IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 28, 2019, accepted January 17, 2020, date of publication January 23, 2020, date of current version February 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2969054

A Motor Imagery EEG Feature Extraction Method Based on Energy Principal Component Analysis and Deep Belief Networks

LIWEI CHENG[1], DUANLING LI[1,2], GONGJING YU[3], ZHONGHAI ZHANG[3], XIANG LI[3], AND SHUYUE YU[3]

[1]School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]College of Mechanical and Electrical Engineering, Shaanxi University of Science and Technology, Xi'an 712000, China
[3]Beijing Aerospace Measurement and Control Technology Company, Ltd., Beijing 100041, China

Corresponding authors: Liwei Cheng (liwei_cheng89@163.com) and Duanling Li (duanlingli@bupt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51775052, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2019JM-181, and in part by the Beijing Key Laboratory of Space-Ground Interconnection and Convergence.

ABSTRACT The motor imagery electroencephalography (MI-EEG) reflects the subjective motor intention, which has received increasing attention in rehabilitation. How to extract the features of MI-EEG accurately and quickly is the key to its successful application. Based on the analysis and comparison of the existing feature extraction algorithms, a feature extraction method based on principal component analysis (PCA) and deep belief networks (DBN) is proposed, namely PCA-DBN. Firstly, the second-order moment is used to analyze the time-domain of MI-EEG, select the effective time interval. Secondly, PCA is used to analyze the selected time-domain interval and obtain the principal component feature points. Then, feature points are imported into DBN to realize the final feature extraction. Finally, use the softmax classifier to complete task classification. Perform algorithm validation on the BCI Competition II Data set III and BCI Competition IV Data sets 2b, classification accuracies are 96.25% and 91.71%, kappa values are 0.925 and 0.8342. The paired-sample t-test with FDR correction is carried out on the verification results, and the comparison with some better classification algorithms shows that the algorithm has better performance. In the end, this method is used to extract the features of laboratory data, the optimal classification accuracy is 97.69% and kappa value is 0.9538, the validity of the method is further verified.

INDEX TERMS Deep belief networks, motor imagery electroencephalogram, principal component analysis, second-order moment, softmax classifier.

I. INTRODUCTION

Brain-computer interface [1] (BCI) technology is a new human-computer interaction mode that directly connects the human brain to external devices without peripheral nerves and muscle tissues. This special way of interaction makes it have a great application prospect in the fields of motor function rehabilitation, information communication, and equipment control. As a new interdisciplinary subject, BCI has become a hot spot of interdisciplinary research in biomedicine, communication engineering, computer technology, etc. [2]. Motor imagery electroencephalography (MI-EEG) is an important branch of BCI research. Because the MI-EEG itself is a weak signal of non-stationary, non-periodic, and time-varying complex. To quickly and effectively identify the components of different consciousness activities, feature extraction and recognition classification play a very important role, which directly affects the accuracy of EEG signal task classification.

At present, there are several methods of MI-EEG feature extraction: autoregression model (AR) [3], wavelet packet transformation (WPT) [4], discrete wavelet transform (DWT) [5], common spatial patterns (CSP) [6], power spectral density (PSD) [7], and principal component analysis (PCA) [8]. AR, the parameters change with the input of each sample point, better reflecting the state of the brain, and it requires less computation and does not require prior knowledge of the relevant frequency band, but it is not

The associate editor coordinating the review of this manuscript and approving it for publication was Benyun Shi.

good for nonstationary signals. As a time-frequency analysis method, WT is suitable for non-stationary signal processing. WPT can decompose the low and high frequency information at the same time, DWT can extract more abundant features. But both of them need too much computational information when the signal quantity is large, and the extracted characteristic information has a large redundancy. CSP algorithm is based on the simultaneous diagonalization of two covariance matrices to design the optimal spatial filter, there is no need to extract the characteristics of the selected frequency band, but it needs multi-channel analysis and is easily disturbed by noise. PSD estimates the power spectral density of the EEG signal by using a finite length signal, which can reflect the energy change of the EEG signal, but the signal time information will be lost. When the EEG data length is short, the signal statistical characteristics obtained by this method are not obvious. PCA is an effective method of dimension reduction, the feature dimension reduction of EEG signals can be achieved with shorter training time and lower algorithm complexity. But some of the extracted features are still redundant information for classification, this part of redundant information needs to be removed to improve classification accuracy.

In this situation, it is an effective method to extract second features from redundant information. Among them, the use of a deep learning model can eliminate redundant information, retain better features and ultimately achieve higher classification accuracy. Common deep learning models are convolutional autoencoder (CAE) [9], convolutional neural network (CNN) [10], neural network (NN) [11], sparse autoencoder (SAE) [12], deep belief networks (DBN) [13]. Among them, DBN has been well studied in redundant information elimination and feature extraction and recognition. Weilong Zheng et al. introduced the differential entropy characteristics of multi-channel EEG into DBN as input, higher classification accuracy was achieved [14]. Li *et al.* proposed an emotional state recognition model based on DBN for EEG signals, this model can process the signals of each EEG channel and effectively extract key information from thousands of features [15]. Li *et al.* applied DBN to EEG fatigue detection, through the training of the initial feature, the model to better identify the performance was obtained [16]. Wulsin *et al.* used DBN to detect the characteristics of abnormal EEG signals and got good results [17]. Altan *et al.* proposed the classification of brain activity based on DBN, the difference between positive and negative tasks in patients with cerebral apoplexy was highly accurate [18].

