This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2019.2953121, IEEE Transactions on Neural Systems and Rehabilitation Engineering

Deep Channel-Correlation Network for Motor Imagery Decoding from Same Limb

Xuelin Ma, Shuang Qiu, Wei Wei, Shengpei Wang, and Huiguang He, *Senior Member, IEEE*

***Abstract*—Motor imagery (MI) is an important brain-computer interface (BCI) paradigm, which can be applied without external stimulus. Imagining different joint movements from the same limb allows intuitive control of the outer devices. However, few researches focused on this field, and the decoding accuracy limited the applications for practical use. In this study, we aim to use deep learning methods to explore the ceiling of the decoding performance of three tasks: the resting state, the MI of right hand and right elbow. To represent the brain functional relationships, the correlation matrix that consists of correlation coefficients between electrodes (channels) was calculated as features. We proposed the Channel-Correlation Network to learn the overall representation among channels for classification. Ensemble learning was applied to integrate the output of multiple Channel-Correlation Networks. Our proposed method achieved the decoding accuracy of up to 87.03% in the 3-class scenario. The results demonstrated the effectiveness of deep learning method for decoding MI of different joints from the same limb and the potential of this fine paradigm to be applied in practice.**

***Index Terms*—Channel-Correlation Network, Electroencephalography (EEG), Ensemble Learning, Fine Motor Imagery, Same Limb**

I. Introduction

BRAIN-computer interface (BCI) system establishes a direct communication and control channel between the brain and the external environment without relying on the peripheral nerves and muscles [1]. Electroencephalography (EEG), which detects electrical activity of the brain, has been widely used for BCI due to its high temporal resolution, cost-effectiveness, portability and noninvasive nature [2]. Compared with steady state visual evoked potentials (SSVEPs) [3] and P300 potentials [4] based BCIs, a motor imagery (MI)-based BCI depends on voluntary modulation of the subjects rather than external stimulus [5]. The MI based BCI systems have been successfully applied for the control of a wheelchair [6], a robotic quadcopter [7] and upper limb exoskeleton [8], as well as for post stroke rehabilitation [9], [10].

This work was supported by National key R&D Program of China (2018YFC2001300), National Natural Science Foundation of China (No. 81701785, 61976209), CAS Scientic Equipment Development Project (YJKYYQ20170050), Youth Innovation Promotion Association CAS and Strategic Priority Research Program of CAS (Grant No.XDB32040200). (Corresponding author: Huiguang He)

X. Ma and S. Qiu contributed equally to this work.

X. Ma, S. Qiu, W. Wei, S. Wang, and H. He are with the Research Center for Brain-Inspired Intelligence, Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing 100190, China (e-mail: {maxuelin2015, shuang.qiu, weiwei2018, wangshengpei2014, huiguang.he}@ia.ac.cn.)

H. He is also with the Center for Excellence in Brain Science and Intelligence Technology, CAS, Beijing, China

X. Ma, W. Wei, S. Wang, and H. He are with the University of Chinese Academy of Sciences (UCAS), Beijing 100049, China.

Most MI studies focused on decoding the motor intention from different limbs. The Common Spatial Pattern (CSP) method and its various extensions that based on a decomposition of the raw EEG signals into spatial patterns, have been widely used to extract the discriminative EEG features [11]–[13]. Also, power spectral density (PSD), atomic decompositions, and time-frequency (TF) energy distributions have also been extracted and used in the framework of MI classification [14]. For decoding of MI tasks, Support Vector Machine (SVM) [15]–[18] and Linear Discriminant Analysis (LDA) [16], [17], [19] were the most commonly used classifiers. Recently, a novel multi-class Support Matrix Machine was proposed and combined with time-domain-parameter features for MI classification, which obtained an accuracy of 91.6% in four MI tasks (left-hand, right-hand, feet and tongue) [20]. Yi et al. investigated the difference of the EEG patterns between simple limb MI and compound limb MI involving several parts of limbs, and reported an average accuracy of 70.43% in the 7-class scenario [21], [22].

With the development of MI, there is a growing interest in the decoding of fine imaging movement from the same limb. Salehi et al. [23] and Alazrai et al. [18] did some research on motor imagery recognition of finger gestures. In order to decode five imaginary finger movements, Alazrai et al. adopted Choi-Williams time-frequency distribution (CWD) to construct TF features and obtained an average classification accuracy of 85.5% in 18 subjects [18]. For decoding wrist movements (flexion, extension, pronation, and supination), Vuckovic et al. [24] utilized Gabor transform to extract a series of Gabor coefficients as features, and applied the Davis-Bouldin index to select feature. An average pair-wise accuracy of 80% was obtained [24]. In 2015, Edelman et al. proposed a novel EEG source imaging approach to decode four complex wrist MI tasks, and improved the 4-class accuracy from 69.1% to 81.4% [25]. In terms of the MI decoding of elbow gestures, Zhang et al. reported that it is possible to use an EEG model from one type of MI (e.g. elbow extension and flexion) for classifying EEG from other types of MI activities (e.g. open a drawer) [19].

In 2015, Yong et al. first proposed a 3-class BCI system that discriminated rest, imaginary grasp movement and imaginary elbow movement from the same limb. The average classification accuracy was 60.7% through the traditional method that combined the CSP method and SVM [17]. Later, they extracted the time-domain features and utilized the SVM to improve the classification accuracy to 74.2% in the 3-class scenario [15]. This paradigm of different joints from the same limb can provide intuitive control of outer equipments

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2019.2953121, IEEE Transactions on Neural Systems and Rehabilitation Engineering

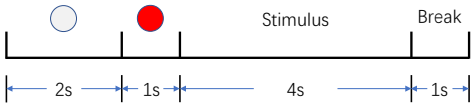


Fig. 1. Timing of one trial of our experimental paradigm.

(e.g. exoskeleton) without paying attention to any artificial relevance between imaging movement and neuroprosthetic movement. However, only these two papers from one group reported this paradigm with the best accuracy of 74.2%. One remaining challenge for its practical use is to develop novel methods to improve the classification accuracy.

Recently, deep learning method [26] has been increasingly used in the field of EEG and obtained good performance in many research areas, including emotion recognition [27], [28], seizure detection [29], [30], sleep stage classification [31], [32], and MI decoding of different limbs [33]–[36]. Compared with CSP+SVM method, Ma et al. proposed a spatial-temporal Recurrent Neural Networks (RNN) to improve the intention decoding accuracy of different limbs by 8.25% [36]. Schirrmeister et al. designed and evaluated Convolutional Neural Network (CNN) of different architectures for MI decoding through end-to-end learning, and their decoding performance reached that of the widely-used filter bank common spatial patterns (FBCSP) algorithm [34]. In 2017, Tabar et al. concatenated the TF representations of C3/CZ/C4, and then, CNN was applied to further extract features from that [33]. Their approach yielded an accuracy of 90% in the 4-class scenario. In 2017, Hajinoroozi et. al. [37], [38] used the covariance matrices as input of the square-size filter CNN model and found it outperformed the Riemannian methods. To the best of our knowledge, the deep learning methods have not yet been applied for decoding fine MI of different joints of the same limb.

