Received February 20, 2019, accepted March 9, 2019, date of publication March 26, 2019, date of current version April 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2907644

Robustness Analysis of Identification Using Resting-State EEG Signals

YANG DI[1], XINGWEI AN[2], FENG HE[1,2], SHUANG LIU[2], YUFENG KE[2], AND DONG MING[1,2], (Member, IEEE)

[1]Laboratory of Neural Engineering and Rehabilitation, Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin 300072, China

[2]Tianjin International Joint Research Center for Neural Engineering, Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin 300072, China

Corresponding authors: Feng He (heaven@tju.edu.cn) and Dong Ming (richardming@tju.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1300302, and in part by the National Natural Science Foundation of China under Grant 81630051 and Grant 61603269.

ABSTRACT The brain activity pattern can be presented by Electroencephalogram (EEG), which is considered as an alternative to traditional biometrics. Researchers have done conducted studies on EEG-based identification, while few of them discussed the effect of time robustness which is very important for the identification system. In this study, we compared and analyzed the two runs EEG signals of resting-state of eye open/closed (REO/REC). The time intervals between two runs were at least two weeks. Here are 17 participants joined in this study. Each of them took two runs experiment. Each run contains four sessions, each session includes 150 seconds of REO/REC. Spectral and statistical analyses were used to extract feature. Three classifiers, Euclidean distance, SVM, and LDA, were used to get classification accuracies and to compare the performance between features of each run and two runs. The results of two runs PSD values of both REO and REC conditions show that there is a similarity within each subject and a difference between subjects. The classification accuracies of three methods of each run are almost 99%. The classification accuracies using two runs data as training set can also reach up to 97% while using each of two-run data as training set is nearly 80%. Thus, the features of most subjects have cross-time robustness and could be used as identification. This study will have an important role in EEG-based identification system.

INDEX TERMS Electroencephalography(EEG), identification, resting-state, robustness.

I. INTRODUCTION

Personal identification has a great impact on the security of personal information. It also affects social public safety, national security and many other aspects. Traditional identification ways using accounts and passwords could not guarantee the information security, since the password is too easy to leak today. To some extent, biometrics identification can compensate for this disadvantage. It uses inherent complicit features of users and is considered as an alternative identification way [1]. Widely used reliable systems based on biometric modalities such as fingerprint, gait, voice and iris, which are uniqueness, ubiquitous, easy acquisition, and persistence, are crucial to information security nowadays [9].

The associate editor coordinating the review of this manuscript and approving it for publication was Navanietha Krishnaraj Krishnaraj Rathinam.

Electroencephalography (EEG), which is used to record the electrical activity of the brain, has received much attention and interest in many areas. It is also regarded as an alternative promising way for biometrics identification. Comparing with other biometric modalities, EEG signals cannot be reproduced and imitated by imposters especially when people are under hijacked or other dangerous situations. This makes EEG-based biometrics identification more reliable and securer. Currently, resting-state spontaneous EEGs, including resting of eye open (REO) and eye closed (REC), are most used EEG signals for personal identification [2], [13], [17] [19].

EEG-based identification methods employ many approaches to extract temporal-spatial features and optimize classifiers to improve the performance of biometrics identification system [3], [15], [16], [20]. EEG-based identification systems have been investigated for decades [27], [31], [32].

First study focusing on the difference of EEG morphology of monozygotic can be traced back to 1936, and the results of latter study had shown similarity on spectral features of EEG signals between twins.

Spectral features are mostly used in EEG-based identification. In 1999, Poulos *et al.* first tried to conduct a system of identity authentication based on EEG signals [20], [21]. They used four different subjects' EEG data and employed a Learning Vector Quantizer networks as the classifier. AR parameters of the alpha rhythm activity were estimated from EEG signals. The classification accuracy can reach 84%, which shows EEG signals has genetic information and different subjects have their own spectral characteristics.

Ward and Obeid [29] reviewed various biometrics identification methods over past few years and listed some challenges and future perspectives. Although biometrics identification seems to be an alternative way to traditional methods, some problems are also existing, such as data acquisition, protocol design, performance evaluation and stability of system. In this review, the author also emphasized the importance of time-robustness for the performance of biometrics identification system.

A stable identification system must have the ability that it can apply the template generated by previous dataset to future datasets. The results in previous works, however, showed that the robustness of EEG signals will get worse with time goes by, and few literatures conducted longitudinal examinations to investigate the stability of system performance. Npflin, M. *et al.* conducted two runs experiments and extracted the features of PSD, peak height and peak frequency. Resting-state EEG data of 20 subjects with eye closed are used, and the performance of identification system have a significant decreased. Jae-Hwan Kang, J. H.*et al.* used public dataset which includes EEG signals of 109 subjects and employed spectral analysis, nonlinear analysis and network analysis to extract various EEG features [8]. The results show there is a significant decreased between resting-state of different subjects.

In this paper, we aim to investigate the time robustness of EEG-based identification using EEG signals of resting of eyes open (REO)/eyes closed (REC) as biometric features. We acquired signals of two independent runs. The interval of two runs was at least two weeks. We try to analysis the time robustness through comparing the classification performances using different parts of the whole dataset. Here, we extract different spectral EEG features as the inherent biological characteristic. We also implement visualization and statistical analysis for spectral analysis. Three classification methods, including Euclidean distance, support vector machines (SVM) [10] and linear discriminant analysis (LDA), are adopted to evaluate the identification performance of spectral features.

Most of previous works focused on REC and single classifier [11], [22]–[24], here we focus on cross-time robustness for the inherent subject-to-subject features. Thus, besides the analysis of spectral features with single-run EEG signals, we mainly compared and analyzed the performance of mixed spectral feature with two-run EEG data. We also try to compare the difference of spectral features between runs. In the present study, the results reveal that there has a significant difference between single run EEG signal of different subjects and strong similarity between two runs EEG signals of most subjects, and the result also shows several subjects have dissimilarity between two runs EEG signals.

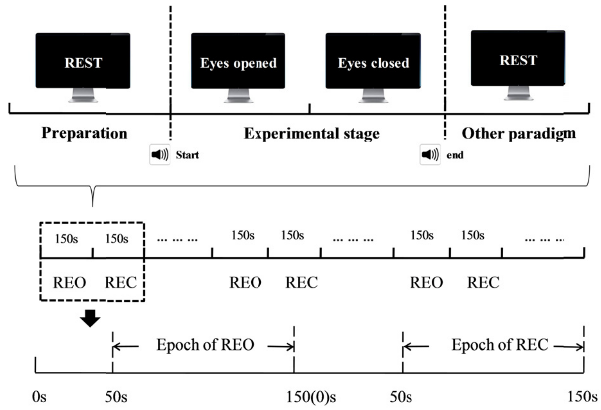


FIGURE 1. Each run of the experiment includes three sessions and each session consists of two sub sessions, which are resting-state of eye open(REO) and resting-state of eye closed (REC).

