

Research Article

An ECoG-Based Binary Classification of BCI Using Optimized Extreme Learning Machine

Xinman Zhang ¹, Qi Xiong ¹, Yixuan Dai,¹ Xuebin Xu,² and Guokun Song³

¹School of Automation Science and Engineering, Faculty of Electronic and Information Engineering, MOE Key Lab for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

²Guangdong Xi'an Jiaotong University Academy, No. 3, Daliangdesheng East Road, Foshan, Guangdong 528000, China

³Sichuan Gas Turbine Research Institute of AVIC, No. 6 Xinjun Road, Xindu District, Chengdu, Sichuan Province, China

Correspondence should be addressed to Qi Xiong; 284983937@qq.com

Received 12 January 2020; Revised 29 May 2020; Accepted 3 June 2020; Published 30 June 2020

Academic Editor: Danilo Comminiello

Copyright © 2020 Xinman Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the accuracy of brain signal processing and accelerate speed meanwhile, we present an optimal and intelligent method for large dataset classification application in this paper. Optimized Extreme Learning Machine (OELM) is introduced in ElectroCorticoGram (ECoG) feature classification of motor imaginary-based brain-computer interface (BCI) system, with common spatial pattern (CSP) to extract the feature. When comparing it with other conventional classification methods like SVM and ELM, we exploit several metrics to evaluate the performance of all the adopted methods objectively. The accuracy of the proposed BCI system approaches approximately 92.31% when classifying ECoG epochs into left pinky or tongue movement, while the highest accuracy obtained by other methods is no more than 81%, which substantiates that OELM is more efficient than SVM, ELM, etc. Moreover, the simulation results also demonstrate that OELM will significantly improve the performance with p value being far less than 0.001. Hence, the proposed OELM is satisfactory in addressing ECoG signal.

1. Introduction

The development of brain-computer interfaces (BCI) has undergone extensive growth in recent years, with the aim of providing an effective method for human-computer interaction without neuromuscular transmission [1]. The ultimate goal of BCI research is to establish a direct communication system that translates human intentions, which is reflected by specific brain signals, into a control command for output devices [2].

According to the method by which users derive their neural signals, BCI can be classified into noninvasive, invasive, and partially invasive ones. ElectroCorticoGram (ECoG), which is obtained by putting electrical electrodes directly on the cortex, has attracted substantial and increasing interest and has been the dominant signal used for invasive BCIs due to its high spatial and temporal resolution [3]. In 2004, the first online ECoG BCI study by Leuthardt et al. provided initial evidence that ECoG signals contain

information about the direction of hand movements, which was one of the earliest demonstrations to show that specific details of motor function can be accurately inferred without measurements from individual neurons [4]. These properties of ECoG offer the possibility of a new nonmuscular communication and control channel, a practical BCI system. By converting ECoG to machine instructions that control peripheral devices, BCI enables users to interact with the outside world with their own thought.

Nowadays, there exists a considerable part of groups which cannot manually control machines. Besides, the study on the human brain in military applications and living entertainment has a greater command. Therefore, the research and development of an effective method is very promising. So far, scientists from a multitude of disciplines have achieved good results in this field. For instance, American Wadsworth Center has designed a multi-classification BCI based on P300 potentials, so paralyzed people with atresia can input 36 characters, numbers, and

spaces through the signal corresponding to specific brain activities, rather than using their own fingers [5]. Besides, BCI research team of Science and Technology University from Austria has also designed an algorithm for different motor imagery signal classification, so that patients with paralyzed arm can achieve a simple action of drinking. In China, the BCI research team from Tsinghua University has designed an automatic dialing system that controls telephone, which is connected to a computer for real-time dial by interpreting brain thinking to corresponding numbers [5].

In recent years, many innovative methods are commonly used in processing binary-class ECoG signals of motor imagery [6, 7]. For example, support vector machine (SVM) is a supervised learning model that is used for pattern recognition, classification, and regression analysis. However, when processing training samples in large size, it is difficult to implement and cannot produce satisfactory results whether in precision or speed. Therefore, a newly different classification method named extreme learning machine (ELM) is proposed by scholar Guang-Bin Huang from Nanyang Technological University [8]. ELM has better generalization performance and fast executing time than SVM. We adopt a newly developed method which is called Optimized Extreme Learning Machine (OELM) to distinguish the imagined movements between left pinky and the tongue on the basis of ELM. It serves as a classification algorithm that classifies two different kinds of ECoG signals quickly but accurately. We expect that results obtained by OELM show superior performance in both classification accuracy and speed. Especially when processing signals in large size, it is supposed to save a lot of time.

This paper is organized as follows: Section 2 describes the data acquisition and description, including how data are acquired and distributed for training and testing. In Section 3, we focus on basic BCI algorithm research and implementation, where we present that OELM serves as the classifier. Prior to that, we use CSP to extract features. Principles and procedures of these algorithms are described in detail. Section 4 carries out experimental results and analysis on ELM and OELM, respectively. Finally, the paper is concluded in the last section.

2. Data Acquisition and Description

Based on the existing research, there is substantial theoretical and empirical evidence that ECoG could support a clinically and functionally reliable BCI with a high level performance. Thus, it is reasonable to envisage that an ECoG-based implant could substantially enhance the functional capability of a disabled patient by enabling their ability to modulate their environment, communicate, or control a prosthesis [9]. In order to evaluate the proposed classification algorithms, dataset I of BCI competition III is adopted in this study. It is provided by the University of Tübingen, Germany, Dept. of Computer Engineering (Prof. Rosenstiel) and Institute of Medical Psychology and Behavioral Neurobiology (Niels Birbaumer), etc. [10]. Compared to signals acquired from the scalp, such as

electroencephalography (EEG), and intraparenchymal single neuronal recordings, ECoG recordings show characteristics that make them especially suited for basic neuroscience research. These characteristics include high spatial resolution and signal fidelity, resistance to noise, and substantial robustness over long recording periods. Thus, we regard the ECoG data as the most suitable datasets to validate the OELM.

All ECoG data were collected during two imagined movements of either left pinky or tongue. Recordings were performed with a sampling rate of 1000 Hz from 64 platinum electrodes, whose size is approximately 8×8 cm [10]. The electrodes are placed as an array which covers a specific area of cortex. Considering that brain potential is weak and prone to interference, we are supposed to characterize the incoming electrosensory signals. Electronic sensors are put inside the electrodes and all fitted with serial current-limiting resistors to measure afferent signal intensity and guarantee the high fidelity [11].

There are 378 trials in total. Every single trial represents either an imagined tongue or finger movement and is recorded for 3 seconds duration. To avoid visually evoked potentials being reflected by data, recording intervals start 0.5 seconds after visual cue has ended. After being amplified and filtered, the recorded potentials are stored as microvolt values [10]. We measure minute differences in the voltage between neurons among each trial.

Among the total 378 trials, 278 labeled trials are used to train classifiers, whereas other 100 unlabeled testing trials are available for measuring generalization performance of the trained classifier. Concerning on the situation that training data and testing data are picked up from the same subject with the same task, but on a different date with about one week apart, the design of our classifier remains challenging. More detailed description can be found in [10]. The task in this study is to try to correctly classify the testing dataset (100), where all these samples belong to either negative or positive class. Our goal is to turn those tiny voltage measurements into two imagery robotic movements of left pinky or tongue.