In this study, we aim to use deep learning methods to improve classification performance of imagining different joint movements from the same limb. We collected EEG data of resting state, imaginary grasp movement, and imaginary elbow movement of the same limb. We proposed the Channel-Correlation Network, where two convolutional layers with channel-size filters were stacked for further representation learning from the correlation matrix features between electrodes (channels), and a dense layer was used to give the output. The outputs of multiple Channel-Correlation Networks were then integrated by ensemble learning strategy for final decoding result. The performance of the proposed method was evaluated and compared from different aspects.

II. MATERIALS

A. Subjects

Twenty-five right-handed healthy subjects (19 males and 6 females, 19-27 years old) participated in this experiment. None of them has experience of using MI-based BCI. The study was approved by the ethical committee of Institute of automation, Chinese Academy of Sciences. All subjects provided informed consent in advance.

B. Experimental Paradigm

The subjects sat in a comfortable chair at 1-meter distance in front of a computer screen. The experimental paradigm is shown in Fig. 1. Each trial (8s) began with a white circle at the center of the monitor for 2s. After that, a red circle appeared as a cue for 1s to remind the subjects of the following target. The target indication ("Hand" or "Elbow") was presented on the screen for 4s. We adopted the kinesthetic, but not visual, motor imagery paradigm in our research. The subjects were asked to concentrate on performing the indicated motor imagery task while avoiding any motion during imagination. After the imagination, "Break" was presented for 1s before next trial. Data for each subject were collected on the same day using the same cap. The experiments contained 7 sessions, involving 5 sessions consisting of 40 trials (20 trials per movement imagination) and 2 sessions consisting of 50 trials each for resting state. The sequence of two MI tasks was randomized. There were breaks of 5 to 10 minutes between sessions. Thus, there are totally 300 trials (100 trials for each type of mental state) in the dataset for the following study.

C. Data Collection and Preprocessing

EEG data were acquired using a Neuroscan SynAmps2 amplifier at a sampling rate of 1000 Hz on a 64-channel electrode cap according to the standard 10/20 System. The left mastoid reference was used for EEG measurement. Electrode impedances were kept below $10k\Omega$ during the experiment. The preprocessing of the acquired data was done with the EEGLAB toolbox (v14.1.1_b) [39] of MATLAB (R2015a) software. Common average reference (CAR) was used here in the pre-processing. The high-pass filter at 0.1Hz and low-pass filtered at 40Hz were adopted. The data were downsampled to 200Hz for reducing computational cost. A plug-in of EEGLAB, named Automatic Artifact Removal toolbox (AAR) [40], was used for automatic removal of ocular and muscular artifacts in the EEG.

III. METHOD

In this section, we first described the event-related spectral perturbation (ERSP) method for revealing the Event-Related Desynchronization (ERD) pattern from raw EEG data. Then we introduced the sliding window approach and feature extraction method for obtaining correlation and coherence feature. Importantly, we introduced the proposed Channel-Correlation Network as well as the ensemble learning for classification.

A. Event-related spectral perturbation

ERSP was first put forward by Makeig [41] in 1993 and has been widely adopted for measuring the average time course of relative changes in the spontaneous EEG amplitude spectrum induced by MI [42]. ERSP computes the data power spectrum within sliding time windows centered at time t in each trial, and then computes the average across trials. The mean ERSP of frequency f at time point t is defined as:

$$ERSP\,(f,t) = \frac{1}{n}\sum_{k=1}^{n}|F_k\,(f,t)|^2, \tag{1}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2019.2953121, IEEE Transactions on Neural Systems and Rehabilitation Engineering

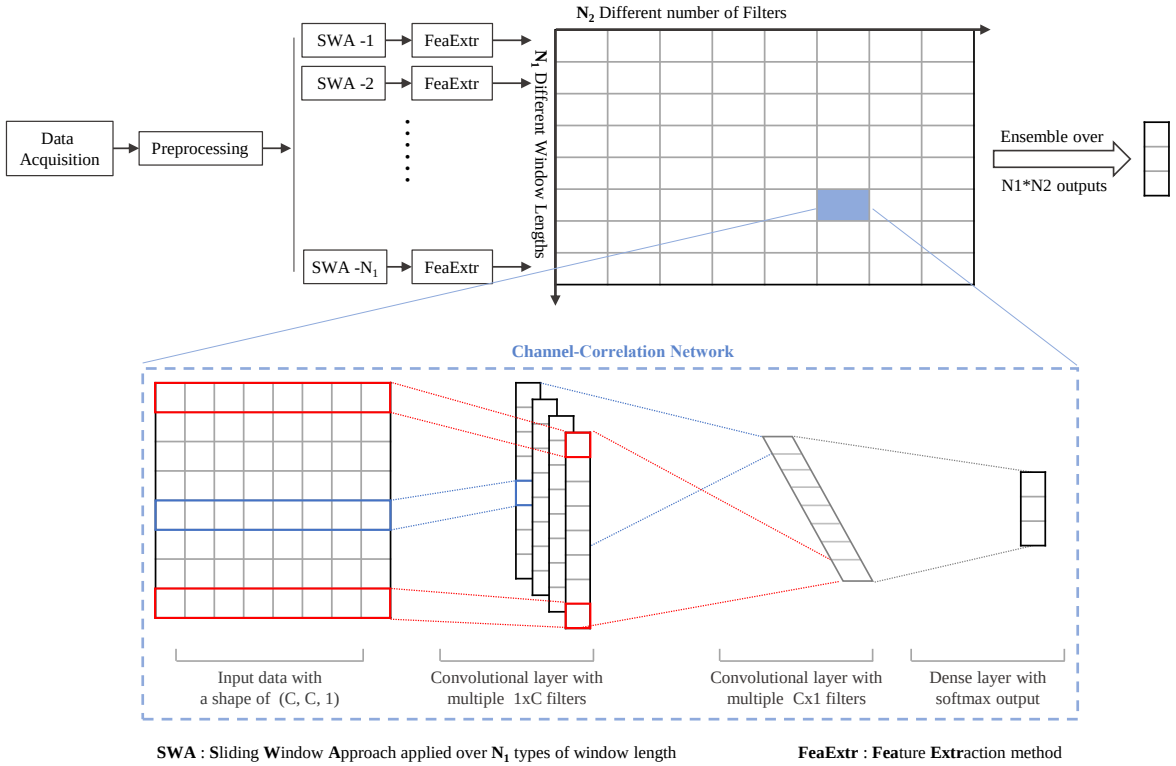


Fig. 2. A diagram of the overall framework for decoding three mental tasks. It consists of data acquisition, preprocessing, sliding window approach (SWA), feature extraction (FeaExtr), our proposed Channel-Correlation classification model, and ensemble method. The outputs of N1*N2 models were integrated to give the final result. The proposed Channel-Correlation Network is shown in the blue dashed box.

where n is the total number of trials, and $F_k(f,t)$ is the spectral estimation of frequency f at time point t in trial k.

In this study, $F_k(f,t)$ was computed using fast Fourier transform (FFT) after applying a Hanning window. The mean log power from -1s to 0s (1s before the stimulus) was used as the spectral baseline. The mean ERSP change across event-related trials (epochs) was estimated by subtracting the spectral baseline from the original log power value. For every mental task, we took -1s to 5s as the time interval and 1Hz to 35Hz as the frequency interval for presentation. In this study, the time-frequency ERD/ERS maps from three key electrode positions C3, Cz, and C4 were presented for analysis.

