drift of the motor imagery signals. To solve this dilemma, we proposed a novel deep learning approach with data augmentation method to improve the classification accuracy of motor imagery signals and avoid overfitting. We transformed EEG data into tensors as inputs of the neural network by complex Morlet wavelets. And we proposed two models—CNN and WNN—to extract feature and classify motor imagery signals. The empirical mode decomposition approach was utilized to generate artificial EEG frames to train the networks. In the experiments, the filter size and hyper-parameters were investigated, and we also investigated epoch size for both two networks. The best value for batch size was found as 300. To evaluate the performance of the proposed approach, both CNN model and WNN model were compared with the tensor decomposition methods. The experimental results show that our new approaches achieve better results than the tensor decomposition method. And we also validate our proposed methods in the BCI Competition II dataset III. The accuracy of the winner algorithm of the competition was 89.3%, and the accuracy performance of our proposed methods was 90.1%. Especially, for the WNN model, the larger generated artificial EEG frames sometimes cause the network showing the exception of no convergence. Therefore, to verify the feasibility of the new proposed WNN model, we applied the actual motor imagery dataset to exclude the influence of the artificial EEG frames. According to the results, it can be concluded that the WNN model has higher classification accuracy and faster convergence rate than that CNN does.

Also, we proved our WNN model is valid for the classification of SSVEP.

However, relatively low computational efficiency of the WNN is the limitation in our proposed methods. In the future study, we will improve the low computational efficiency by finding better network frames for parallel computation or the approximate representations of the input dictionaries of WNN model.

ACKNOWLEDGMENT

(Zhiwen Zhang, Feng Duan, and Jordi Solé-Casals contributed equally to this work.)

REFERENCES

[1] V. Gandhi, G. Prasad, D. Coyle, L. Behera, and T. M. McGinnity, "EEG-based mobile robot control through an adaptive brain–robot interface," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 9, pp. 1278–1285, Sep. 2014.

[2] J. Li, J. Liang, Q. Zhao, J. Li, K. Hong, and L. Zhang, "Design of assistive wheelchair system directly steered by human thoughts," *Int. J. Neural Syst.*, vol. 23, no. 3, p. 1350013, 2013.

[3] M. Nakanishi, Y. Wang, Y.-T. Wang, Y. Mitsukura, and T.-P. Jung, "A high-speed brain speller using steady-state visual evoked potentials," *Int. J. Neural Syst.*, vol. 24, no. 6, pp. 767–785, 2014.

[4] J. Li *et al.*, "Evaluation and application of a hybrid brain computer interface for real wheelchair parallel control with multi-degree of freedom," *Int. J. Neural Syst.*, vol. 24, no. 4, p. 1450014, 2014.

[5] Y. He *et al.*, "An integrated neuro-robotic interface for stroke rehabilitation using the NASA X1 powered lower limb exoskeleton," in *Proc. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 3985–3988.

[6] K.-K. Shyu *et al.*, "Total design of an FPGA-based brain–computer interface control hospital bed nursing system," *IEEE Trans. Ind. Electron.*, vol. 60, no. 7, pp. 2731–2739, Jul. 2013.

[7] L. da Silva-Sauer, L. Valero-Aguayo, A. de la Torre-Luque, R. Ron-Angevin, and S. Varona-Moya, "Concentration on performance with P300-based BCI systems: A matter of interface features," *Appl. Ergonom.*, vol. 52, pp. 325–332, Jan. 2016.

[8] W. Yi, S. Qiu, H. Qi, L. Zhang, B. Wan, and D. Ming, "EEG feature comparison and classification of simple and compound limb motor imagery," *J. NeuroEng. Rehabil.*, vol. 10, no. 1, p. 106, 2013.

[9] C. J. Bell, P. Shenoy, R. Chalodhorn, and R. P. N. Rao, "Control of a humanoid robot by a noninvasive brain–computer interface in humans," *J. Neural Eng.*, vol. 5, no. 2, pp. 210–214, 2008.