The detailed block scheme of the BCI system is presented in Figure 1 [12].

In Figure 1, the grid of 8×8 ECoG platinum electrodes is placed on the contralateral (right) motor cortex. Signals are obtained from the brain through the primary electrosensory afferents. We transmit ECoG signals to computer for subsequent data processing. Peripheral devices can then be controlled by the generated motor commands [13].

3. Basic BCI Algorithm Research and Implementation

In this section, first we adopt CSP to extract features of ECoG signal and then transfer them to feature classification module, using ELM to train and test the corresponding data. Then an improved classification algorithm named Optimized Extreme Learning Machine (OELM) is proposed and put into practice on the basis of ELM. Here the basic

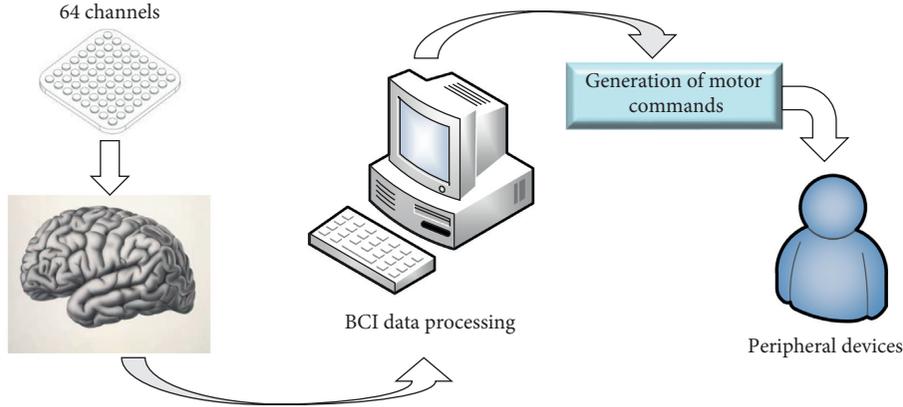


FIGURE 1: The main procedure of the BCI system.

principles of these three algorithms are introduced. Also, the detailed process and procedure are presented.

3.1. Common Spatial Pattern Algorithm Principle. Common spatial pattern (CSP) is an effective method for feature extraction in discriminating two kinds of data. In recent years, CSP has become greatly popular to extract ECoG features. It is a signal processing method based on two or more different brain potentials, and it filters the signal in space [14]. The fundamental idea is that after filtering, spatial energy of two kinds of signals holds the biggest differences. That is to find a projection direction that discriminates two classes of ECoG data, by maximizing the variance of one class while minimizing the variance of the other. The basic principles are as follows.

Assuming that X_1 and X_2 , respectively, represent two different types of ECoG signals, and their dimensions are both $N \times T$, where N and T represent the number of channels and measurement samples corresponding to a single channel, respectively. Two covariance matrices C_1 and C_2 are calculated as follows:

$$C_1 = \frac{X_1 X_1^T}{\text{trace}(X_1 X_1^T)}, \quad (1)$$

$$C_2 = \frac{X_2 X_2^T}{\text{trace}(X_2 X_2^T)}, \quad (2)$$

where X^T represents the transpose of matrix X , while $\text{trace}(X)$ gives the sum of diagonal elements of X . In order to get the experimental data with higher accuracy and lower error rate, multiple experiments have been carried out. Finally, take $\overline{C}_1, \overline{C}_2$ as the average value of C_1, C_2 , then the covariance matrix is obtained by using principal component analysis as shown in the following formula:

$$C = \overline{C}_1 + \overline{C}_2 = U \Lambda U^T, \quad (3)$$

where U is the matrix composed by eigenvectors of the mixed covariance matrix C , while Λ is a diagonal matrix composed by eigenvalues correspondingly.

The whitening transformation matrix P is represented by the following formula:

$$P = \sqrt{\Lambda^{-1}} U^T. \quad (4)$$

Then, matrices C_1 and C_2 carry on the whitening transformation, respectively. Results are obtained as shown in the following formula:

$$\begin{aligned} S_1 &= P \overline{C}_1 P^T, \\ S_2 &= P \overline{C}_2 P^T, \end{aligned} \quad (5)$$

where S_1 and S_2 share common eigenvectors. It can be proved that if two diagonal matrices are summed, the result is an identity matrix. That is,

$$\Lambda_1 + \Lambda_2 = I. \quad (6)$$

From formula (6), it is not difficult to find out that the maximum variance of one class leads to the minimum variance of the other. Thus, the two spatial filters are designed according to such property. Sorting the eigenvalues in descending order, then take out the maximum m value in Λ_1 . The first spatial filter F_1 is constructed using the m eigenvectors. In the same way, the maximum m eigenvalues in Λ_2 are taken out, and the corresponding eigenvectors are used to construct the second spatial filters F_2 [15].

After two kinds of filters are obtained, the original multichannel ECoG signals are divided into two categories, which can be formulated as follows:

$$\begin{aligned} Z_1 &= F_1 X, \\ Z_2 &= F_2 X. \end{aligned} \quad (7)$$

Finally, the characteristics of two kinds of signals are constructed:

$$\begin{aligned} f_1 &= \log_2 \left(\frac{\text{var}(Z_1)}{\text{var}(Z_1) + \text{var}(Z_2)} \right), \\ f_2 &= \log_2 \left(\frac{\text{var}(Z_2)}{\text{var}(Z_2) + \text{var}(Z_1)} \right). \end{aligned} \quad (8)$$

In formula (8), in order to make the two types of features more close to normal distribution, the base of logarithmic operation is set to 2.

3.2. *Optimized Extreme Learning Machine.* With its fast speed and high precision, Optimized Extreme Learning Machine (OELM) is more effective in discriminating two classes of ECoG data than conventional algorithms. As the basis of OELM, Extreme Learning Machine (ELM) has been paid more attention by researchers around the world in recent years. It is a neural network essentially, which is composed of input layer, hidden layer, and output layer [16]. Different from other traditional learning algorithms for a neural type of SLFNs, ELM aims to reach not only the smallest training error but also the smallest norm of output weights [17].

Assuming that a single-hidden layer feed-forward neural network (SLFN) is given, and it has N hidden layer nodes. Different from the previous algorithm that all parameters need to be tuned by feed-forward neural network, ELM can accurately learn N different observation values with no need to adjust the weights between input neurons and initial hidden layer bias in practical application [18]. In fact, many simulation results also show that ELM is not only fast in classification speed, but also can yield very high recognition accuracy for its universal approximation capability [19]. The steps of ELM are presented as follows.

Firstly, randomly given N sample pairs (x_j, o_j) , where x_j and o_j , respectively, denote the input and the output, $x_j = [x_{j1}, x_{j2}, x_{j3}, \dots, x_{jm}]^T \in R^m$ and $o_j = [o_{j1}, o_{j2}, o_{j3}, \dots, o_{jl}]^T \in R^l$. Given the number of hidden layer nodes in the network is \tilde{N} , then the corresponding SLFN is expressed as follows:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(x_j) = O_j, \quad (9)$$

where $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ is the weight between the i th hidden neurons and the m th output neurons and $g(x) = [g_1(x), \dots, g_{\tilde{N}}(x)]$ is the output vector of the hidden layer with respect to the input x . It is worth noting that $g(x)$ actually maps the data from input space to hidden layer space, and thus, $g(x)$ is indeed a feature mapping. o_j is the output of the j th neuron. For additive nodes with activation function g , g_i is defined as follows:

$$g_i = g(a_i \cdot x_i + b_i), \quad (10)$$

where $a_i = [a_{i1}, a_{i2}, \dots, a_{im}]$ is the weight vector connecting the i th hidden neuron and the input neurons and b_i is the bias of the i th hidden neuron. The selection of activation function is not unique.