To explore more clear information about the ERD band range, we averaged the ERSP values across the imagination period (0ms to 4000ms), so as to obtain the power changes of EEG with frequency in different mental tasks. Then the mean power changes were computed by averaging over all subjects. In addition, topographical distribution can help us figure out which areas are involved when ERD occurs during different mental tasks. For each of 62 electrodes (except HEO, VEO), the average ERSP values in the alpha (7-14Hz) and beta (14-30Hz) frequency band and time interval (0-4s) were calculated separately. We plot the topo-distribution maps according to the electrode positions and the corresponding group-level ERSP values among all 25 subjects.

B. Feature Extraction

In this study, we applied the sliding window approach with a fixed stride of 10 points to divide the trial-wise samples into slice samples. Each clip had a fixed length of S with (S-10) points overlapping between continuous neighbors. The value range of S was set according to the feature extraction methods.

To represent the brain functional relationships, the correlation and coherence for each channel pair were calculated [43]. The correlation matrix and coherence matrix would be used as input of our model separately.

1) Pearson Correlation Coefficient: Given a slice of data with C channels and S time points $x_k \in \mathbb{R}^{C\times S}$, the correlation coefficient matrix R can be calculated as follows:

$$R_{ij} = \frac{Cov_{ij}}{\sqrt{Cov_{ii} * Cov_{jj}}} \in \mathbb{R}, \tag{2}$$

where $1 \le i \le C$, $1 \le j \le C$. The covariance matrix element $Conv_{ij}$ is the covariance of d_i and d_j. The element $Conv_{ii}$ and $Conv_{jj}$ represents the variance of d_i and d_j, respectively. Therefore, the correlation matrix R is a square matrix of order C. To evaluate the influence of window length, the slice length S of correlation features was set over a wide range of values: $[10, 20, 30, 40, 50, 60, 100, 150, 200, 250, 300, 350, 400]$.

2) Coherence Frequency Band Feature: Given a slice data $x_k \in \mathbb{R}^{C\times S}$, the magnitude squared coherence of d_i and d_j can be calculated as follows:

$$Coh_{ij} = \frac{|P_{ij}|^2}{(P_{ii} * P_{jj})} \in \mathbb{R}^{1\times N_{freq}}, \tag{3}$$

where $1 \le i \le C$, $1 \le j \le C$, N_{freq} is the number of sample frequencies, P_{ii} and P_{jj} are power spectral density estimates of d_i and d_j, and P_{ij} is the cross spectral density estimate of d_i and d_j. The PSD was estimated using Welch's [44] method.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2019.2953121, IEEE Transactions on Neural Systems and Rehabilitation Engineering

TABLE I
CHANNEL-CORRELATION NETWORK ARCHITECTURE, WHERE C = NUMBER OF CHANNELS, F_1 = NUMBER OF THE FILTERS IN BLOCK 1, F_2 = NUMBER OF THE FILTERS IN BLOCK 2, AND N = NUMBER OF CLASSES, RESPECTIVELY.

Block	Layer	# filters	size	# params	Output	Activation	Regularization	Options
1	Input				(C, C)			
	Reshape				(1, C, C)			
	Conv2D	F_1	(1, C)	$(1 * C + 1) * F_1$	(F_1, C, 1)	Linear	L2 norm	mode = valid
	Activation				(F_1, C, 1)	Relu	L1 norm	
	BatchNorm			$4 * F_1$	(F_1, C, 1)			
	Dropout				(F_1, C, 1)			$p = 0.3$
2	Conv2D	F_2	(C, 1)	$(F_1 * C * 1 + 1) * F_2$	(F_2, 1, 1)	Linear	L2 norm	mode = valid
	Activation				(F_2, 1, 1)	Relu	L1 norm	
	BatchNorm			$4 * F_2$	(F_2, 1, 1)			
	Flatten				(F_2)			
Classifier	Dense	(F_2+1) * N			N	Softmax	L2 norm	

To explore the alpha (7-14Hz) and beta (14-30Hz) frequency band, the magnitude squared coherence within these two band were averaged separately to obtain the Alpha Matrix and Beta Matrix:

$$Alpha_{ij} = \frac{1}{N_\alpha} \sum_{freq \in \alpha} Coh_{ij} \in \mathbb{R}, \tag{4}$$

$$Beta_{ij} = \frac{1}{N_\beta} \sum_{freq \in \beta} Coh_{ij} \in \mathbb{R}. \tag{5}$$

Therefore, the Alpha Matrix and Beta Matrix were both $C \times C$ matrices. Here, we set the window length of the coherence feature in the range of $[200, 250, 300, 350, 400]$, which ensured sufficient points to calculate the time-frequency transform.

C. Channel-Correlation Network

As a kind of well-developed deep learning framework, the CNN has been applied in various fields [26], including EEG signal decoding tasks [34], [45]. CNNs have been proved to be capable of dealing with 2-D or 3-D matrices using custom convolution filter kernels. In this paper, we designed the CNN model with two 1-D convolutional layers, where the filter size was set identical with the number of the EEG electrodes (channels) to learn from all the channels simultaneously. This model is referred as **Channel-Correlation Network**.

As shown in Fig. 2, the proposed Channel-Correlation Network takes the symmetric square matrices of order C as input, and outputs the predicted labels. The input of the model can be denoted as $x \in \mathbb{R}^{C \times C}$. The first 1-D convolutional layer of Channel-Correlation network contains F_1 filters. These filters are all in the size of row-like $1 \times C$, which can be applied over x to extract the global information from all the C electrodes. The output of each filter has a shape of $C \times 1$ and the output of the first CNN layer is size of $C \times 1 \times F_1$:

$$\text{CNN-layer-1: } f_{cnn_1} = C_{1D\text{-row}}(x),\ f_{cnn_1} \in \mathbb{R}^{C \times 1 \times F_1} \tag{6}$$

In a similar way, F_2 column-like $C \times 1$ filters were applied over f_{cnn_1} to integrate all the C channel further in the second convolutional layer:

$$\text{CNN-layer-2: } f_{cnn_2} = C_{1D\text{-col}}(f_{cnn_1}),\ f_{cnn_2} \in \mathbb{R}^{1 \times 1 \times F_2} \tag{7}$$

The output of the second convolutional layer were flattened before being fed into the dense layer:

$$\text{Flatten-layer: } f_{Flat} = Flatten(\forall f_{cnn_2}) \in \mathbb{R}^{F_2 \times 1} \tag{8}$$

A dense layer with size n (n is the number of the classes) followed by softmax activation function were applied to generate the class labels:

$$\text{Dense-layer: } f_{Dense} = Dense(f_{Flat}) \in \mathbb{R}^{n \times 1} \tag{9}$$

$$\text{Output-layer: } f_{out} = Softmax(f_{Dense}) \in \mathbb{R}^{n \times 1} \tag{10}$$

In this paper, the numbers of filters in the two convolutional layers were set identical, which was denoted as F. To verify the effect of the filter number, F was set in a wide range from 30 to 150 with a stride of 20. In the training phase, the slice data from training trial samples were shuffled and fed to the network. In the inference phase, the predicted labels of multiple slice data from the identical trial sample determined the corresponding predicted trial label through the majority voting strategy.