II. METHODS AND MATERIALS

A. PARTICIPANTS

There are 17 subjects (8 males) participated in this experiment with average age of 21(±3). They are all volunteers from Tianjin University and have normal vision (or corrective vision). All participates have signed the consent form which includes notice in experiment, using of data and their individual rights before conducting the experiment. The whole experiment for each participant include two runs and the time interval of two runs experiment was at least two weeks.

B. PARADIGM

Fig 1 shows the experimental setup. Each run of the experiment consists of three sessions. Subjects can relax at the interval of two sessions and press 'space' to start next session of the experiment at any time. For each session, there are three subsessions within each session which are preparation stage, experimental stage and relaxation stage. In first stage of preparation, subjects sit in front of the screen and wait for audio hint for begin. The second stage which is named as 'Experimental stage' includes two conditions: resting-state of eyes open (REO) and resting-state of eyes closed (REC). 'Experiment stage' begins and ends with audio hint. Each condition of this stage lasts 150 seconds.

C. DATA ACQUISITION

EEG signals was acquired using an EEG Cap with 64 Ag/AgCl electrodes placed at the standard positions of the international 10-20 system. The top of the head is set as the reference and 'AFz' channel is set as ground. A 64-channels amplifier (Neuroscan) were used here for data acquisition with the sample rate of 1000 Hz. Only 20 channels (Fz, F3/4, F7/8, Cz, C3/4, Pz, P3/4, PO7/8, TP7/8, Oz, O1/2, M1/2) were recorded for data analysis. The study was approved by the local ethical committee at Tianjin University.

D. PREPROCESS

All preprocess of EEG analysis was conducted using MATLAB R2013b with the EEGLAB toolbox (v13.4.4.4bversion). The EEG data were firstly re-referenced to the average of left and right mastoids ('M1' and 'M2') first. Then, we filtered the EEG data with a pass filter of 1-40Hz to increase the signal-to-noise ratio, and were down-sampled from 1000Hz to 100Hz. For both REO and REC, we extracted the EEG data behind from the label of 50 seconds to 150 seconds according to the audio hint for each session of both REO and REC as an epoch. The EEG data from the label of 0 seconds to 50 seconds, which is not used as features, were used as baselines. Each epoch was then divided into 10 segments (a fragment of 10seconds) in order to increase the EEG sample amount of EEG data of same each subjects and this is helpful for the next analysis and classification The data size of each subject after preprocess is C×T×S, where C means channels (C = 18) and S denotes all trials of one experiment (S = 30). T denotes the length of a single trial (T = 1000).

III. FEATURE AND CLASSIFICATION

A. SPECTRAL ANALYSIS

We introduced the method of power spectral density (PSD) to extract the spectral feature. Power spectral density(PSD) is a typical method to describe the distribution of power in EEG signal analysis [6]. The PSD of EEG signals is estimated in each epoch for each channel by performing a fast Fourier transform (FFT) with a Hamming window.

We estimated the average PSD of all channels with following bands: 1. theta band (4-8Hz), 2. alpha band (8-13Hz), 3.low beta band (13-20Hz), 4. high beta band (20-30Hz), 5. gamma band (30-40Hz), 6.all frequency bands (1-40Hz), and we used these features for further analysis and classification [7], [12], [14].

Spectral estimation is to estimate the distribution of power over frequency based on a finite set. the autocorrelation function is the average measure of the characteristics of the signal in the time domain. It is used to describe the correlation between the values of the EEG random signal X(t) at any two different times. We define autocorrelation function $\mathbf{R_x}(\boldsymbol{\tau})$ as:

$$R_x(\tau) = E\left[X(t)X(t+\tau)\right] \tag{1}$$

If the $\boldsymbol{\tau} = \mathbf{0}$, $\mathbf{R_x}(\mathbf{0})$ represent the mean square value of $\mathbf{X(t)}$:

$$R_x(0) = \lim \frac{1}{T}\int_0^T X^2(t)\,dt \tag{2}$$

With stochastic process, frequency range often be defined by Fourier transform. For time series, whether the Fourier transform exists depends on whether it is absolutely integrable. Constraint condition is defined as follows:

$$\int_{-\infty}^{\infty} |X(t)|\,dt < \infty \tag{3}$$

For stationary stochastic process, autocorrelation function meets this formula. So we can get the FFT of autocorrelation function, $\mathbf{S_x}(\boldsymbol{\omega})$ represent the change of signal of power with itself frequency. It can be defined as:

$$S_x(\omega) = \lim_{T\to\infty}\left\{E\left[\frac{1}{T}\left|X_T(w)\right|^2\right]\right\} \tag{4}$$

For estimation of the spectrum, the periodgram is the easiest method. This method is to do the Discrete Fourier transform (DFT) of stochastic process sampling and calculate square of amplitude. Welch's method is an improvement to periodgram. It divides the data into some segments which can overlap, then estimate each segment and average. Here, we set parameters overlap 50% and hamming window.

B. CLASSIFICATION

For personal identification, we compared three commonly used classification methods, which are Euclidean distance, SVM and LDA. Euclidean distance and LDA are basic and common methods in personal identification of resting-state EEG signals. SVM proved to be useful in small-sample in some fields.

1) EUCLIDEAN DISTANCE

The Euclidean Distance between two EEG trials $\mathbf{X_1}(\mathbf{t}) = \left\{\mathbf{x_1^1}, \mathbf{x_1^2}, \mathbf{x_1^3}, \cdots, \mathbf{x_1^n}\right\}$ and $\mathbf{X_2}(\mathbf{t}) = \left\{\mathbf{x_2^1}, \mathbf{x_2^2}, \mathbf{x_2^3}, \cdots, \mathbf{x_2^n}\right\}\}$ are defined as :

$$d\left(X_1(t), X_2(t)\right) = \sqrt{\sum\nolimits_{i=1}^{n}\left(x_1^i - x_2^i\right)^2} \tag{5}$$

where $\mathbf{d}\left(\mathbf{X_1}(\mathbf{t}), \mathbf{X_2}(\mathbf{t})\right)$ in (5) is Euclidean distance of signal $\mathbf{X_1}(\mathbf{t})$ and $\mathbf{X_2}(\mathbf{t})$, $\mathbf{i}$ and $\mathbf{j}$ represent time and trial in $\mathbf{x_j^i}$ respectively.