All the equations above can be written compactly as follows:

$$H\beta = O, \quad (11)$$

where $\beta = [\beta_1, \beta_2, \dots, \beta_N]^T$ is the vector of the output weights between the hidden layer of N nodes and the output node. H is the hidden layer output matrix:

$$H = \begin{bmatrix} g(x_1) \\ \vdots \\ g(x_N) \end{bmatrix} = \begin{bmatrix} g_1(x_1) & \cdots & g_N(x_1) \\ \vdots & \vdots & \vdots \\ g_1(x_N) & \cdots & g_N(x_N) \end{bmatrix}. \quad (12)$$

Assuming that the smaller the norms of output weights are, the better the generalization performance the networks tend to have. ELM is to minimize the training error as well as the norm of the output weights. The following mathematical model is established:

Minimize:

$$\begin{aligned} & \|H\beta - T\|^2, \\ & \|\beta\|, \end{aligned} \quad (13)$$

where T is the expected output. The minimal norm least square method is used as follows:

$$\beta = H^{-1}T, \quad (14)$$

where H^{-1} is the inverse of matrix H [19].

Since ELM has no need of iteration, the learning speed is much faster than traditional classification algorithms. By continuously adjusting the number of hidden layer nodes, the learning ability and classification accuracy can both achieve an optimal value. According to the ELM algorithm above, a widespread type of activation functions g can be used in ELM so that ELM can approximate any target continuous function T . With this universal approximation capability, the bias in the optimization constraints of SVM can be removed [20], which explains for the better generalization performance and lower computational complexity of ELM.

Although ELM outperforms SVM in both classification speed and accuracy, it is very unstable when processing high-dimensional but small samples, which results from the random assignment of input weights. Therefore, we proposed an improved algorithm based on ELM, which is called Optimized Extreme Learning Machine (OELM), where projection of feature signal is introduced. OELM improves some existing disadvantages of ELM: ELM needs more hidden neurons than BP; the generalization performance of the ELM depends on the proper selection of constant parameters, especially for a small number of training samples. According to reference [21], singular value decomposition (SVD), as a linear dimensionality reduction, aims at mapping the original data to a lower-dimensional space using a projection matrix. To reduce the complexity of our network, we assign the result of SVD to the input layer. Steps of OELM are presented as follows on the basis of ELM.

Firstly, the characteristic of input signal is represented by matrix $X_{N \times m}$, where N and m , respectively, represent the number of samples and attributes of the signal.

Secondly, do SVD to input matrix. SVD is widely used in image processing, signal classification, pattern recognition, and so on. In this experiment, SVD is represented as follows:

$$\text{svd}(X) = PSQ^T, \quad (15)$$

where P and Q are the left and right singular matrices of input matrix X . The singular value matrix S is composed of singular values which are arranged in descending order. Select the d singular vector elements in S corresponding to the largest singular values, which is used to approximate

input matrix X . Finally, the optimal rank of X is obtained as follows:

$$\text{best}_d(X) = P_d S_d Q_d. \quad (16)$$

Next, in the low-dimensional space composed by Q_d , high-dimensional data are represented as follows:

$$XQ_d = P_d S_d, \quad (17)$$

where Q_d is known as the projection vector.

Then, in order to overcome the defect of poor performance in processing high-dimensional small sample, we set the input layer weights for projection vector, instead of giving random value. That is, $W = Q_d$. This improvement simplifies the complexity of the network. The output of the hidden layer can be obtained by the following formula after determining input layer weights:

$$H = g(XQ_d) = g(P_d S_d), \quad (18)$$

where $g(x)$ is a single-hidden layer neural network transfer function.

At last, the output layer weights can be calculated by means of a linear expression as follows:

$$H\beta = T. \quad (19)$$

The training module is over after the above five basic steps. In fact, we need not learn input weights once again in OELM and only output weights need to be calculated by least square. The deterministic assignment of the input layer overcomes the defect of ELM, whose classification accuracy changes dynamically due to the random assignment of input weights [22]. Results also show that the classification performance of OELM is very stable and relatively explicit compared with the results of ELM.

3.3. Description of the Classification Procedure. Based on the description given in Sections 3.1 and 3.2, features of brain signals can be extracted by CSP at every sampling point. We adopt the extracted features to train OELM classifier in the training dataset, and then the trained classifier is applied to classify the features extracted at the same sampling point in the testing dataset.

The detailed procedure of OELM classification strategy is described in Algorithm 1.

With algorithms mentioned above, experiments can be conducted then. Figure 2 illustrates the detailed block diagram of the proposed ECoG processing scheme.

Firstly, we sample and collect signals from the brain. In this experiment, we increase the hidden layer node by one degree from 10 to 60. Sampling points are simply set to 1500, 2000, and 2500, respectively; then CSP is applied to extract features after acquiring ECoG signals from 64 electrodes; at last, 16 different activation functions are chosen to OELM for classification. All experiments are performed on several sets of parameters with 278 trials to train OELM and the remaining 100 to test the performance of the trained OELM.

4. Experiment Results and Discussion

We conduct all experiments under the exactly same development environment. Experiments are performed on the same computer with Intel Core i3 2 processor of 2.4 GHz with 4 GB memory and implemented by MATLAB 7.12 (2011a, 64 bit). Section 4.1 lists several parameters which evaluate the performance of the system, while Section 4.2 shows the distribution of signal's amplitude obtained by CSP. Results of the mentioned methods are presented and discussed in Section 4.3. Bold value in tables indicates the best result among the listed.

4.1. Performance Evaluation. To evaluate the performance of a BCI, there are five main indicators as follows:

- (1) *Classification accuracy.* It refers to the correct classification rate. It is a fundamental indicator to judge whether the BCI system meets the requirements or not. By comparing the true label of each trial with the label obtained from the testing stage, we get the number of equaled pairs and divide it by total trials to obtain classification accuracy.
- (2) *Training time.* Time (seconds) spent on training classifier.
- (3) *Testing time.* Time (seconds) spent on predicting all testing trials.
The shorter the training and testing time, the better the system performance.
- (4) *Number of hidden layer nodes.* The less the hidden layer nodes, the lower the complexity of the network.
- (5) *p value.* We used *t*-test to compare the discrimination of classification accuracies corresponding to ELM and OELM:

$$[H, p] = t \text{ test}(X, Y). \quad (20)$$

Formula (20) performs a paired *t*-test of the hypothesis, where two matched samples in vectors X and Y come from distributions of ELM and OELM with equal means and return the result of testing in H . $H=0$ indicates that the hypothesis ("equal means") cannot be rejected at the 5% significance level. $H=1$ indicates that the hypothesis can be rejected at the 5% level.

However, high classification accuracy of BCI may sacrifice the training or testing time, due to the high complexity of algorithm. In turn, the rapid classification speed is generally obtained at the expense of lowering accuracy [23]. Therefore, we aim to get a compromise of high classification accuracy and less time consumption to meet the system requirements.