D. Ensemble Learning

Ensemble learning, also known as multi-classifier system or committee-based learning, trains combinations of base models to obtain better performance [46]. Benefitting from its stable and superior performance, ensemble learning has been widely used by the champions of many data mining competitions (e.g. Kaggle competitions).

In our study, the window length S of sliding window approach and the filter number F of Channel-Correlation Network were set in value ranges with $N1$ and $N2$ possible values, respectively. Thus, the different combinations of S and F would result in $N1 \times N2$ different classifiers (as shown in Fig. 2). These different classifiers could be used as basic classifiers in ensemble learning. The predicted labels of these $N1 \times N2$ classifiers were integrated with voting-based ensemble methods to give the final result.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2019.2953121, IEEE Transactions on Neural Systems and Rehabilitation Engineering

IV. EXPERIMENTS

A. Experimental Setup

The Channel-Correlation Networks were implemented with the Keras framework and trained on a Nvidia 1080Ti GPU from scratch in a fully-supervised manner. The Adam algorithm [47] was used to optimize the cross-entropy loss function with the default parameters mentioned in the original paper [47]. The L_1/L_2 norm regularization, Batch Normalization [48] and Dropout [49] techniques were used to avoid model overfitting. The L_1 and L_2 norm were both $1 * e^{-7}$. The dropout probability was 0.3. In addition, an exponential decay of learning rate (reduced by a factor of 1,000 after 600 iterations) was applied and the iteration number was fixed at 600 during training.

B. Compared Methods

1) baseline methods: The CSP [11] and FBCSP [50] spatial filter methods followed by LDA classifier were adopted as the baseline methods and denoted as CSP+LDA and FBCSP+LDA, respectively. Since the CSP [11] is designed for two-class problem, in our case of three-class decoding tasks, a one-vs-one strategy was appointed for both CSP and FBCSP. The CSP/FBCSP methods and LDA classifier were implemented in Python using the MNE [51] and scikit-learn [52] libraries respectively with the default parameter settings.

2) state-of-the-art: We gave a brief introduction to the state-of-the-art compared methods. We gave priority to the source code provided by the authors and ran the compared methods directly on our dataset. If the source code is not provided, we re-implemented them according to the descriptions in the corresponding papers. We organized our data appropriately according to the input format required by these models. Specifically, the sampling rate of input data, the cropping strategy for training and the hyper-parameter setting of models were consistent with the article cited.

- Tavakolan et. al. [15]: In 2017, Tavakolan et. al. [15] used time-domain features for classification and reported a state-of-the-art result for the MI decoding of different joints from the same limb. In their study, three types of time-domain features were computed based on signal amplitudes: Auto-Regression (AR) model coefficients, the waveform complexity (WL), and root mean square (RMS). And, SVM was used for classification.
- parallel CNN-RNN [45]: The parallel convolutional recurrent network structure contains two parts, CNN and RNN, for spatial and temporal feature extraction respectively. The parallel model extracts the spatial and temporal features of EEG signals in parallel and fuses the extracted features for final intention recognition. We adopted the source code (based on TensorFlow framework) provided by the authors of [45] (https://github.com/dalinzhang/Cascade-Parallel).
- FBCSP-C2CM [53]: Sakhavi et al. introduced an envelope representation based on FBCSP method for EEG data. After that, they utilized two convolutional layers for learning temporal information followed by one convolutional layer for learning spatial information (C2CM). We used the raw EEG envelope (R1) as the representations for the EEG (described in the section III.B of [53]).
- EEGNet [54]: The EEGNet contains three one-dimensional convolutional layers. The EEGNet-8,2 model (8 temporal filters, 2 spatial filters) with kernel length of 32 samples was used in this paper. We used the 2s window width and 8 data points stride to slide the trial samples and applied the maximum voting method on the slice sample predictions to get the trial predicted label and corresponding accuracies. We adopted the source code (based on PyTorch framework) provided by the authors of [54] (https://github.com/vlawhern/arl-eegmodels).
- Deep ConvNet and Shallow ConvNet [34]: The Deep ConvNet had four convolution-max-pooling blocks, with a special first block designed to handle EEG input, followed by three standard convolution-max-pooling blocks and a dense softmax classification layer. Both Deep ConvNet and Shallow ConvNet were designed for decoding imagined or executed tasks from raw EEG. We adopted the source code (based on PyTorch framework) provided by the authors of [34] (https://github.com/TNTLFreiburg/braindecode).
- Square-CNN [38]: In convolutional layers, squared kernels of size $3 * 3$ were applied to learn the local features in the covariance matrices. The square-CNN model contained two convolutional layers, a fully connected layer and a softmax layer for classification.

In order to eliminate the impact of data partitioning on experimental results, we first used a category balanced five-fold cross-validation method to split the training/test data. And then, all experiments used these consistent training and testing data. In each fold, 20% of the trial samples for each session were used as test data, and the remaining 80% of the trial samples were used as training data. This ensured that all trial samples were available and only used once as test data. In the case of slice-wise input format (for examples, the EEGNet and the proposed Channel-Correlation Network), the divided trial sample was segmented into multiple slice samples and input into the network using a sliding window method. This ensured that the slice samples from the identical trial sample were either assigned to the training set or to the test set. The reported accuracies were calculated based on trial labels.

C. Statistical Analysis

In this paper, Friedman test was conducted to analyze the effect of different single-model methods, feature types, window lengths and the number of convolutional filters on classification accuracies. Post hoc analysis was conducted with a Bonferroni correction. The significant level was set at 0.05. All results are presented as $mean \pm std$ deviation.

V. RESULT

A. ERD/ERS during MI

The TF maps (Fig. 3) show obvious ERD patterns in both alpha (8-13Hz) and beta (20-25Hz) bands from 0 to 4s after MI task onset for two MI except rest. In order to investigate the activation of motor areas, the averaged topographical

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2019.2953121, IEEE Transactions on Neural Systems and Rehabilitation Engineering

TABLE II
COMPARISON OF SINGLE-MODEL ACCURACY.