2) LDA

Linear discriminant analysis (LDA) is a generalization of Fisher's linear discriminant. LDA approaches the problem by assuming that the conditional probability density functions $\mathbf{P}\left(\vec{\mathbf{x}}\middle|\,\mathbf{y}\right) = \mathbf{0}$ and $\mathbf{P}\left(\vec{\mathbf{x}}\middle|\,\mathbf{y}\right) = \mathbf{1}$ are normally distributed with mean and covariance parameter$\left(\vec{\boldsymbol{\mu}}_\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{0}\right)$ and $\left(\vec{\boldsymbol{\mu}}_\mathbf{1}, \boldsymbol{\Sigma}_\mathbf{1}\right)$. Consider the observations $\vec{x}$ for each sample of an event with known class $\mathbf{y}$,

$$\Sigma_b = \frac{1}{c}\sum\nolimits_{i=1}^{c}\left(\mu_i - \mu\right)\left(\mu_i - \mu\right)^T \tag{6}$$

where $\boldsymbol{\mu}$ is the mean of the class means and $\mathbf{i}$ is the label of classes. The $\vec{\boldsymbol{\omega}}$ in this case will be calculated by:

$$S = \frac{\vec{\omega}^T\,\Sigma_B\vec{\omega}}{\vec{\omega}^T\,\Sigma\vec{\omega}} \tag{7}$$

where $\boldsymbol{\Sigma}$ mean the covariance of each class and $\vec{\boldsymbol{\omega}}$ is an eigenvector of $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\mathbf{b}$ to corresponding eigenvalue.

3) SVM

Support vector machines (SVM) are supervised learning models in machine learning with associated learning algorithms which analyze data used for classification [4], [5]. Given training data $\mathbf{x_i} \in \mathbf{R^m}, \mathbf{i} = \mathbf{1}, \ldots\ldots, \mathbf{l}$, and $\mathbf{y} \in \mathbf{R^l}$ as an indicator vector such that $\mathbf{y_i} \in \{\mathbf{1}, \mathbf{-1}\}$, solves the following formulation:

$$\begin{aligned} &\min_{\omega,b,\xi} \frac{1}{2}\omega^T\omega + C\sum\nolimits_{i=1}^{l}\xi_i \\ &\text{subject to } y_i\left(\omega^T\emptyset\left(x_i\right)+b\right) \geq 1-\xi_i, \\ &\qquad \xi_i \geq 0, i = 1, \ldots\ldots, l, \end{aligned} \tag{8}$$

where $\emptyset\left(\boldsymbol{x}_\mathbf{i}\right)$ maps $\boldsymbol{x}_\mathbf{i}$ into a higher-dimensional space and $\mathbf{C} > \mathbf{0}$is the regularization parameter. Due to the possible high dimensionality of the vector variable $\boldsymbol{\omega}$, we transform it to following dual problem:

$$\begin{aligned} &\min_{\alpha} \frac{1}{2}\alpha^T Q\alpha - e^T\alpha \\ &\text{subject to } y^T\alpha = 0, \\ &\qquad 0 \leq \alpha_i \leq C, i = 1, \ldots\ldots l, \end{aligned} \tag{9}$$

where $\boldsymbol{e} = [\mathbf{1}, \ldots\ldots, \mathbf{1}]^{\boldsymbol{T}}$ and $\mathbf{Q}_{\boldsymbol{ij}} \equiv \mathbf{y}_{\boldsymbol{i}}\mathbf{y_j}\boldsymbol{K}\left(\boldsymbol{x_i}, \boldsymbol{x}_\mathbf{j}\right)$, and $\boldsymbol{K}\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right) \equiv \boldsymbol{\emptyset}(\boldsymbol{x}_\mathbf{i}^{\boldsymbol{T}})\emptyset(\boldsymbol{x_j})$ is the kernel function. Using primal-dual relationship, the $\boldsymbol{\omega}$ satisfies:

$$\omega = \sum\nolimits_{i=1}^{l} y_i\alpha_i\emptyset(x_i) \tag{10}$$

and the decision function is:

$$sgn\left(\omega^T\emptyset\left(x_i\right)+b\right) = sgn(\sum\nolimits_{i=1}^{l} y_i\alpha_i K\left(x_i, x\right)+b) \tag{11}$$

LIBSVM were used in MTLAB to solve the issues, such as SVM optimization problems, theoretical convergence, multi-class classification and so on, for subsequent analysis. LIBSVM is a library for support vector machines by Chang and Lin [34].

4) ACCURACY

For comparison of classification performance, the accuracy formulation was used as follows:

$$\text{Accuracy} = \frac{\#\ correctly\ predicted\ index}{\#\ total\ testing\ index} \times 100\% \tag{12}$$

IV. RESULT

In this section, the experimental results are presented using methods described above. 18-channel PSD values of each single trial and the Pearson's correlation coefficient among trials are presented to show the spectral features within and between subjects. Classification accuracies are also presented here to show the results with different classifier. In the following, we will firstly present the mentioned results of both conditions (REO and REC) using only first-run data, and then we will show the results of both runs. At the end of this section, we will present the cross-time accuracy of both conditions, where we use the first-run experimental dataset as training dataset and the second-run experimental dataset as testing dataset.

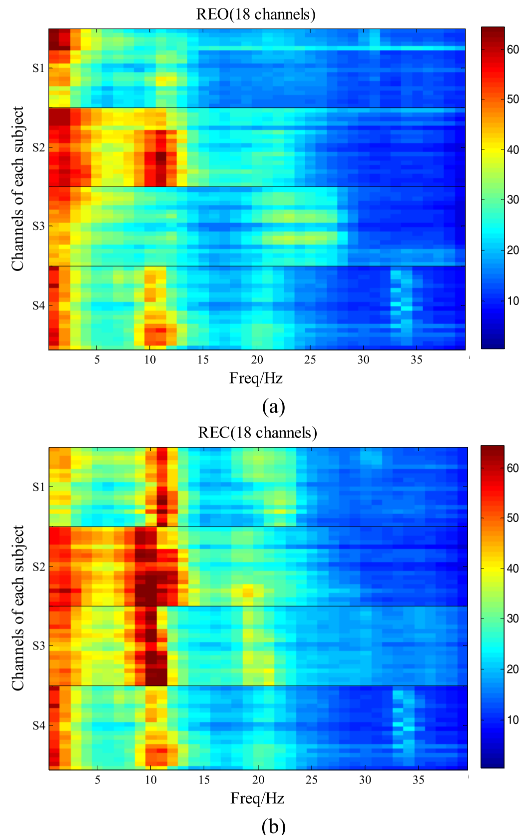


FIGURE 2. (a) 18-channel PSD of four subjects under condition REO. (b) 18-channel PSD of four subjects under condition REC.