4.2. Preprocessing Stage. After extracting features with CSP, we can calculate the amplitude of all signals from 64 electrodes. To show it in an intuitive way, we randomly choose nine continuous trials (trial 6 to trial 14) from 278 training trials and then plot trials 9 and 10 as shown in Figure 3,

- (1) Let $P = \text{train_data}(:, 2: \text{size}(\text{train_data}, 2))'$, which denotes the sample of training data; $T = \text{train_data}(:, 1)'$, which denotes the given target label of training data sample.
- (2) Preprocessing the whole data of classification.
- (3) Processing targets of training and targets of testing.
- (4) Calculate the singular value decomposition of P : $[u, s, v] = \text{svd}(P', \text{econ}')$.
- (5) Set number of hidden neurons N' .
- (6) Set input weights. Let $\text{InputWeight} = u(:, 1: N')$.
- (7) Calculate hidden neuron output matrix under specific activation function $g(x)$.
- (8) Calculate output weights $\text{OutputWeight} = \text{pinv}(H') * T'$.
- (9) Input the testing data test_data . Then calculate hidden neuron output matrix H_{test} .
- (10) Calculate the actual output of testing data $TY = (H_{\text{test}}' * \text{OutputWeight})'$.
- (11) Calculate CPU time (seconds) spent by OELM predicting the whole testing data.
- (12) Calculate the testing classification accuracy.

ALGORITHM 1: Given training dataset train_data , which consists of samples and their corresponding target labels. For binary classification, target labels can be only -1 or 1 . Set activation function $g(x)$.

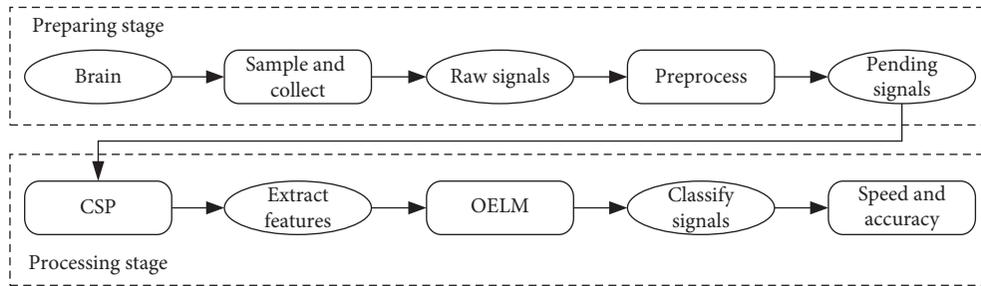


FIGURE 2: Detailed block scheme of CSP and OELM.

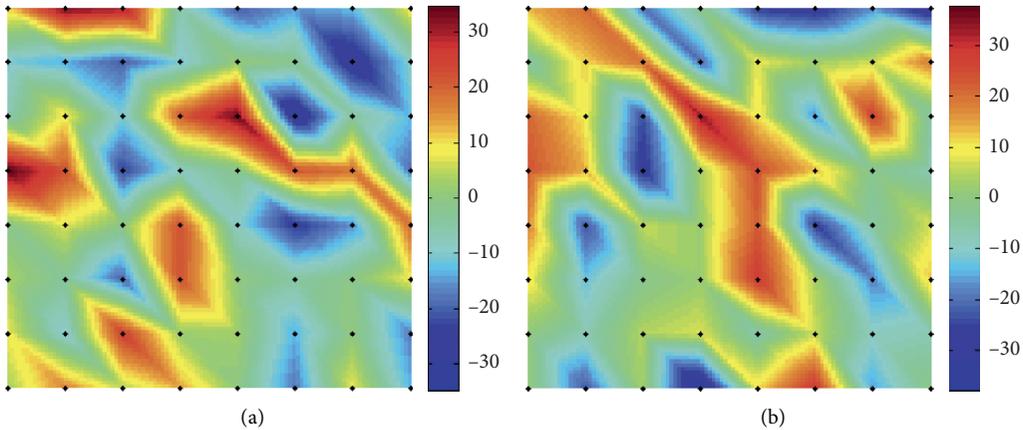


FIGURE 3: Relative amplitude contribution of (a) trial 9 and (b) trial 10. Black dots denote 64 channels, and each color bar represents the range of signal voltage.

which illustrate the relative contribution of signal's amplitude corresponding to these two trials.

As shown in Figure 3, trials 9 and 10 have rather different contour line distribution, while the region in red of Figure 3(a) basically shows blue of Figure 3(b), which represents low voltage potential. To highlight the difference further, we select a relatively middle channel (30th) and plot the amplitude of these two trials at every sampling point in Figure 4.

In Figure 4, the mean amplitude value of trial 10 is higher than 20, while the mean value of trial 9 is lower than 0. This strongly illustrates that CSP can extract features effectively. We can easily notice that in Figure 3, region of interest (ROI) is concentrated over the sensorimotor cortex area [24], where locates the 22nd, the 29th to the 32nd, and the 37th to the 40th channel. We hand-select those more discriminative channels and pick out trials 9 and 10, whose level curves have the largest difference among the other one, and plot their

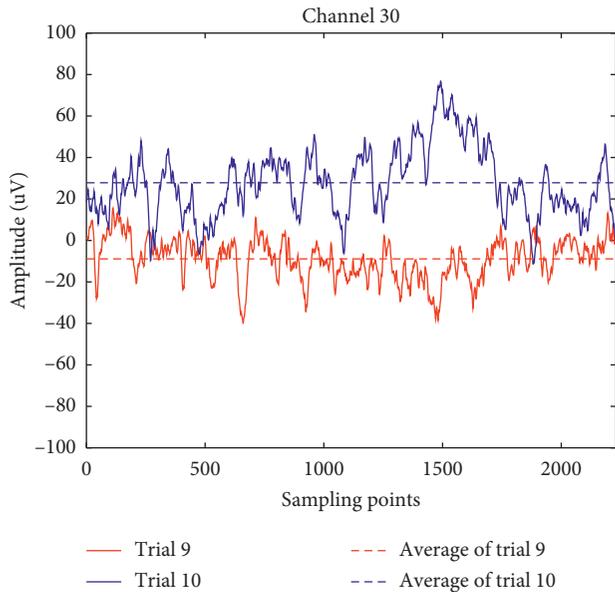


FIGURE 4: Trial 9 and 10's amplitude of channel 30 in training data.

amplitude-sampling points curves corresponding to those nine channels, as shown in Figure 5.

In Figure 5, the curves in red represent the voltage of trial 9 while the curves in blue denote that of trial 10 from raw training data. The straight lines parallel to the horizontal axis show the mean values corresponding to each trial. Channels are regarded as “good” if they are visibly discriminable on average between two classes [25]. In the whole 278 training trials, we list the exact mean voltage of those channels in Table 1 and the last column presents label for each trial.

As presented in Table 1, label of trial 9 is 1 while label of trial 10 is -1 , which explains the difference of curves among those nine channels in Figure 5. We transfer the features extracted by CSP to SVM, ELM, and OELM, and the results are calculated as follows.

4.3. Results on Extreme Learning Machine. We set sampling points to 1500, 2000, and 2500 roughly at first. Experimental results are obtained under 5 different activation functions, which are “Sine” (sin), “Sigmoidal” (sig), “Hard limit” (hardlim), “Triangular basis” (tribas), and “Radial basis” (radbas). We list the optimal results corresponding to specific hidden layer nodes in Table 2.