Method	Year	Training Mode	Sampling Rate	Window Length	Stride	Accuracy (%)	Median (%)
CSP [11] +LDA	2000	Trial-wise	200Hz	Trial (4s)	-	63.91±4.26	64.67
FBCSP [50] +LDA	2008	Trial-wise	200Hz	Trial (4s)	-	66.69±2.80	66.67
Tavakolan et. al. [15]	2017	Trial-wise	250Hz	Trial (4s)	-	63.28±4.52	63.67
FBCSP-C2CM [53]	2018	Trial-wise	250Hz	Trial (4s)	-	**69.96±2.10**	70.00
Square-CNN [38]	2017	Slice-wise	200Hz	10 points (50ms)	10 points	65.55±6.29	66.33
Parallel CNN-RNN [45]	2018	Slice-wise	250Hz	10 points (40ms)	10 points	**72.31±7.96**	74.00
Ours-Corr+CNN-10	-	Slice-wise	200Hz	10 points (50ms)	10 points	68.88±4.04	69.00
Deep ConvNet [34]	2017	Slice-wise	250Hz	522 points (∼2s)	1 point	59.65±6.48	60.67
Shallow ConvNet [34]	2017	Slice-wise	250Hz	534 points (∼2s)	1 point	63.03±6.79	64.67
Square-CNN [38]	2017	Slice-wise	200Hz	400 points (2s)	10 points	66.73±4.96	66.67
EEGNet [54]	2018	Slice-wise	128Hz	256 points (2s)	8 points	69.25±6.57	71.00
Ours-Corr+CNN-400	-	Slice-wise	200Hz	400 points (2s)	10 points	**73.64±4.01**	73.00
Ours-Corr+CNN-150	-	Slice-wise	200Hz	150 points (0.75s)	10 points	**75.03±3.37**	74.33

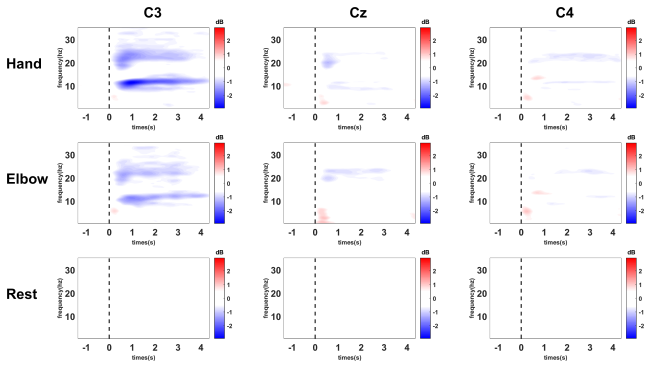


Fig. 3. Averaged time-frequency maps among 25 subjects at three electrode locations (C3/CZ/C4) during three state (imagining "Hand" movement, imagining "Elbow" movement and "Rest") and. Blue indicates ERD.

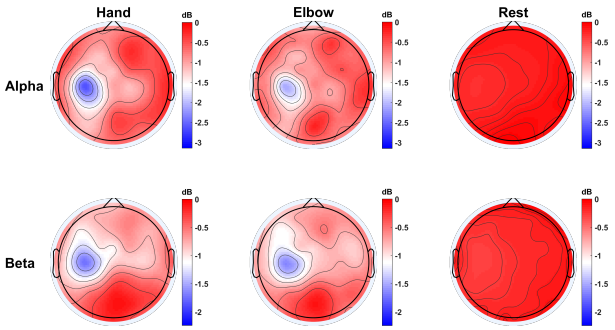


Fig. 4. Averaged topographical distribution of power in the upper alpha band (10-12Hz) and upper beta band (23-25Hz) three state (imagining "Hand" movement, imagining "Elbow" movement and "Rest") and Blue indicates ERD.

distribution in the alpha and beta band were computed and depicted in Fig. 4. We can see obvious and similar ERD over the contralateral motor areas for the imagining movement tasks of hand and elbow. It presents a significant contralateral dominance. The maps during resting state does not appear ERD patterns. This result means that both of the hand and elbow MI activated the motor areas of the brain.

Fig. 5 shows the comparison of power changes with frequency for electrode C3. As revealed in Fig. 5, significant difference between the imagining of hand and elbow movements was found from 10.9Hz to 12.1Hz.

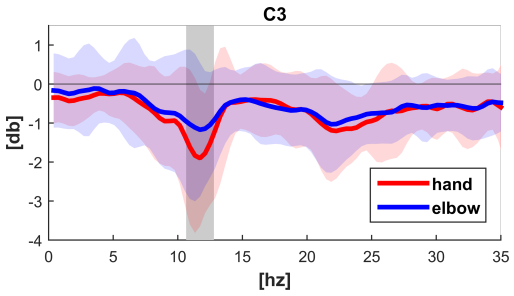


Fig. 5. The comparison of power changes with frequency during hand and elbow motor imagery for electrode C3. The paired t-test was calculated and the significant difference between two conditions are shaded by grey blocks ($p < 0.05$).

B. Single-Model Comparison

The detailed single-model comparison between the proposed Corr+CNN method and the compared methods is shown in Table II. In the trial-wise, 10 points, and 2s input configurations, different methods have significant influence on the accuracy ($\chi^2(3) = 35.204, p < 0.001$, $\chi^2(2) = 15.657, p < 0.001$, and $\chi^2(4) = 57.133, p < 0.001$, respectively). In the trial-wise input configuration, FBCSP-C2CM method significantly outperformed other methods (Bonferroni correction of 6, all $p < 0.05$). In the 10 points input configuration, the Parallel CNN-RNN method significantly outperformed Tavakolan et. al. [15] and Corr+CNN-10 (Bonferroni correction of 3, both $p < 0.001$). In the 2s input configuration, our proposed method outperformed other four compared methods (Bonferroni correction of 10, all $p < 0.05$). The optimal result (75.03±3.37%) of our single model was obtained with Corr+CNN-150 for filter numbers of 130. The Corr+CNN-150 method significantly outperformed FBCSP-C2CM, the Parallel CNN-RNN, and the Corr+CNN-400 methods (Wilcoxon sign rank test, all $p < 0.05$). The confusion matrices of these single-model methods are shown in Fig. A in the Supplementary Information.

C. Comparison between Correlation and Coherence

To compare the performance of coherence and correlation features, we extracted features under different conditions. For frequency domain, we computed the coherence matrix of EEG

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2019.2953121, IEEE Transactions on Neural Systems and Rehabilitation Engineering

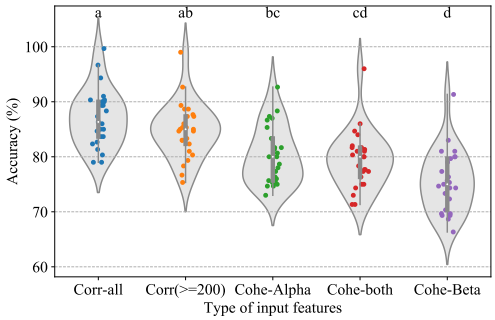
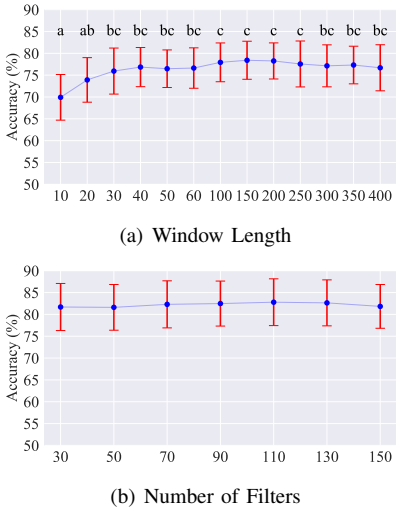


Fig. 6. Accuracy comparison between Coherence and Correlation features. *Cohe-beta*, *Cohe-alpha* and *Cohe-both* represent coherence features on beta band (14-30Hz), alpha band (7-14Hz) and both of them (7-30Hz), respectively. *Corr(*$\geq$ 200*)* and *Corr-all* represent the correlation features of window length in 200$\sim$400 points and 10$\sim$400 points, respectively. Factors marked with the same lowercase letter indicate no significance between each other.