A. SINGLE RUN DATASET

Figure 2 shows the PSD values of 18 channels for four subjects at 1-40Hz (left: resting-state of eyes open (REO), right: resting-state of eyes closed (REC)). The average values of 30 trials of each channel were calculated and arranged into a matrix. The X-axis of the figure indicates frequency from 0 Hz to 50 Hz, and Y-axis indicates 18 channels of each subject, which means that each line represents the averaged PSD values of one channel (channels' order: F7, F3, FZ, F4, F8, C3, CZ, C4, TP7, TP8, P3, PZ, P4, PO7, PO8, O1, OZ, O2).

As Fig 2 shows, there are significantly high energy between 1-5 Hz bands and around 10Hz for both REO and REC conditions. The frequency band at around 10Hz of REC condition is higher than that of REO condition and the frequency band at around 20Hz also shows a difference. These frequency bands are also commonly used frequency bands for individual classification of resting EEG.

Figure 3 depicts the trial-to-trial Pearson's correlation coefficient of same four subjects both for REO and REC conditions. Firstly, the PSD values of channel-averaged signal were calculated for each single trial to extract the trial-to-trial Pearson's correlation coefficient. Here we show the correlation results of four subjects. Each subject has thirty trials. It shows a good correlation within subject, which was significantly higher than between subjects.

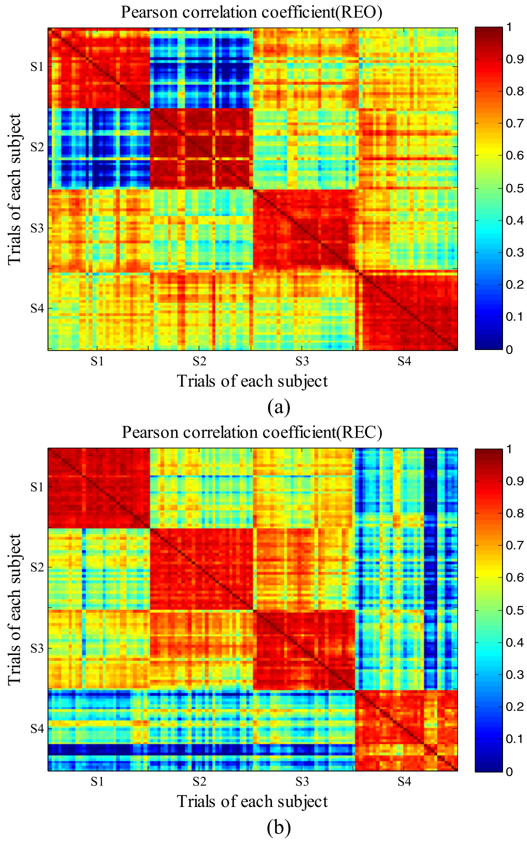


FIGURE 3. (a) Pearson correlation coefficient of REO (b) Pearson correlation coefficient of REC.

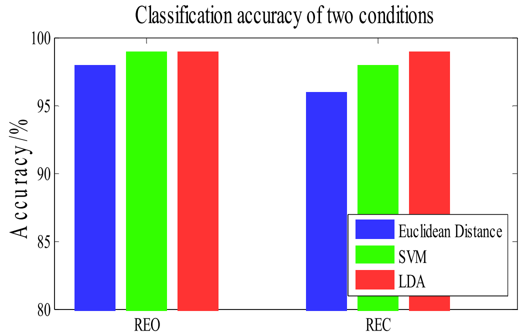


FIGURE 4. Classification accuracy of two conditions (REO and REC) by three methods (ED, SVM and LDA). Training set and test set are from single run experiment.

The training data and testing data were all extracted from the first-run experiment, and 10-fold cross-validation was used to classify. We used average PSD values of 18-channel for each trial as the feature. Figure 4 shows the classification accuracy using three different methods (Euclidean Distance, LDA and SVM) of two conditions (REO and REC). The three bars on the left shows the results in condition REO, and the right part shows the accuracies in condition REC. Different color indicates different classification method. Classification accuracy in each condition can all reach more than 95%. REO condition has higher classification accuracies than REC condition. SVM and LDA have better classification performance than ED.

B. TWO-RUN DATASET

For personal identification, time-robustness is very important. Thus we recorded from two runs experiment. The interval of two runs was at least two weeks. Up analysis are all results of the first run experiment. From this subsection, we will also analyze the second-run data. We will first show the spectral features of each subject in each run, and then we will rebuild the training and testing datasets using mixed datasets from two runs.

Similar to the previous description, the averaged 18-channel PSD values of each subjects were selected as features as shown in Figure 5. It depicts 18-channel PSD values of four subjects over time in different conditions. The up two rows show the PSD values in resting of eyes open (REO) condition and eyes closed (REC) condition among first-run and second-run experiment. For the resting of eyes opened, the energies at 5 Hz and 10 Hz are relatively high than other frequencies both for first and second runs. For the resting of eyes closed, the energy of about 5Hz in two cross-time experiments is lower than that of eyes opened. The energy at 10Hz is obviously higher than that of eyes opened, and moreover, high energy appears at around 20Hz.

Figure 6 shows Pearson's correlation coefficient of PSD values between first-run and second-run experiments. The left subfigure shows the results for condition REO, and the right part represent the coefficient of condition REC. Each line represents the Pearson's correlation coefficient between subjects using the averaged PSD values of 18 channels as the feature. X-axis represents the features extracted from the first-run dataset for each subject, and Y-axis represents the features extracted from the second-run dataset. Coefficient values under 0.8 will be presented the same color as 0.8. Different color represents different coefficient value. As we expected, the correlation coefficient of same subject is much higher than it of different subjects under two conditions. The results show that there are significant similarities between two runs dataset of same subject.

Figure 7 depicts the classification accuracy of three methods using two runs dataset. We use 10-fold cross-validation and mix two experimental data for training and testing. We calculate the average PSD values of 18 channels for each subject and 30 trials were used for each subject finally. Three classification methods all here achieved more than 90% in both two conditions (REO and REC). Considering that the training and test sets contain data from the first and the second runs, this result is not surprising. To a certain extent, this indicates that the PSD features of the resting state have cross-time robustness, and their features are also well tested for different classifiers The results obtained by using the three classifiers show that the PSD values also has certain stability for different classifiers.