Table 2 gives the performance of 5 activation functions based on ELM with sampling point increasing from 1500 to 2500. It is noteworthy here that the maximum of classification accuracy (80.77%) is obtained with “Sine” (sin) and “Sigmoid” (sig) function. Besides, we can also get a relatively shorter time, especially when the activation function takes “sig,” the average training time and average testing time are both 0.0006 seconds. Compared with experimental results under other conditions, system performance is optimal.

It is known that training and testing time are both relatively good; however, classification accuracy is still not good enough. The reasons are as follows: on the one hand, as the feature extraction method, CSP does not take the

frequency domain information into account, which will generate unrepresentative ECoG features, and affect classification accuracy later [25]; on the other hand, we apply ELM in feature classification. Due to its random assignment of input weights, it is difficult to get global optimum in the process of finding the optimal weight [26]. These two points make sense as CSP combined with ELM cannot reach satisfactory results. Then, OELM is proposed and put into practice as shown in the next section.

4.4. Results on Optimized Extreme Learning Machine. For OELM, the kinds of activation functions are up to 16. Similarly, we set sampling points to 1500, 2000, and 2500. So 48 experiments are completed and we list the optimal results under every parameter setting in Table 3 without the “Radial basis” (radbas), “Triangular basis” (tribas), “Hard limit” (hardlim), “Cosine” (cos), “CosineH” (cosh), and “arcCosineH” (acosh), whose classification accuracies are lower than 70%.

Table 3 suggests that in binary classification, a fixed parameter setting for each activation function is chosen which works well among all experiments. Compared with the optimal accuracy (80.77%) in Table 2, the best of Table 3 is 84.62%, which wins with 3.85%. It is noteworthy here that the average training and testing time obtained by OELM are very stable, which are slightly faster than that obtained by ELM. When activation function takes “sin” and “sig,” hidden layer node numbers both set to 53, and sampling point is 1500, classification accuracy can reach to 84.62% and 83.33%, respectively, with average training time being 0.0049 seconds and 0.0052 seconds, while average testing time being lower than 0.0001 seconds and 0.0012 seconds, respectively.

It also need to be emphasized here that the activation functions “sin,” $g(x) = \sin(x)$, and “sig,” $g(x) = S(x) = 1/(1 + e^{-x})$, are used in ELM and OELM classifiers for their better performance. Since the singular value decomposition (SVD) can reduce data dimension effectively, it is better to combine SVD with these two activation functions. This, together with a closer look at results from Table 2, suggests that “sin” and “sig” are more valuable for our binary decision tasks than others.

We still expect a further improvement of classification accuracy, and proper selection of sampling points and hidden layer nodes may markedly functioning [27]. We are supposed to lessen the number of hidden layer nodes, which can effectively reduce the complexity of network. Figure 6 compares performance at different sampling points setting. For each case, we calculate the least number of hidden layer nodes and plot it in secondary axis.

Figure 6 suggests that when sampling points set to 2150, the result is relatively optimal. The classification accuracy can improve to 92.31% when adopting OELM as the classifier with either “sin” or “sig” as the activation function, while the number of hidden layer nodes both being 33, which is relatively less compared with that of ELM, whose nodes are more than 38. Figure 7 depicts the classification accuracy at every hidden layer node number.

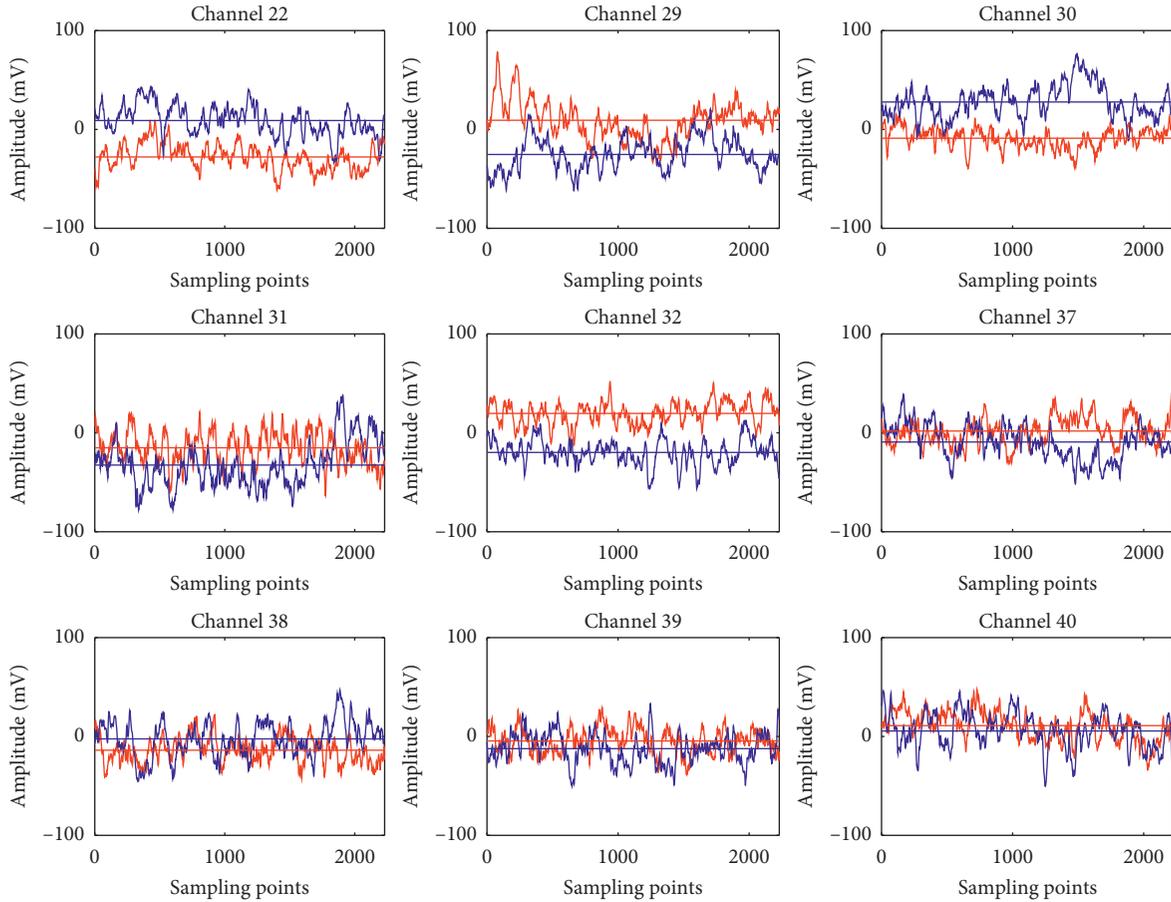


FIGURE 5: Amplitude of nine channels from trials 9 and 10 of training data.

TABLE 1: Mean voltage of nine selected channels of nine trials.