(a) Window Length

(b) Number of Filters

Fig. 7. Parameter sensitivity analysis of (a) window length and (b) the number of filters for the Corr+CNN method. Factors marked with the same lowercase letter indicate no significantness between each other.

data in alpha, beta frequency band, and both of them (from 7Hz to 30Hz). For time domain, we extracted the correlation matrix on window length from 10 to 400 points and that on window length from 200 to 400 points. In the latter case of time domain, the size of the window length was consistent with the coherence feature. These features were taken as input of the proposed CNN model separately, and the comparison of their ensemble accuracy was shown in Fig. 6. The confusion matrices of these five methods are shown in Fig. B in the Supplementary Information.

The features had significant influence on the classification performance of Channel-Correlation Network ($\chi^2(4) = 69.492, p < 0.001$). The post-hoc analysis (Bonferroni correction of 10) showed that, in the case of 200$\sim$400 window length, the correlation feature (median of 85.00%) did not significantly outperformed the coherence feature on alpha band (median of 80.00%) but significantly higher than that on beta band (median of 75.00%) or full band (median of 80.00%). And the accuracy of Cohe-alpha-CNN was significantly higher than that of Cohe-beta-CNN. These results suggested that coherence feature in alpha band may play an important role than that in beta band in decoding fine MI of hand and elbow from the same limb. Moreover, the accuracy of Corr-all-CNN (median of 86.33%) was significantly higher than than those of three Coherence-based features.

D. Evaluation of Ensemble Strategy

To analyze the influence of ensemble learning on the accuracy, we took Corr+CNN method as an example to compare the classification accuracies obtained with and without ensemble learning. The detailed averaged accuracies of Corr+CNN method among 25 subjects are listed in the Tab. III.

For each number of filters (each row in Tab. III), the classification result (the last cell in each row) integrating Corr+CNN models with different window lengths by ensemble learning, was higher than that of models without ensemble learning. Similarly, for each window length (column in Tab. III), the ensemble classification result over different number of filters, outperformed that obtained without ensemble learning. And the classification accuracy integrating all Corr+CNN models with different window lengths and different number of filters (the last in the Tab. III) was superior to that with integrating base on window length or number of filters, as well as that of without ensemble. For Beta+CNN and Alpha+CNN method, ensemble learning also revealed great advantage for classification accuracy. Therefore, for our proposed Channel-Correlation Network, the ensemble learning method could greatly improve the MI decoding performance of different joints from the same limb.

E. Parameter Sensitivity Analysis

Fig. 7(a) shows the classification accuracies of Corr+CNN methods under different window length. For the Corr+CNN method, the window length had a significant influence on the classification performance of Channel-Correlation Network ($\chi^2(12) = 106.368, p < 0.001$). The post-hoc analysis (Bonferroni correction of 78) showed that the classification results of 10 and 20 points (median of 69.30%, 75.00%, respectively) were significantly lower than those of larger window lengths. The averaged accuracy increased gradually from 10 to 40 points and became stable from 40 points on. From 30 points to 400 points, the classification accuracy had no significant difference between each other. The performances under 10 and 20 points window length were still higher than 70%. Therefore, all classifiers with these different window lengths could be taken into account in the ensemble learning.

Fig. 7(b) shows the classification accuracies under different number of filters. For the Corr+CNN method, the number of filters had no significant influence on the classification performance of Channel-Correlation Network ($\chi^2(6) = 10.278, p = 0.113$). It demonstrates that our proposed Channel-Correlation Network can work robustly with different numbers of filters ranging from 30 to 150.

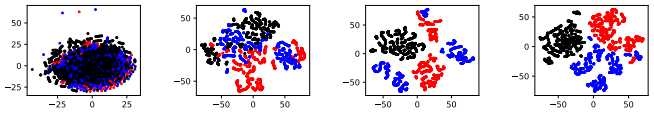
F. Data Visualization

T-distributed Stochastic Neighbor Embedding (t-SNE) [55] was used to embed the data into two dimensions for drawing a scatter plot. Fig. 8 shows an example of visualization from

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2019.2953121, IEEE Transactions on Neural Systems and Rehabilitation Engineering

TABLE III
AVERAGE ACCURACIES WITHOUT ENSEMBLE LEARNING AND WITH DIFFERENT ENSEMBLE STRATEGIES. THERE ARE 7 DIFFERENT FILTER NUMBERS AND 13 DIFFERENT WINDOW LENGTHS. THE BACKGROUND OF THE CELLS IS SET TO A DIFFERENT DEPTH DEPENDING ON THE ACCURACY.

Acc(%) WinLen / #Filters	10	20	30	40	50	60	100	150	200	250	300	350	400	Win-Ensemble
30	68.27	71.35	71.93	72.79	72.75	72.72	74.81	74.39	73.85	73.72	73.48	73.13	72.91	81.69
50	68.25	71.31	71.97	73.28	73.20	73.87	74.36	74.04	74.20	74.13	72.99	72.84	73.17	81.61
70	67.75	70.67	72.51	72.65	73.80	73.00	74.57	75.00	74.04	73.75	73.69	73.91	73.09	82.29
90	68.88	71.07	72.87	72.91	73.44	73.88	74.39	74.63	74.45	73.69	73.20	73.57	73.60	82.48
110	67.95	70.75	72.47	73.16	73.08	73.12	74.16	74.71	74.77	74.04	73.80	72.97	73.07	**82.80**
130	68.27	70.65	72.99	73.52	73.15	73.37	74.39	**75.03**	74.49	73.27	74.11	74.37	73.64	82.63
150	68.71	70.72	73.08	72.44	72.47	73.00	73.95	74.64	74.75	74.15	73.56	73.80	73.40	81.83
Filter-Ensemble	69.93	73.92	75.93	76.85	76.48	76.64	77.93	**78.43**	78.27	77.56	77.13	77.32	76.68	**87.03**



(a) Raw Data (b) Correlation Feature (c) Output of the First Layer (d) Output of the Second Layer

Fig. 8. An example of t-SNE Visualization for raw data (a), correlation matrix feature (b), the output of the first convolutional layer (c) and the second convolutional layer (d), from one subject. Red, Blue and Black points represent hand movement, elbow movement and resting state, respectively.

one subject. There is a large amount of overlap between the different categories in the raw data (Fig. 8(a)). After extracting correlation matrix feature, the samples of each class clustered into multiple block mass obviously but still cannot be linearly separated (Fig. 8(b)). As the correlation matrix features are processed over the convolutional layers, the separation between three classes after the first convolutional layer is clearly visible in Fig. 8(c), and, the output of the second convolutional layer is linearly separable in Fig. 8(d). Thus, the visibility of separation increases from raw data to the output of Corr+CNN method. Both of correlation feature extraction method and Channel-Correlation Network have great importance in decoding MI tasks of different joints from the same limb.

VI. DISCUSSION

In this study, correlation matrix and coherence matrix were calculated as features. And we proposed a novel Channel-Correlation Network to classify resting state, imagining right hand movement and imagining right elbow movement. Combining the correlation feature and Channel-Correlation Network performed best and obtained a decoding accuracy up to 87.03% in the 3-class scenario.