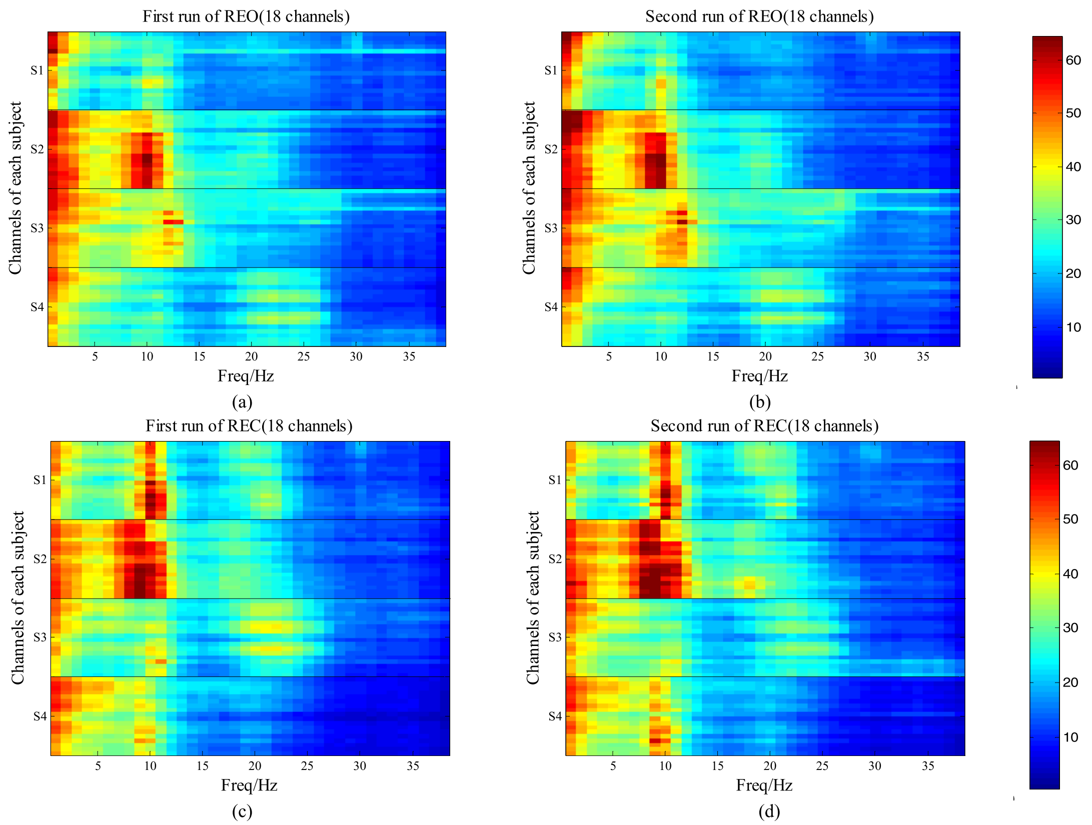


FIGURE 5. 18-channel PSD of four subjects of two runs data. X-axis represent frequency (Hz) at 1-40Hz. (a) 18-channel PSD of REO in first run. (b) 18-channel PSD of REC in second. (c) 18-channel PSD of REC in first run. (d) 18-channel PSD of REC in second run.

C. TWO RUNS DATASET

Here we try to use first-run dataset to train and second-run dataset for test to investigate whether there is robustness across time between two independent runs.

Figure 8 depicts the classification accuracy for cross-time data. We use first-run data as a training set and second-run as a test set. Then exchange them, the second-time data as a training set, and the first-time data as a test data. We still use PSD values as features to classify and analysis. As can be seen from the figure, for the full-band power spectral density of 1-40Hz, the classification accuracy obtained by three methods are poor, only 30%-40%. Then we made a selection of the frequency domain range, characteristics of five bands (theta, alpha, low beta, high beta, gamma) are used as features to classify and made combination of these frequency domain bands. The result can reach up to 80%.

V. DISCUSSION

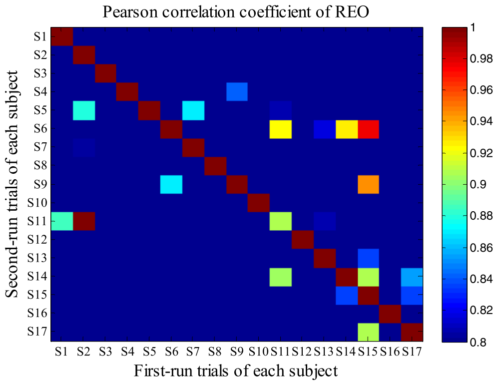
In this paper, we conduct the analysis of spectral features with single-run EEG signals and spectral features between runs. We also compare the classification accuracies using different classifiers with different features.

A. PSD FEATURE

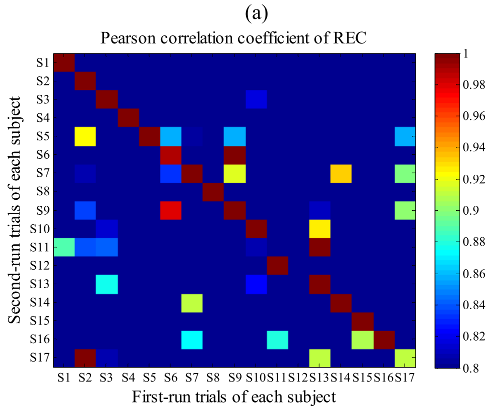
For single-run PSD of different subjects, there is a difference during the frequency range of 1-40Hz. As we can see from Fig 2, the PSD of same subject has a similarity between 18 channels and there is an obvious boundary between different subjects. The difference mainly concentrated around 5Hz and 10Hz of REO. Some subjects have high energy around at 5Hz, 10Hz or both of them. The PSD of forehead channels of most subjects at 1-5Hz is significantly higher than other channels. Compared with REO, REC also has a significant difference at 20Hz. The PSD at 1-5 Hz of most subjects is lower than it under condition REO.

Most cross-time analysis researches limited their analysis with single-experimental resting-state data of public databases [25], [26]. In this article, we conducted two experiments on some subjects for analysis and the interval of two experiments at least two weeks. Our ultimate goal is to find an independent identification method for resting state. That is, we can get a good performance by using one independent run data for training and another independent run data for testing. To achieve this goal, we calculated PSD values of two runs data of two conditions and the result shows that there seems be a significant similarity between two runs. The correlation of PSD between the two runs data which showed in Fig 5 is mainly concentrated at 1-10Hz for condition REO and also at 20Hz for condition REC.

Noted that subject four has a high PSD at 20 Hz of condition REO between two runs data, which is different from PSD of other subjects in Fig 5. The PSD of subject one and



(a)



(b)

FIGURE 6. (a) Pearson correlation coefficient of REO between two runs data. (b) Pearson correlation coefficient of REC between two runs data.