Trial	#22	#29	#30	#31	#32	#37	#38	#39	#40	Label
Trial 6	-10.34	-8.20	-22.98	22.68	35.47	22.58	-0.79	7.58	-0.94	1
Trial 7	13.73	5.08	-9.34	12.27	-18.63	-1.60	-16.07	-13.51	27.03	1
Trial 8	17.08	17.29	-19.81	-29.63	20.85	5.23	-54.27	-11.33	-7.63	-1
Trial 9	-27.86	9.27	-8.89	-15.15	19.61	1.80	-13.88	-4.54	10.96	1
Trial 10	9.00	-25.32	27.76	-32.60	-19.84	-9.29	-2.45	-12.32	5.55	-1
Trial 11	-9.65	-7.49	14.66	-22.26	-11.35	-15.80	-5.19	14.67	4.24	-1
Trial 12	3.00	-24.68	-4.28	7.94	-14.20	-3.41	2.82	-22.35	-13.26	-1
Trial 13	-30.10	33.16	3.62	18.61	-2.22	17.61	13.30	-13.73	26.55	1
Trial 14	-24.01	3.48	-28.78	14.30	23.73	2.27	5.63	-23.09	18.06	-1

For each activation function, different numbers of hidden layer nodes, ranging from 10 to 60, are applied to ELM and OELM. Thereafter, classification accuracy corresponding to each set of parameters is calculated on testing data. Tables 4 and 5 compare the performance of ELM with that of OELM, with activation functions being “sig” and “sin.”

In Tables 4 and 5, “Std” and “RMSE” mean standard deviation and means root mean square error, respectively. According to formula (20), we set testing accuracy obtained by ELM as parameter X while that of OELM as Y in t -test function.

Both results show that H is equal to 1, which means two testing accuracy samples are rather different. p values are far less than 0.0001, which indicates that the mean values of the two samples are equal with the probability of less than 0.01%. This confirms that OELM obtains significant improvement in classification accuracy in comparison with ELM.

According to results in Table 4, the proposed OELM yields a maximum accuracy of 92.31%, which increases 16.67% compared with the optimal result (75.64%) of ELM under the same parameters. Likewise, in Table 5, the proposed OELM outperforms ELM with 12.82%. On average, the increase of classification accuracy achieved by OELM is

TABLE 2: Objective classification performance of ELM.

Activation function	Parameters		Evaluation metrics		
	Sample point	Hidden layer nodes	Classification accuracy (%)	Average training time (s)	Average testing time (s)
sin	1500	41	75.64	0.0009	<0.0001
	2000	58	80.77	0.0043	0.0012
	2500	57	76.92	0.0055	0.0012
sig	1500	34	78.21	0.0040	0.0034
	2000	49	79.49	0.0018	0.0018
	2500	32	80.77	0.0006	0.0006
hardlim	1500	25	67.95	0.0018	0.0012
	2000	50	78.21	0.0015	0.0031
	2500	18	74.32	0.0012	0.0012
tribas	1500	40	67.95	0.0015	0.0049
	2000	57	71.79	0.0015	<0.0001
	2500	53	69.23	0.0018	0.0012
radbas	1500	43	71.79	0.0021	0.0006
	2000	58	69.23	0.0043	0.0006
	2500	50	75.64	0.0012	0.0012

TABLE 3: Objective classification performance of OELM.

Activation function	Parameters		Evaluation metrics		
	Hidden layer node number	Classification accuracy (%)	Average training time (s)	Average testing time (s)	
sin	53	84.62	0.0049	<0.0001	
sig	53	83.33	0.0052	0.0012	
tan	53	84.62	0.0061	0.0012	
asin	53	84.62	0.0049	0.0012	
atan	53	84.62	0.0058	0.0006	
acos	52	80.77	0.0077	<0.0001	
asinh	53	84.62	0.0070	<0.0001	
atanh	53	84.62	0.0098	0.0006	
tanh	53	84.62	0.0055	0.0012	
sinh	53	84.62	0.0077	<0.0001	

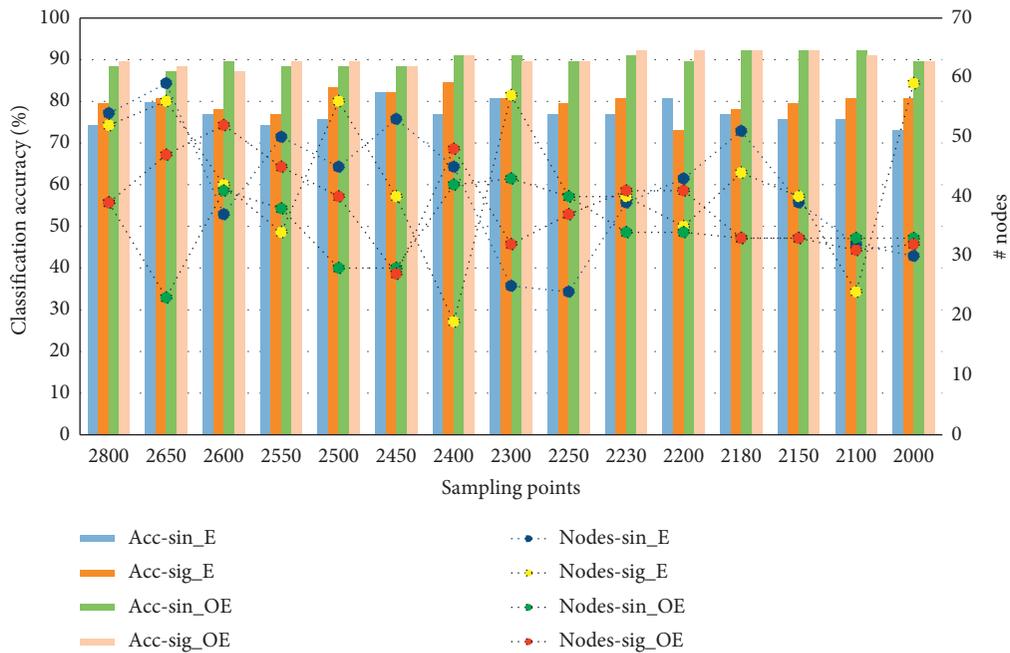


FIGURE 6: Bar plot visualizes the performance corresponding to each sampling point number.