[10] H. Cho, M. Ahn, S. Ahn, M. Kwon, and S. C. Jun, "EEG datasets for motor imagery brain computer interface," *Giga Sci.*, vol. 6, no. 7, pp. 1–8, 2017.

[11] L. F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, and D. Álvarez, and R. Hornero, "Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain–computer interfaces," *Neurocomputing*, vol. 159, no. 1, pp. 186–196, 2015.

[12] G. Pfurtscheller and F. L. Da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, 1999.

[13] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 1994.

[14] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[15] P. J. García-Laencina, G. Rodríguez-Bermudez, and J. Roca-Dorda, "Exploring dimensionality reduction of EEG features in motor imagery task classification," *Expert Syst. Appl.*, vol. 41, no. 11, pp. 5285–5295, 2014.

[16] X. Guo, L. Wang, X. Wu, and D. Zhang, "Dynamic analysis of motor imagery EEG using kurtosis based independent component analysis," in *Advances in Cognitive Neurodynamics ICCN*. Dordrecht, The Netherlands: Springer, 2008, pp. 381–385.

[17] Z. Jin, G. Zhou, D. Gao, and Y. Zhang, "EEG classification using sparse Bayesian extreme learning machine for brain–computer interface," *Neural Comput. Appl.*, pp. 1–9, 2018.

[18] Y. Jiao *et al.*, "Sparse group representation model for motor imagery EEG classification," *IEEE J. Biomed. Health Inform.*, to be published.

[19] Suraj, R. K. Sinha, and S. Ghosh, "Jaya based ANFIS for monitoring of two class motor imagery task," *IEEE Access*, vol. 4, pp. 9273–9282, 2016. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7778152

[20] A. H. Phan and A. Cichocki, "Tensor decompositions for feature extraction and classification of high dimensional datasets," *IEICE Nonlinear Theory Appl.*, vol. 1, no. 1, pp. 37–68, 2010.

[21] A. B. Das, M. I. H. Bhuiyan, and S. M. S. Alam, "Classification of EEG signals using normal inverse Gaussian parameters in the dual-tree complex wavelet transform domain for seizure detection," *Signal, Image Video Process.*, vol. 10, no. 2, pp. 259–266, 2016.

[22] C.-Y. Chiu, C.-Y. Chen, Y.-Y. Lin, S.-A. Chen, and C.-T. Lin, "Using a novel LDA-ensemble framework to classification of motor imagery tasks for brain-computer interface applications," in *Proc. Int. Comput. Symp. (ICS)*, vol. 1, 2014, pp. 136–142.

[23] L. He, D. Hu, M. Wan, Y. Wen, K. M. von Deneen, and M. Zhou, "Common Bayesian network for classification of EEG-based multiclass motor imagery BCI," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 6, pp. 843–854, Jun. 2016.

[24] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, Mar. 2012, Art. no. 39.

[25] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005," *J. Neural Eng.*, vol. 2, no. 4, pp. L14–L22, 2005.

[26] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, p. 016003, 2016.

[27] X. An, D. Kuang, X. Guo, Y. Zhao, and L. He, "A deep learning method for classification of EEG data based on motor imagery," in *Proc. Int. Conf. Intell. Comput.* Springer, 2014, pp. 203–210.

[28] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted Boltzmann machines," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 566–576, Jun. 2017.

[29] A. Paris, G. K. Atia, A. Vosoughi, and S. A. Berman, "A new statistical model of electroencephalogram noise spectra for real-time brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1688–1700, Aug. 2017.

[30] F. Lotte, "Generating artificial EEG signals to reduce BCI calibration time," in *Proc. 5th Int. Brain-Comput. Interface Workshop*, 2011, pp. 176–179.

[31] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[32] F. Riaz, A. Hassan, S. Rehman, I. K. Niazi, and K. Dremstrup, "EMD-based temporal and spectral features for the classification of EEG signals using supervised learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 28–35, Jan. 2016.