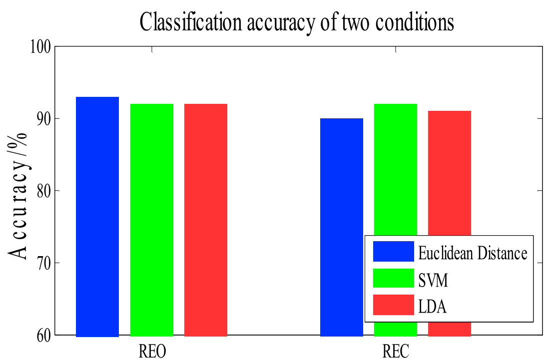


FIGURE 7. Classification accuracy of two condition (REO and REC). Two-run dataset was used here. We mixed them and used part of mixed dataset as training set or test set.

four are much lower than other two subjects at around 10Hz for all channels. For condition REC, the PSD of subject four at 20Hz is significantly lower than other subjects. this is a little different from the results we got in previous section. The results reveal that the PSD of all channels have a similarity within the same subject and difference with different subjects.

B. PEARSON CORRELATION COEFFICIENT

For single-run Pearson correlation coefficient of each subject, the red zone of diagonal means the coefficient much close to '1', which represent strong correlation between

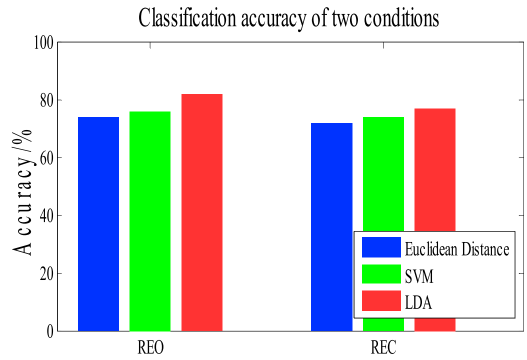


FIGURE 8. Classification accuracy of two condition (REO and REC). Two runs data were used to classify. One run was used as training set and another was used as test set.

trials within subject. For condition REO or REC, the result in Fig. 3 shows that there is a strong correlation between trials of average channels within subject and lower correlations with other subjects. This means that we can use the average-channel PSD of each subject as features for the following analysis and classification.

Next, we calculated the Pearson correlation coefficient of two runs data. The PSD of average channels are used as features and each subject of each run has 30 trials. The results in Fig 6 shows a certain correlation between two runs data for most subjects. The result is better than what we expected before. For condition REO, only first run data of subject 11 shows lower correlation with second run data. But it also has a high correlation coefficient (more than 0.8). For condition REC, the result is not as good as condition REO, the correlation coefficients of four subjects are not the highest on the diagonal and the coefficients of two subjects (subject 11 and subject 12) even below 0.8.

C. CLASSIFICATION RESULTS

Three methods were used here to classify. As we expected, for classification result of single run dataset, all three methods get good results which can reach more than 95%. These results are also consistent with the results which mentioned in some researches. It also reveals that there is a significant difference between single run data of different subjects.

Fig 7 shows the classification accuracy of two runs dataset of different subjects. Here, we use 10-fold cross-validation. Two runs dataset are used as training data and test data. which include data from both the first run and the second run. The classification accuracy rate can reach more than 90%, considering that training set and test set both consist of first and second experimental data. To some extent, the classification results are accord with our expectations that there is indeed a certain time robustness between experimental data from different times for each subject.

The PSD features at different times, however, seems to have some differences. The result in Figure 8 shows the best classification accuracies that we use one run dataset as training set and another as test set.

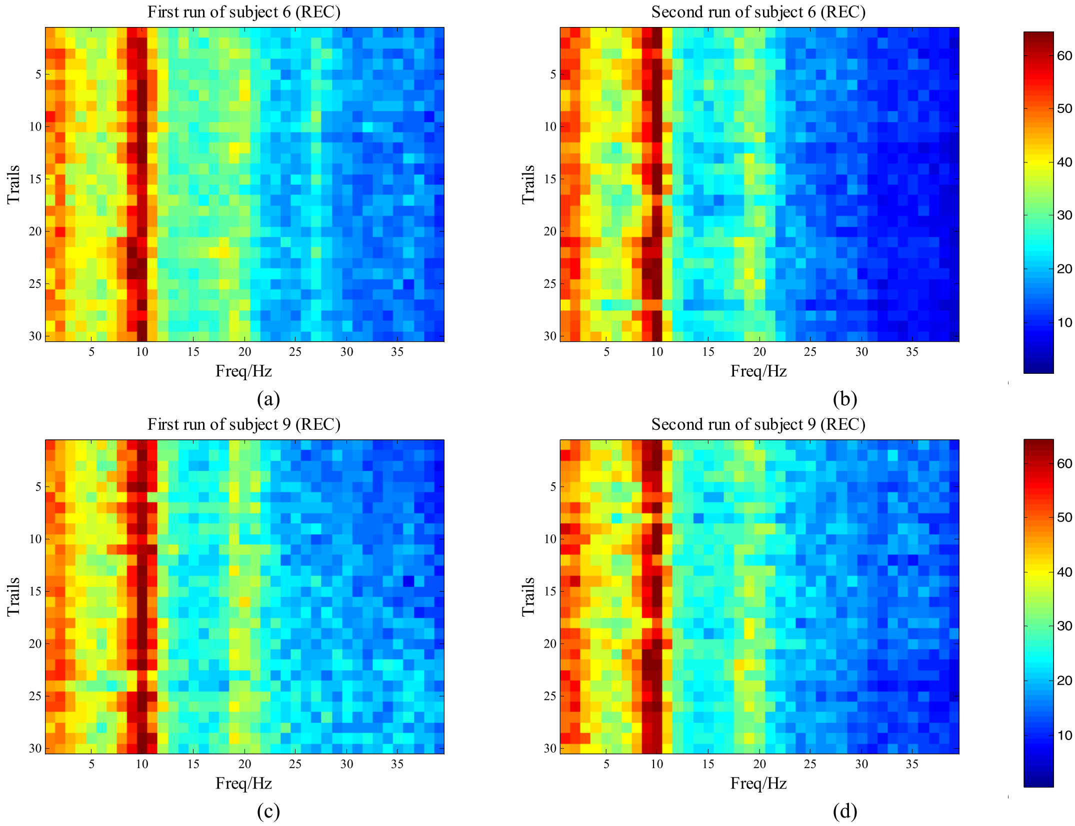


FIGURE 9. **Average-channel PSD of two misclassified subjects (6 and 9). X-axis is frequency (Hz) at 1-40Hz. (a) (b) Two runs data (30 trials) of subject 6. (b) (d) Two runs data (30 trials) of subject 9.**

We first use PSD of 1-40Hz as features to classify and the result shows poor accuracy just about 50% This seems not a good result. Then we try to optimize the features by using the various bands and their combinations. The goal is to find a frequency band that can better represent the common features between the same subject while increasing the gap between the different subjects. The classification result can reach 80% when using PSD of range 5-35Hz as features.