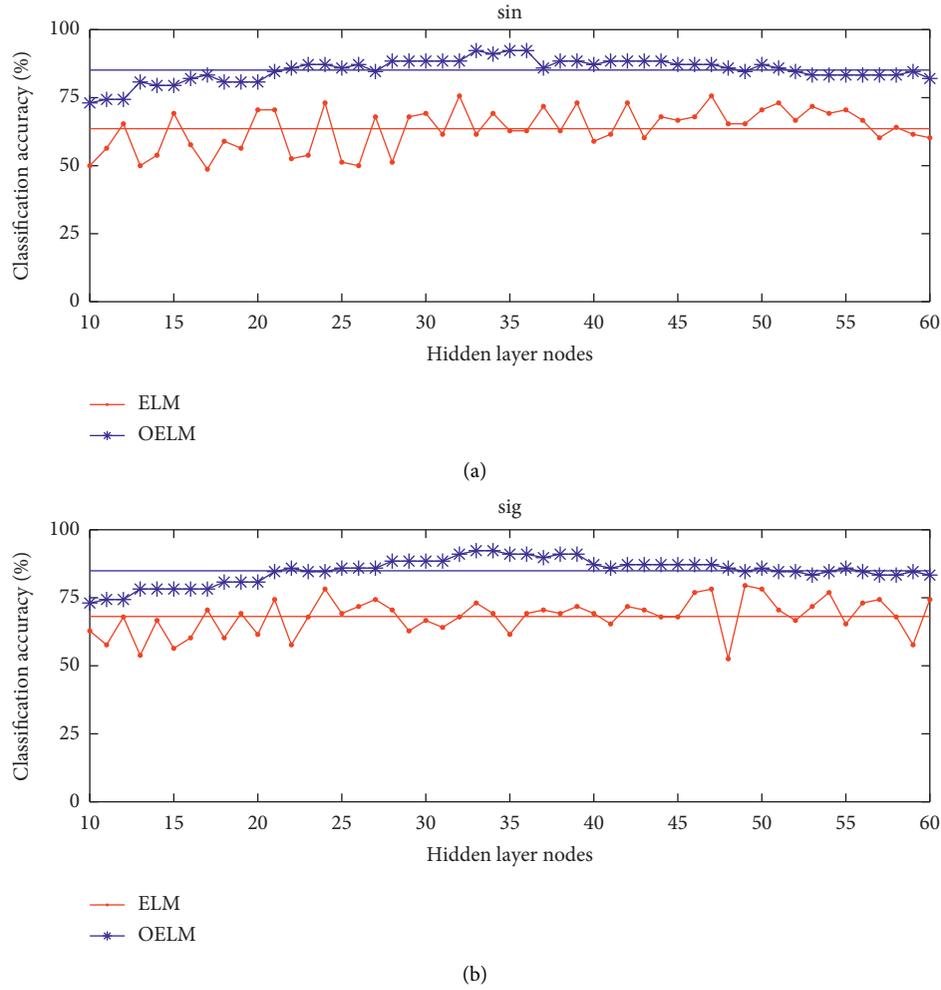


FIGURE 7: Classification accuracy of ELM and OELM with the same parameters.

TABLE 4: Performance comparison of ELM and OELM for activation function sig.

Algorithms	Time (s)		Testing accuracy (%)				p value	H	#nodes
	Training	Testing	Max	Mean	Std	RMSE			
ELM	<0.0001	<0.0001	75.64	66.26	6.07	34.61	$p < 0.0001$	1	40
OELM	<0.0001	<0.0001	92.31	84.92	4.61	15.92			

TABLE 5: Performance comparison of ELM and OELM for activation function sin.

Algorithms	Time (s)		Testing accuracy (%)				p value	H	#nodes
	Training	Testing	Max	Mean	Std	RMSE			
ELM	0.0156	<0.0001	79.49	64.88	6.80	36.11	$p < 0.0001$	1	39
OELM	0.0312	<0.0001	92.31	85.14	4.26	15.60			

21.97% and 23.80% in comparison with ELM under the same condition. As observed from Tables 4 and 5, generally speaking, ELM and OELM obtain similar performance in classification speed. However, the number of hidden nodes required by ELM is larger than that needed by OELM, meaning that the complexity of OELM is much lower than ELM [28].

Figure 8 shows the maximum value, mean value, standard deviation, and root mean square error of testing accuracy obtained by ELM and OELM.

Figure 8 reveals that for whether “sin” or “sig” function, the maximum and mean values of classification accuracy obtained by OELM are rather higher than those achieved by ELM.

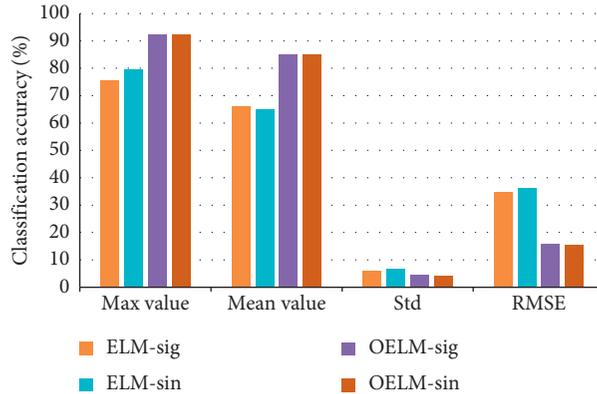


FIGURE 8: Evaluation index of ELM and OELM under the two activation functions.

TABLE 6: Classification results comparing the proposed system with other systems.

Contributor	Research lab	Feature extraction	Classifier	Accuracy (%)
Qingguo Wei et al.	Tsinghua University	CSSD and FDA	Linear SVM	91
Paul Hammon	University of California, San Diego	ICA and AR	SVMs	87
Florian Knoll et al.	Graz University of Technology	AAR and band power	LDA	84
Kiyong Yang et al.	University of Southern California	Correlation coefficient	SVM	81
Archis Gore	Fergusson College, Pune	Band power	NNW and FDA	79
Hyunjin Yoon et al.	University of Southern California	FSS	SVM	65
Xi-Chen Sun	Peking University	Time-frequency	Clustering network	54
Miharu Nishino et al.	University of Tokyo	Cross-correlation	1-Nearest neighbor	44
Proposed method	Xi'an Jiaotong University	CSP	OELM	92.31

TABLE 7: Classification results comparing the proposed method with other methods.

Contributor	Research lab	Feature extraction	Classifier	Accuracy (%)
Liu Yang et al.	National University of Defense Technology	CSP	LDA	86
Zhou Zongtan et al.	National University of Defense Technology	CSP	LDA	84
Bin An et al.	University of Science and Technology of China	CSP	SVM	48
Proposed method	Xi'an Jiaotong University	CSP	OELM	92.31

4.5. Comparison with Other Classification Methods. In order to evaluate the performance of our whole BCI system, we compare our method with those proposed in [29] in Table 6, which share the same dataset. Table 6 lists several experiment results, their feature extraction methods, and classifiers. The last column shows the classification accuracies corresponding to each system.

Table 6 indicates that our method obtains an accuracy of 92.31%, which is 1.31% higher than Qingguo Wei's method, the best of the others. It can be concluded that our BCI system outperforms those listed with CSP to extract the features and OELM to classify them afterwards. The performance of the proposed algorithm is also compared with several other classification methods listed in [29] in Table 7, which all share the same dataset and adopt CSP as the feature extraction algorithm.

Table 7 reveals that our method obtains an accuracy of 92.31%, which is 6.31% higher than Liu Yang's method, the best of the others. The results prove that OELM shows higher classification accuracy than SVM and LDA under the condition of the same feature extraction method (CSP). We

also evaluate our method in terms of computation time, which includes training and testing time.

The optimal classification accuracy, training time, and testing time corresponding to each experiment that we carried out are shown in Table 8.

As seen from Table 8, CSP combined with OELM achieves the highest accuracy, which is 11.54% higher than ordinary ELM. On the whole, OELM outperforms SVM in both accuracy and speed. Compared with accuracies obtained by SVM and ELM, the accuracy of OELM is more competitive. Defects in the application of SVM can be successfully overcome by OELM for its high accuracy and fast speed. The BCI system can generate better results with OELM than with other state-of-the-art methods when analyzing and processing ECoG signals [30]. As seen from Table 8, ELM is comparable with OELM in speed; however, OELM runs much faster than SVM by a factor up to thousands, whether in training or testing module. Furthermore, OELM can achieve the maximum testing rate 92.31% with 33 nodes, which is significantly higher than all the results so far listed in the ranking of the BCI competition III, using some

TABLE 8: Results of different algorithms.

Algorithms	Classification accuracy (%)	Training time (s)	Testing time (s)
CSP and SVM	64.10	25.5313	10.0781
CSP and ELM	80.77	0.0006	0.0006
CSP and OELM	92.31	<0.0001	<0.0001

popular algorithm such as SVM [31]. It can thus be concluded from the results displayed in these figures and tables that OELM is much more suitable and competitive for binary-class signals of motor imagery.