[33] J.-R. Huang, S.-Z. Fan, M. F. Abbod, K.-K. Jen, J.-F. Wu, and J.-S. Shieh, "Application of multivariate empirical mode decomposition and sample entropy in EEG signals via artificial neural networks for interpreting depth of anesthesia," *Entropy*, vol. 15, no. 9, pp. 3325–3339, 2013.

[34] S. D. Hawley, L. E. Atlas, and H. J. Chizeck, "Some properties of an empirical mode type signal decomposition algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 3, Mar./Apr. 2008, pp. 3625–3628.

[35] J. Dinarès-Ferran, R. Ortner, C. Guger, and J. Solé-Casals, "A new method to generate artificial frames using the empirical mode decomposition for an EEG-based motor imagery BCI," *Frontiers Neurosci.*, vol. 12, p. 308, May 2018.

[36] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

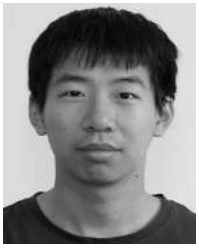
[37] F. Chollet. (2015). *Keras*. [Online]. Available: https://github.com/fchollet/keras

[38] M. Abadi *et al.* (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: https://tensorflow.org

[39] K. Amunts and K. Zilles, "Architectonic mapping of the human brain beyond brodmann," *Neuron*, vol. 88, no. 6, pp. 1086–1107, 2015.

[40] R. M. Sanchez-Panchuelo, J. Besle, A. Beckett, R. Bowtell, D. Schluppeck, and S. Francis, "Within-digit functional parcellation of Brodmann areas of the human primary somatosensory cortex using functional magnetic resonance imaging at 7 tesla," *J. Neurosci.*, vol. 32, no. 45, pp. 15815–15822, 2012.

[41] C. H. Kasess, C. Windischberger, R. Cunnington, R. Lanzenberger, L. Pezawas, and E. Moser, "The suppressive influence of SMA on M1 in motor imagery revealed by fMRI and dynamic causal modeling," *NeuroImage*, vol. 40, no. 2, pp. 828–837, 2008.
[42] A. Solodkin, P. Hlustik, E. E. Chen, and S. L. Small, "Fine modulation in network activation during motor execution and motor imagery," *Cerebral Cortex*, vol. 14, no. 11, pp. 1246–1255, 2004.
[43] C. R. Genovese, N. A. Lazar, and T. Nichols, "Thresholding of statistical maps in functional neuroimaging using the false discovery rate," *NeuroImage*, vol. 15, no. 4, pp. 870–878, 2002.
[44] D. Lee, S.-H. Park, and S.-G. Lee, "Improving the accuracy and training speed of motor imagery brain–computer interfaces using wavelet-based combined feature vectors and Gaussian mixture model-supervectors," *Sensors*, vol. 17, no. 10, p. 2282, 2017.
[45] M. Ahn and S. C. Jun, "Performance variation in motor imagery brain–computer interface: A brief review," *J. Neurosci. Methods*, vol. 243, pp. 103–110, Mar. 2015.
[46] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, "SVM and kernel methods—Matlab toolbox," Perception Syst. Inf., INSA Rouen, Rouen, France, Tech. Rep., 2005. [Online]. Available: http://asi.insa-rouen.fr/enseignants/~arakoto/toolbox/
[47] S. Lemm, C. Schäfer, and G. Curio, "BCI competition 2003-data set III: Probabilistic modeling of sensorimotor μ rhythms for classification of imaginary hand movements," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1077–1080, Jun. 2004.
[48] Y. Ren and Y. Wu, "Convolutional deep belief networks for feature extraction of EEG signal," in *Proc. Int. Joint Conf. Neural Netw.*, 2014, pp. 2850–2853.