In order to investigate the reasons why we get the lower classification accuracy in two runs dataset, we have a simple analysis of the classification results. For most of the subjects, most of their trials were correctly classified while some other subjects, almost lost all trials. Here, we choose two subjects which are all misclassified in classification. Figure 9 shows PSD values of two subjects (subject six and subject nine) under condition REC. As we can see, two runs PSD values of two subjects have a significant similarity. They all have a high PSD value at around 1-10Hz and 20Hz and the distribution of frequency energy is very similar. We can also prove this results in Fig 6 that There is a significant similarity between two runs data of two subjects (S6 and S9). In the future work, we will focus on the analysis of these subjects, experiments will be conducted on the same subject to determine whether it is a problem with the signals itself and other features of the signals will also be analyzed to achieve identification.

VI. CONCLUSION

In this paper, we focus on the cross-time robustness of REO and REC between two runs experiments. The Power Spectral Density was used here to extract features of dataset. We used Pearson correlation coefficient to weight the similarity between trials. Three methods (ED, SVM and LDA) were used to analysis and classification. The results show that for single run dataset, there is a significant difference between PSD of all subjects under condition REC and REO. So, we conducted the second run experiment on a number of 17 subjects and compared the similarity between two runs data of same subject. The results show that there is an obvious similarity for most subjects of REO and REC. The results indicate that there is a cross-time robustness of REO and REC between same subject and the identification system that using one run data as a template while using others to verification independently is feasible.

ACKNOWLEDGMENT

(Yang Di and Xingwei An contribute equally to this work.)

REFERENCES

[1] B. C. Armstrong, M. V. Ruiz-Blondet, N. Khalifian, K. J. Kurtz, Z. Jin, and S. Laszlo, "Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics," *Neurocomputing*, vol. 166, pp. 59–67, Oct. 2015.

[2] R. J. Barry, A. R. Clarke, S. J. Johnstone, and C. R. Brown, ''EEG differences in children between eyes-closed and eyes-open resting conditions,'' *Clin. Neurophysiol.*, vol. 120, no. 10, pp. 1806–1811, 2009.

[3] J. Berkhout and D. O. Walter, ''Temporal stability and individual differences in the human EEG: An analysis of variance of spectral values,'' *IEEE Trans. Bio-Med. Eng.*, vol. BME-15, no. 3, pp. 165–168, Jul. 1968.

[4] C. M. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). New York, NY, USA: Springer-Verlag, 2006.

[5] C.-J. Lin, ''Errata to 'A comparison of methods for multiclass support vector machines,''' *IEEE Trans. Neural Netw.*, vol. 13, no. 4, pp. 1026–1027, Jul. 2002.

[6] F. H. Duffy and H. Als, ''A stable pattern of eeg spectral coherence distinguishes children with autism from neuro-typical controls—A large case control study,'' *BMC Med.*, vol. 10, no. 1, p. 64, 2012.

[7] T. Gasser, P. Bächer, and J. Möcks, ''Transformations towards the normal distribution of broad band spectral parameters of the EEG transformation en distribution normale des paramétres spectraux d'une bande large de l'EEG,'' *Electroencephalography Clin. Neurophysiol.*, vol. 53, no. 1, pp. 119–124, 1982.

[8] J.-H. Kang, Y. C. Jo, and S.-P. Kim, ''Electroencephalographic feature evaluation for improving personal authentication performance,'' *Neurocomputing*, vol. 287, pp. 93–101, Apr. 2018.

[9] J. Klonovs, C. Petersen, H. Olesen, and A. Hammershoj, ''ID proof on the go: Development of a mobile EEG-based biometric authentication system,'' *IEEE Veh. Technol. Mag.*, vol. 8, no. 1, pp. 81–89, Mar. 2013.

[10] M. R. N. Kousarrizi, A. A. Ghanbari, M. Teshnehlab, M. A. Shoorehdeli, and A. Gharaviri, ''Feature extraction and classification of EEG signals using wavelet transform, SVM and artificial neural networks for brain computer interfaces,'' in *Proc. Int. Joint Conf. Bioinformatics, Syst. Biol. Intell. Comput. (IJCBS)*, Shanghai, China, Aug. 2009, pp. 352–355.

[11] L. Ma, J. W. Minett, T. Blu, and W. S-Y. Wang, ''Resting state EEG-based biometrics for individual identification using convolutional neural networks,'' in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2015, pp. 2848–2851.

[12] C. Lee, J.-H. Kang, and S.-P. Kim, ''Feature slection using mutual information for EEG-based biometrics,'' in *Proc. Int. Conf. Telecommun. Signal Process.*, Jun. 2016, pp. 673–676.

[13] M. Näpflin, M. Wildi, and J. Sarnthein, ''Test–retest reliability of resting EEG spectra validates a statistical signature of persons,'' *Clin. Neurophysiol.*, vol. 118, no. 11, pp. 2519–2524, 2007.

[14] D. J. Mcfarland, L. M. Mccane, S. V. David, and J. R. Wolpaw, ''Spatial filter selection for EEG-based communication,'' *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 3, pp. 386–394, 1997.

[15] C. Miyamoto, S. Baba, and I. Nakanishi, ''Biometric person authentication using new spectral features of electroencephalogram (EEG),'' in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Feb. 2009, pp. 1–4.

[16] R.-R. Nir, A. Sinai, E. Raz, E. Sprecher, and D. Yarnitsky, ''Pain assessment by continuous EEG: Association between subjective perception of tonic pain and peak frequency of alpha oscillations during stimulation and at rest,'' *Brain Res.*, vol. 1344, pp. 77–86, Jul. 2010.

[17] R. Palaniappan and D. P. Mandic, ''Biometrics from brain electrical activity: A machine learning approach,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 738–742, Apr. 2007.

[18] R. Palaniappan, ''Two-stage biometric authentication method using thought activity brain waves,'' *Int. J. Neural Syst.*, vol. 18, no. 1, pp. 59–66, 2008.

[19] R. B. Paranjape, J. Mahovsky, L. Benedicenti, and Z. Koles, ''The electroencephalogram as a biometric,'' in *Proc. Can. Conf. Elect. Comput. Eng.*, May 2001, pp. 1363–1366.

[20] M. Poulos, M. Rangoussi, N. Alexandris, and A. Evangelou, ''Person identification from the EEG using nonlinear signal classification,'' *Methods Inf. Med.*, vol. 41, no. 1, pp. 64–75, 2002.

[21] M. Poulos, M. Rangoussi, V. Chrissikopoulos, and A. Evangelou, ''Parametric person identification from the EEG using computational geometry,'' in *Proc. IEEE Int. Conf. Electron. Circuits Syst.*, Sep. 1999, pp. 1005–1008.

[22] K. V. R. Ravi and R. Palaniappan, ''Leave-one-out authentication of persons using 40 Hz EEG oscillations,'' in *Proc. Int. Conf. Comput. Tool.*, Nov. 2005, pp. 1386–1389.