CSP and FBCSP features have been widely adopted in previous studies for decoding MI of different limbs [11]–[13]. However, they performed poor in our study for decoding imagery hand and elbow movements. This may be due to the similar spatial distributions during MI of hand and elbow from the same upper limb. This has been shown in Fig. 3 and Fig. 4, and has also been reported in previous studies [15]. Therefore, CSP-based spatial filtering method was less effective in discriminating two classes of fine MI.

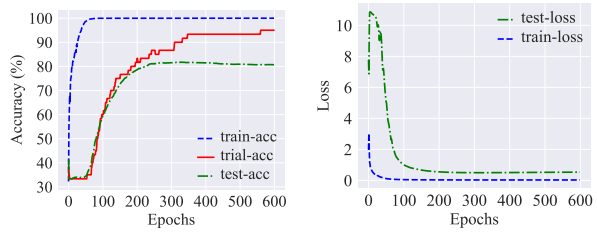
Our proposed Channel-Correlation Network obtained comparable or superior performances than many other compared methods in the decoding of resting state, MI of hand and elbow from the same limb. It can be attributed to several aspects.

Firstly, the correlation matrix represents the global functional relationships of the brain. This feature not only includes information from single channel, but also adds the connectivity information from each channel pair, which may be better than the features extracted from every single channel. From the visualization of the raw data (Fig. 8(a)) and the correlation features (Fig. 8(b)), it is obvious that it became more linearly separable after extracting the correlation matrix features.

Secondly, the designed Channel-Correlation Network can further extract useful information for classification from the correlation coefficient matrix. The correlation matrix feature with 62×62 size was taken as input of our model. Through the first two convolutional layers, a global representation was got to make three mental tasks maximally separable. The dense layer with softmax activation was used to generate the final output. Our single Channel-Correlation Network with 400-points (2s) input configuration ($75.63 \pm 3.98\%$) outperformed other compared baseline methods and deep-learning based methods. It revealed the effectiveness of our proposed single Corr+CNN method. In addition, our proposed Channel-Correlation Network was small, whose trainable parameter number ranged from 0.06M to 1M. It is much less than that of the compared parallel CNN-RNN method [45] (about 10M). Moreover, cooperating with the exponential decline strategy of learning rate, the proposed Channel-Correlation Networks converged stably, as shown in Fig. 9. The proposed Corr+CNN method can work robustly with various window length and various number of filters. Thus, the ensemble learning can be used easily to further improve the classification results. These results demonstrate that, our proposed Channel-Correlation Network can be used as a base learner for ensemble learning in the EEG MI decoding scenario.

Lastly, we integrated the output of totally 91 Corr+CNN classifiers to get the final result and the classification accuracy was significantly improved in this way. Ensemble learning over multiple Corr+CNN classifiers based on different window lengths and numbers of filters can improve the final classification results by reducing the variance of the estimate of the output class posteriors. Furthermore, the ensemble learning is unlikely to overfit, so the generalization performance is good. Therefore, Corr+CNN achieved a high accuracy of 87.03% for decoding three mental tasks. Interestingly, in the proposed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2019.2953121, IEEE Transactions on Neural Systems and Rehabilitation Engineering



(a) Accuracy Curves (b) Loss Curves

Fig. 9. Learning curves drawn from the training and testing phases of a Channel-Correlation Network. (a) shows the training slice samples accuracy (blue dotted line), testing slice samples accuracy (green dotted line) and testing trial samples accuracy (red solid line) growth with training times. (b) shows the training and testing loss (blue and green dotted line, respectively) decaying with the training times.

Channel-Correlation method, the classification accuracy of imagining hand movement was significantly higher than that of imagining elbow movement. Generally, hand movements are more familiar to persons than elbow movements. So, subjects may imagine hand movements more easily, which caused higher accuracy. In addition, the brain motor areas activated by the hand movement are larger than that of elbow movement [56]. This may contribute to that imaging hand movements could be better discriminated.

To verify the online compatibility of Channel-Correlation Networks, we used the first four task sessions and the first rest session as training data to retrain the networks from scratch. These online models were then tested on the last task session and the first 20 trial samples of the last rest session, keeping the category-balance in the testing data. In this simulate online experiment, the Corr+CNN method obtained an average decoding accuracy of $83.93 \pm 9.21\%$ over all the 25 subjects in the 3-class scenario. This demonstrates that our Corr+CNN method could be used for online systems. We do, however, anticipate a lower accuracy in real online experiment as the hyper-parameters of our ensemble models were manually tuned using 5-fold cross-validation, which included the test data of the online simulation.

Imagining moving different limbs for controlling often leads to the cognitive disconnection between the motion intention and the end-effector [25]. For example, opening and closing the hand orthosis by imagination of right hand and both feet movement respectively were unnatural [57]. The MI of fine movement from the same limb can provide a more intuitive control of outer equipments (e.g. exoskeleton) without paying attention to any artificial relevance between imaging movement and neuroprosthetic movement. The proposed fine MI-based BCI system can be applied to effectively controlling the robotic system with more degree of freedom, as well as providing intuitive and targeted training for the rehabilitation of stroke patients.

Only two imaginary joint movements were included and analyzed in this study. The MI of wrist and shoulder from the same limb, should be added to study in future research. There have been some studies about applying deep learning method for EEG decoding with end-to-end learning [34], [36], [45]. Future study should consider raw data of EEG as input of deep learning method for the MI decoding of different joints from the same limb.

VII. Conclusion

In this study, we used the correlation matrix feature and the coherence matrix feature to represent the functional relationship during mental tasks and developed an ensemble Channel-Correlation Network to enhance MI decoding of different joints from the same limb. The ensemble Channel-Correlation Network with the correlation matrix feature achieved the performance of up to 87.03%, which is encouraging for the pursuit of subsequent work. Comparison experiments and simulated online experiments demonstrated that MI of hand and wrist from the same limb and resting state could be classified with high accuracy and can be enhanced through deep learning methods. This work makes an important developing step toward fine MI for practical BCI system research.

References

[1] J. J. Shih, D. J. Krusienski, and J. R. Wolpaw, "Brain-computer interfaces in medicine," in *Mayo Clinic Proceedings*, vol. 87, no. 3. Elsevier, 2012, pp. 268–279.

[2] W.-P. Teo and E. Chew, "Is motor-imagery brain-computer interface feasible in stroke rehabilitation?" *PM&R*, vol. 6, no. 8, pp. 723–728, 2014.

[3] G. R. Muller-Putz and G. Pfurtscheller, "Control of an electrical prosthesis with an ssvep-based bci," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 361–364, 2008.

[4] J. Polich, "Updating p300: an integrative theory of p3a and p3b," *Clinical neurophysiology*, vol. 118, no. 10, pp. 2128–2148, 2007.

[5] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[6] K. Choi and A. Cichocki, "Control of a wheelchair by motor imagery in real time," in *International conference on intelligent data engineering and automated learning*. Springer, 2008, pp. 330–337.

[7] K. LaFleur, K. Cassady, A. Doud, K. Shades, E. Rogin, and B. He, "Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface," *Journal of neural engineering*, vol. 10, no. 4, p. 046003, 2013.