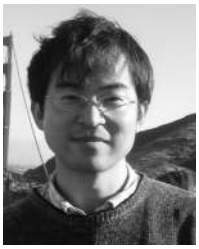
ZHIWEN ZHANG received the B.E. degree in intelligence science and technology from Nankai University, Tianjin, China, in 2016, where he is currently pursuing the master's degree with the College of Artificial Intelligence. His current research interests include EEG signal processing and robot research.

FENG DUAN received the B.E. and M.E. degrees in mechanical engineering from Tianjin University, China, in 2002 and 2004, respectively, and the M.S. and Ph.D. degrees in precision engineering from The University of Tokyo, Japan, in 2006 and 2009, respectively. He is currently a Professor with Nankai University, China. His research interests include cellular manufacture systems, rehabilitation robots, and brain machine interfaces.

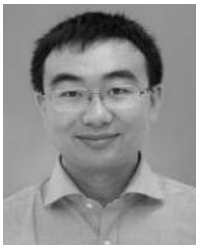
JORDI SOLÉ-CASALS received the M.Sc. degree in telecommunication engineering and the Ph.D. degree with European label from the Polytechnic University of Catalonia, Barcelona, in 1995 and 2000, respectively, and the bachelor's degree in humanities from the Open University of Catalonia, Barcelona, in 2010. In 1994, he joined the Department of Engineering, University of Vic–Central University of Catalonia. He was a Visiting Researcher with the GIPSA-Lab, Grenoble, France, LABSP-RIKEN, Tokyo, Japan, and BMU, Cambridge, U.K., and a Visiting Professor with the Department of Psychiatry, University of Cambridge. He currently continues these relations with these laboratories. He is also the Head of the Data and Signal Processing Research Group and maintains active collaborations with other international groups. His research interests include neurosciences, biomedical signal processing, machine learning, neural networks, source separation, and biometrics.

JOSEP DINARÈS-FERRAN received the M.Sc. degree in telecommunication engineering from the Polytechnic University of Catalonia, Barcelona, in 2010. He is currently pursuing the Ph.D. degree the University of Vic–Central University of Catalonia (UVic-UCC), where he joined the Data and Signal Processing Research Group, UVic-UCC. Since 2016, he has been researching with g.tec medical engineering, Barcelona, Spain. His research interests include brain–computer interface based on EEG for the purpose of neurorehabilitation.

ANDRZEJ CICHOCKI (F'13) received the M.Sc. (Hons.), Ph.D., and Dr.Sc. (Habilitation) degrees from the Warsaw University of Technology, Poland. He worked several years at the University of Erlangen-Nuerenberg, Germany, as an Alexander-von-Humboldt Research Fellow and a Guest Professor. From 1995 to 2018, he was a Team Leader and the Head of the Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Japan. Under the guidance of Prof. Cichocki, the new Laboratory Tensor Networks and Deep Learning for Applications in Biomedical Data Mining is established at SKOLTECH. The mission of the Laboratory is to perform cutting-edge innovative research in the design and analysis of deep neural networks, tensor networks, and multiway component analysis for biomedical applications. He is among the most cited Polish computer scientists and is or has been an associate editor of the top journals in signal processing and neural networks. His publications currently report over 38,000 citations and with an h-index of 85.



ZHENGLU YANG received the Ph.D. degree from The University of Tokyo, Japan, in 2008. From 2008 to 2014, he was a Faculty with the Institute of Industrial Science, The University of Tokyo. He is currently a Professor with the College of Computer Science, Nankai University, China. His research interests include artificial intelligence, database systems, and big data mining.



ZHE SUN received the Ph.D. degree from Yokohama City University, in 2017. He joined RIKEN, in 2015, as a Research Support Assistant. He has been a Research Scientist with RIKEN, since 2017. He is currently a Research Scientist with the Computational Engineering Applications Unit, R&D Group, Head Office for Information Systems and Cybersecurity, National Institute of RIKEN, Japan. From 2014 to 2017, his research topics were the development of spiking neuron model, and spiking neural network to understand end elucidate brain functions. His current research interests include large scale brain simulation and neuromorphic engineering.