[23] D. La Rocca, P. Campisi, and G. Scarano, ''EEG biometrics for individual recognition in resting state with closed eyes,'' in *Proc. Int. Conf. Biometrics Special Interest Group*, Sep. 2012, pp. 1–12.

[24] D. La Rocca, P. Campisi, and G. Scarano, ''Stable EEG features for biometric recognition in resting state conditions,'' in *Biomedical Engineering Systems and Technologies*. Berlin, Germany: Springer, 2014.

[25] D. Rodrigues, G. F. A. Silva, J. P. Papa, A. N. Marana, and X.-S. Yang, ''EEG-based person identification through binary flower pollination algorithm,'' *Expert Syst. Appl.*, vol. 62, pp. 81–90, Nov. 2016.

[26] R. Xu *et al.*, ''Feature extraction of individual differences for identification recognition based on resting EEG,'' in *Cross-Cultural Design. Methods, Practice, and Case Studies*. Berlin, Germany: Springer, 2013.

[27] S. Marcel and J. D. R. Millan, ''Person authentication using brainwaves (EEG) and maximum *a posteriori* model adaptation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 743–752, Apr. 2007.

[28] A. B. L. Tort, R. Komorowski, H. Eichenbaum, and N. Kopell, ''Measuring phase-amplitude coupling between neuronal oscillations of different frequencies,'' *J. Neurophysiol.*, vol. 104, no. 2, pp. 1195–1210, 2010.

[29] H.-L. Chan, P.-C. Kuo, C.-Y. Cheng, and Y.-S. Chen, ''Challenges and future perspectives on electroencephalogram-based biometrics in person recognition,'' *Front Neuroinformat.*, vol. 12, p. 66, Oct. 2018.

[30] C. Ward and I. Obeid, ''Application of identity vectors for EEG classification,'' *J. Neurosci. Methods*, vol. 311, pp. 338–350, Jan. 2019.

[31] W. Woerner, ''Test-retest reliability of spectral parameters of the resting EEG in a field sample: A 5 year follow-up in schoolchildren with and without psychiatric disturbances,'' *Electroencephalogr Clin. Neurophysiol.*, vol. 40, pp. 624–628, 1987.

[32] Q. Wu, B. Yan, Y. Zeng, C. Zhang, and L. Tong, ''Anti-deception: Reliable EEG-based biometrics with real-time capability from the neural response of face rapid serial visual presentation,'' *Biomed. Eng. OnLine*, vol. 17, no. 1, p. 55, 2018.

[33] Y.-C. Yu, S. Wang, and L. A. Gabel, ''A feasibility study of using event-related potential as a biometrics,'' in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2016, pp. 4547–4550.

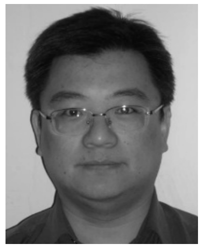
[34] C.-C. Chang and C.-J. Lin, ''LIBSVM: A library for support vector machines,'' *ACM Trans. Intell. Syst. Technol.*, pp. 2:27:1–27:27, 2011.

YANG DI was born in 1995. He received the B.S. degree in biomedical engineering from Nanchang University, Jiangxi, China, in 2013. He is currently pursuing the M.S. degree with the Laboratory of Neural Engineering and Rehabilitation, College of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, China.

His current research interests include machine learning and brain-computer interface.

XINGWEI AN was born in 1986. She received the B.S. degree in biomedical engineering from the Hebei University of Technology, Tianjin, China, in 2008, and the M.S. and Ph.D. degrees in biomedical engineering from Tianjin University, Tianjin, in 2010 and 2015, respectively.

From 2011 to 2013, she was a Visiting Student with the Neurotechnology Group, Berlin Institute of Technology, Berlin, Germany. She is currently an Associate Professor with the Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, China. Her research interests include visual- and auditory-stimuli-based brain-computer interface, machine learning, rehabilitation, and neural engineering.



FENG HE received the B.S. and M.S. degrees in biomedical engineering from Tianjin University, in 1994 and 1998, respectively, and the Ph.D. degree in 2009. After that, he had served as the Chief Technical Officer of Stone Group Tianjin Branch for five years. Since 2003, he has been a Lecturer with the School of Precision Instrument and Opto-Electronics Engineering, Tianjin University, where he is currently the tenure-track Associate Professor. His research interests include neural engineering, biomedical signal detection and processing, and medical instrument design.

SHUANG LIU was born in 1989. She received the B.S. degree in biomedical engineering from Tianjin Medical University, Tianjin, China, in 2012, and the M.S. and Ph.D. degrees in biomedical engineering from Tianjin University, Tianjin, China, in 2018. She is currently an Assistant Professor with the Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, China. Her research interests include the physiological mechanism of emotion, emotion recognition and regulation, and biomarker detection of the depression.

YUFENG KE completed the Ph.D. degree from the Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, China, where he is currently an Assistant Professor with the Academy of Medical Engineering and Translational Medicine. His research interests include brain-computer interfaces, brain stimulation and adaptive human-machine interaction systems.

DONG MING received the B.S. and Ph.D. degrees in biomedical engineering from Tianjin University (TJU), China, in 1999 and 2004, respectively. He was a Research Associate with the Department of Orthopaedics and Traumatology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, from 2002 to 2003, and was a Visiting Scholar with the Division of Mechanical Engineering and Mechatronics, University of Dundee, U.K., from 2005 to 2006. He joined as a Faculty with the College of Precision Instruments and Optoelectronics Engineering, TJU, in 2006, and has been promoted as a Full Professor of biomedical engineering, since 2011. He is currently the Chair Professor with the Department of Biomedical Engineering, TJU, the Director of the Tianjin Neural Engineering Center, the Director of the Academy of Medical Engineering and Translational Medicine, and the Chair of IEEE-EMBS Tianjin Chapter.

His major research interests include neural engineering, rehabilitation engineering, sports science, and biomedical instrumentation and signal/image processing, especially in brain–computer interface, functional electrical stimulation, and gait analysis. He has managed almost 20 national and international research projects. He designed I-NeuroSys Robot TUNEXON-I, which is the first intention controlled artificial neural robot for whole body rehabilitation of stroke patients. It is one of the top 10 technical achievements of China to change the future in 2014. He and his colleagues have successfully carried out the first in-space BCI experiment on the Tiangong II space station. Furthermore, he has been an International Advisory Board Member of the The Foot, and the Editorial Committee Member of the *International Journal of Biomedical Engineering* in China. He has also organized and hosted over 10 international conferences as the Session Chair or Track Chair over the last 10 years and was the General Chair of the 2012 IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems (VECIMS 12).