[8] A. Frisoli, C. Loconsole, D. Leonardis, F. Banno, M. Barsotti, C. Chisari, and M. Bergamasco, "A new gaze-bci-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1169–1179, 2012.

[9] K. K. Ang, K. S. G. Chua, K. S. Phua, C. Wang, Z. Y. Chin, C. W. K. Kuah, W. Low, and C. Guan, "A randomized controlled trial of eeg-based motor imagery brain-computer interface robotic rehabilitation for stroke," *Clinical EEG and neuroscience*, vol. 46, no. 4, pp. 310–320, 2015.

[10] A. Dunsky and R. Dickstein, "Motor imagery training for gait rehabilitation of people with post-stroke hemiparesis: Practical applications and protocols," *Global Journal of Health Science*, vol. 10, no. 11, 2018.

[11] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[12] S. Sun and C. Zhang, "Adaptive feature extraction for eeg signal classification," *Medical and Biological Engineering and Computing*, vol. 44, no. 10, pp. 931–935, 2006.

[13] N. Robinson, A. P. Vinod, K. K. Ang, K. P. Tee, and C. T. Guan, "Eeg-based classification of fast and slow hand movements using wavelet-csp algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2123–2132, 2013.

[14] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 4, pp. 317–326, 2008.

[15] M. Tavakolan, Z. Frehlick, X. Yong, and C. Menon, "Classifying three imaginary states of the same upper extremity using time-domain features," *PloS one*, vol. 12, no. 3, p. e0174161, 2017.

[16] A. Suwannarat, S. Pan-ngum, and P. Israsena, "Comparison of eeg measurement of upper limb movement in motor imagery training system," *Biomedical engineering online*, vol. 17, no. 1, p. 103, 2018.

[17] X. Yong and C. Menon, "Eeg classification of different imaginary movements within the same limb," *PloS one*, vol. 10, no. 4, p. e0121896, 2015.

[18] R. Alazrai, H. Alwanni, Y. Baslan, N. Alnuman, and M. I. Daoud, "Eeg-based brain-computer interface for decoding motor imagery tasks within the same hand using choi-williams time-frequency distribution," *Sensors*, vol. 17, no. 9, p. 1937, 2017.

[19] X. Zhang, X. Yong, and C. Menon, "Evaluating the versatility of eeg models generated from motor imagery tasks: An exploratory investigation on upper-limb elbow-centered motor imagery tasks," *PloS one*, vol. 12, no. 11, p. e0188293, 2017.

[20] Q. Zheng, F. Zhu, J. Qin, and P.-A. Heng, "Multiclass support matrix machine for single trial eeg classification," *Neurocomputing*, vol. 275, pp. 869–880, 2018.

[21] 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," *Journal of neuroengineering and rehabilitation*, vol. 10, no. 1, p. 106, 2013.

[22] W. Yi, S. Qiu, K. Wang, H. Qi, L. Zhang, P. Zhou, F. He, and D. Ming, "Evaluation of eeg oscillatory patterns and cognitive process during simple and compound limb motor imagery," *PloS one*, vol. 9, no. 12, p. e114853, 2014.

[23] S. S. M. Salehi, M. Moghadamfalahi, F. Quivira, A. Piers, H. Nezamfar, and D. Erdogmus, "Decoding complex imagery hand gestures," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*. IEEE, 2017, pp. 2968–2971.

[24] A. Vuckovic and F. Sepulveda, "Delta band contribution in cue based single trial classification of real and imaginary wrist movements," *Medical & biological engineering & computing*, vol. 46, no. 6, pp. 529–539, 2008.

[25] B. J. Edelman, B. Baxter, and B. He, "Eeg source imaging enhances the decoding of complex right-hand motor imagery tasks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 1, pp. 4–14, 2016.

[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[27] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[28] J. Li, Z. Zhang, and H. He, "Hierarchical convolutional neural networks for eeg-based emotion recognition," *Cognitive Computation*, pp. 1–13, 2018.

[29] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals," *Computers in biology and medicine*, vol. 100, pp. 270–278, 2018.

[30] P. Thodoroff, J. Pineau, and A. Lim, "Learning robust features using deep learning for automatic seizure detection," in *Machine learning for healthcare conference*, 2016, pp. 178–190.

[31] A. Supratak, H. Dong, C. Wu, and Y. Guo, "Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel eeg," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998–2008, 2017.

[32] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 4, pp. 758–769, 2018.

[33] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of eeg motor imagery signals," *Journal of neural engineering*, vol. 14, no. 1, p. 016003, 2016.

[34] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[35] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted boltzmann machines," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 25, no. 6, pp. 566–576, 2017.

[36] X. Ma, S. Qiu, C. Du, J. Xing, and H. He, "Improving eeg-based motor imagery classification via spatial and temporal recurrent neural networks," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 1903–1906.

[37] M. Hajinoroozi, J. Zhang, and Y. Huang, "Prediction of fatigue-related driver performance from eeg data by deep riemannian model," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, pp. 4167–4170.

[38] M. Hajinoroozi, J. M. Zhang, and Y. Huang, "Driver's fatigue prediction by deep covariance learning from eeg," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2017, pp. 240–245.

[39] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[40] G. Gómez-Herrero, W. De Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Van Huffel, and W. Van Paesschen, "Automatic removal of ocular artifacts in the eeg without an eog reference channel," in *Proceedings of the 7th Nordic Signal Processing Symposium-NORSIG 2006*. IEEE, 2006, pp. 130–133.

[41] S. Makeig, "Auditory event-related dynamics of the eeg spectrum and effects of exposure to tones," *Electroencephalography and clinical neurophysiology*, vol. 86, no. 4, pp. 283–293, 1993.

[42] G. Pfurtscheller and F. L. Da Silva, "Event-related eeg/meg synchronization and desynchronization: basic principles," *Clinical neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.

[43] J. C. Shaw, "Correlation and coherence analysis of the eeg: a selective tutorial review," *International Journal of Psychophysiology*, vol. 1, no. 3, pp. 255–266, 1984.

[44] P. Welch, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on audio and electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[45] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, R. Boots, and B. Benatallah, "Cascade and parallel convolutional recurrent neural networks on eeg-based intention recognition for brain computer interface." in *AAAI*, 2018.

[46] N. C. Oza and S. Russell, *Online ensemble learning*. University of California, Berkeley, 2001.

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

[48] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[49] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[50] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (fbcsp) in brain-computer interface," in *2008 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2008, pp. 2390–2397.

[51] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen *et al.*, "Meg and eeg data analysis with mne-python," *Frontiers in neuroscience*, vol. 7, p. 267, 2013.

[52] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[53] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 11, pp. 5619–5629, 2018.

[54] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[55] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[56] A. R. Luft, G. V. Smith, L. Forrester, J. Whitall, R. F. Macko, T.-K. Hauser, A. P. Goldberg, and D. F. Hanley, "Comparing brain activation associated with isolated upper and lower limb movement across corresponding joints," *Human brain mapping*, vol. 17, no. 2, pp. 131–140, 2002.

[57] G. Pfurtscheller, C. Guger, G. Müller, G. Krausz, and C. Neuper, "Brain oscillations control hand orthosis in a tetraplegic," *Neuroscience letters*, vol. 292, no. 3, pp. 211–214, 2000.