5. Conclusions

In this paper, a newly intelligent and efficient learning algorithm called Optimized Extreme Learning Machine (OELM) is presented and applied in motor imagery signals classification with CSP to extract features. The proposed method outperforms conventional popular learning algorithms for the extremely fast learning speed and good generalization performance, which is demonstrated with the BCI competition III dataset I. Different sampling points and activation functions are employed in different experiments to analyze the property of OELM. The results show that OELM needs less computational time and obtains better accuracy than SVM and ELM. In conclusion, OELM is a novel and efficient classifier for biometric applications. Although only binary-class classification strategy is discussed in our study, OELM can also be applied to solve multiclassification problem. We believe that this method has great potential for the design of real-time BCI systems.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation (no. 61673316) and project commissioned by Sichuan Gas Turbine Research Institute of AVIC, Major Science & Technology Project of Guangdong Province (no. 2015B010104002), Scientific Research Project of Education Department of Hunan Province (no. 17A148), and Science and Technology Planning Projects of Changde City (no. 2019S019).

References

- [1] Y. Fang, M. Chen, and X. Zheng, "Extracting features from phase space of EEG signals in brain-computer interfaces," *Neurocomputing*, vol. 151, pp. 1477–1485, 2015.
- [2] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks et al., "Brain-computer interface technology: a review of the first international meeting," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 164–173, 2000.
- [3] E. C. Leuthardt, G. Schalk, J. R. Wolpaw, J. G. Ojemann, and D. W. Moran, "A brain-computer interface using electrocorticographic signals in humans," *Journal of Neural Engineering*, vol. 1, no. 2, pp. 63–71, 2004.
- [4] G. Townsend, B. K. LaPallo, C. B. Boulay et al., "A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns," *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1109–1120, 2010.
- [5] R. Ortner, B. Z. Allison, G. Korisek, H. Gaggel, and G. Pfurtscheller, "An SSVEP BCI to control a hand orthosis for persons with tetraplegia," *IEEE Transactions on Neural Systems & Rehabilitation Engineering*, vol. 19, no. 1, pp. 1–5, 2011.
- [6] M. T. Sadiq, X. Yu, Z. Yuan et al., "Motor imagery EEG signals decoding by multivariate empirical wavelet transform-based framework for robust brain-computer interfaces," *IEEE Access*, vol. 7, pp. 171431–171451, 2019.
- [7] M. T. Sadiq, X. Yu, Z. Yuan et al., "Motor imagery EEG signals classification based on mode amplitude and frequency components using empirical wavelet transform," *IEEE Access*, vol. 7, pp. 127678–127692, 2019.
- [8] G.-B. Huang, X. Ding, and H. Zhou, "Optimization method based extreme learning machine for classification," *Neurocomputing*, vol. 74, no. 1–3, pp. 155–163, 2010.
- [9] G. Charvet, F. Sauter-Starace, M. Foerster et al., "WIMAGINE®: 64-channel ECoG recording implant for human applications," in *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2756–2759, Osaka, Japan, July 2013.
- [10] BCI Competition III: 2005," <http://www.bbci.de/competition/iii/>.
- [11] J. Roland, K. Miller, Z. Freudenburg et al., "The effect of age on human motor electrocorticographic signals and implications for brain-computer interface applications," *Journal of Neural Engineering*, vol. 8, no. 4, Article ID 046013, 2011.
- [12] P. S. Hammon and V. R. D. Sa, "Preprocessing and meta-classification for brain-computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 3, pp. 518–525, 2007.
- [13] S. Robinet, P. Audebert, and G. Régis, "A low-power 0.7 32-channel mixed-signal circuit for ECoG recordings," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 1, no. 4, pp. 451–460, 2011.
- [14] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.
- [15] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Transactions on Bio-Medical Engineering*, vol. 55, no. 8, pp. 1991–2000, 2008.
- [16] P. L. Bartlett, "The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 525–536, 1998.
- [17] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, & Cybernetics, Part B Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.
- [18] H.-J. Rong, G.-B. Huang, N. Sundararajan, and P. Saratchandran, "Online sequential fuzzy extreme learning

- machine for function approximation and classification problems,” *IEEE Transactions on Systems, Man, & Cybernetics, Part B Cybernetics*, vol. 39, no. 4, pp. 1067–1072, 2009.
- [19] G.-B. Huang, “An insight into extreme learning machines: random neurons, random features and kernels,” *Cognitive Computation*, vol. 6, no. 3, pp. 376–390, 2014.
- [20] E. Hortal, A. Ubeda, E. Ianez, D. Planelles, and J. M. Azorin, “Online classification of two mental tasks using a SVM-based BCI system,” in *Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering*, pp. 1307–1310, San Diego, CA, USA, November 2013.
- [21] V. P. Oikonomou, G. Liaros, K. Georgiadis et al., “Comparative evaluation of state-of-the-art algorithms for SSVEP-based BCIs,” 2016, <http://arxiv.org/abs/1602.00904>.
- [22] L. Duan, H. Zhong, J. Miao, Z. Yang, W. Ma, and X. Zhang, “A voting optimized strategy based on ELM for improving classification of motor imagery BCI data,” *Cognitive Computation*, vol. 6, no. 3, pp. 477–483, 2014.
- [23] T. N. Lal, T. Hinterberger, W. Guido et al., “Methods towards invasive human brain computer interfaces,” in *Proceedings of the Advances in Neural Information Processing Systems 17 (NIPS)*, Vancouver, Canada, December 2004.
- [24] E. C. Leuthardt, Z. Freudenberg, D. Bundy, and J. Roland, “Microscale recording from human motor cortex: implications for minimally invasive electrocorticographic brain-computer interfaces,” *Neurosurgical Focus*, vol. 27, no. 1, p. E10, 2009.
- [25] Z. Yu, G. Zhou, J. Jing, X. Wang, and A. Cichocki, “Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface,” *Journal of Neuroscience Methods*, vol. 255, pp. 85–91, 2015.
- [26] Y. Miche, A. Sorjamaa, P. Bas, O. Simula, C. Jutten, and A. Lendasse, “OP-ELM: Optimally pruned extreme learning machine,” *IEEE Transactions on Neural Networks*, vol. 21, no. 1, pp. 158–162, 2010.
- [27] G.-B. Huang, L. Chen, and C.-K. Siew, “Universal approximation using incremental constructive feedforward networks with random hidden nodes,” *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879–892, 2006.
- [28] X. Xu, L. Lu, X. Zhang, H. Lu, and W. Deng, “Multispectral palmprint recognition using multiclass projection extreme learning machine and digital shearlet transform,” *Neural Computing and Applications*, vol. 27, no. 1, pp. 143–153, 2014.
- [29] <http://www.bbci.de/competition/iii/results/index.html>.
- [30] A. Bamdadian, C. Guan, K. K. Ang, and J. Xu, “Improving session-to-session transfer performance of motor imagery-based BCI using adaptive extreme learning machine,” in *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, pp. 2188–2191, Osaka, Japan, July 2013.
- [31] B. Blankertz, K. R. Müller, D. J. Krusienski et al., “The BCI competition III: validating alternative approaches to actual BCI problems,” *IEEE Transactions on Neural Systems & Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.