In some studies where the experimental data is BCI Competition II Data set III. Minyou Chen et al. proposed a phase space feature extraction method (AFAPS). This method revealed various information in the original sequence through phase space reconstruction (PSR), then the amplitude-frequency analysis (AFA) method was used to extract the signal characteristics in the phase space. The extracted features not only maintained the continuity of the original information but also contained the nonlinear information in the original information very well [19]. Wanzhong Chen et al. proposed a motor imagery recognition method based on masking empirical modal decomposition (MEMD), the windowed EEG was decomposed by MEMD and hybrid features were then extracted from the first two intrinsic mode functions (IMFs). A good classification result was obtained by importing the feature results in the linear discriminant analysis (LDA) [20]. Önder Aydemir et al. proposed an optimum combination of sub-band power features a method for improving the classification accuracy of motor imagery electroencephalogram. The sub-band power features were extracted from the best time segment of electroencephalogram trials and the proposed training model determined the optimum combination of sub-bands. Better classification accuracy was obtained [21]. Yan Wang et al. proposed a feature extraction method combined spectrum analysis with wavelet packet analysis. The linear discriminant classifier, Self-Organizing Feature Map (SOFM) neural network, and Back Propagation (BP) neural network were compared, and the result that BP neural network had a higher recognition rate [22]. Rajdeep Chatterjee et al. proposed a fuzzified adaptation of discernibility matrix with four variants of dissimilarity measures to deal with continuous data. Then support vector machine (SVM) was used to classify the simplified data set. Compared with the original difference matrix method and the PCA method, the performance of this method was improved greatly [23]. Nguyen *et al.* proposed a multi-sphere approach to SVDD to have a better description for the brain data, this method was a fuzzy clustering approach to optimize SVDD parameters, achieved a better description of the MI-EEG data [24]. Steven Lemm et al. proposed an online classification algorithm for a single experiment based on random wavelet filtering, which had adapted to individual EEG spectra, and reduced the error rate of signal recognition and classification [25], [26]. Kübra Saka et al. proposed a novel Fast Walsh Hadamard Transform based feature extraction method for classification of EEG signals recorded during left/right hand movement imagery. It did not only provided well-discriminative attributes but also the computational time of extracting the features from a single EEG trial was fast [27]. Jie Zhou et al. proposed a method of EEG signal classification based on wavelet envelope analysis and long-short time memory classifier, this method combined Hilbert transform (HT) and DWT, the significant features of the signal were extracted. Then entered into the LSTM classifier for classification [28]. Guo and Wu proposed a dynamic ICA base on sliding window Infomax algorithm to analyze motor imagery EEG, by used the feature patterns based on total energy of dynamic mixing matrix coefficients in a certain time window, better classification accuracy was achieved [29].

In some studies where the experimental data is BCI Competition IV dataset 2b. Gaowei Xu et al. proposed a CNN framework based on VGG-16, this framework consisted of a VGG-16 CNN model pre-trained on ImageNet and a target CNN model with the same structure as VGG-16 except

for the softmax output layer. The accuracy and efficiency of EEG signal classification were improved [30]. Ha and Jeong proposed a novel method based on CapsNet to classify two-class motor imagery signals, this method used short-time Fourier transform (STFT) to transform EEG signals into time-frequency image, then the capsule network was introduced for training and testing. Better and more reliable performance than the traditional CNN method was obtained [31]. Wang *et al.* proposed a method designated frequency domain CSP (FDCSP), this method used the fast Fourier transform (FFT) algorithm to obtain uniform frequency domain samples in the range of 8 to 30Hz as the input signal of CSP, the classification performance was improved [32]. Hai-Jun Rong et al. proposed an incremental adaptive EEG classification method. In this method, an Extended sequential adaptive fuzzy inference system (ESAFIS) was used to evolve its structure dynamically and adapted the classifier automatically online to address the non-stationarity of the EEG signals. A better classification effect was achieved [33]. Bentlemsan *et al.* proposed a method to classify the EEG signal by FBCSP and RF, this method improved the classification accuracy effectively [34]. Bustios and Rosa proposed a novel method to select these subject-specific frequency bands in a new configuration for the Filter Bank CSP approach, this method used exhaustive search to find the best subset of the frequency band that contained the most discriminating patterns, the data sets were classified more accurately [35]. Muthong *et al.* proposed a method to captured all the change points in EEG signals and divided them into a group of signals, the regularized LDA was then applied to each partition signal, the classification accuracy of EEG signals was higher [36]. Tian and Liu proposed a classification method of two-class MI-EEG signals based on CNN, the prediction accuracy of left/right hand motor imagery task was greatly improved [37]. Jun Cai et al. proposed a new model called Convolutional Gated Recurrent Neural Network that combined CNN and Gated Recurrent Unit (GRU). This model extracted the combinatorial features of the preprocessed MI-EEG by CNN, it enriched the GRU input, and then it used GRU to extract some sequence information hidden in the EEG signals to improve the classification accuracy of MI-EEG [38]. Jin *et al.* proposed a sparse Bayesian ELM (SBELM) algorithm based on probabilistic reasoning, combined the advantages of ELM and sparse Bayesian learning, this algorithm could automatically control the complexity of the model, eliminated redundant hidden neurons, and achieved better classification accuracy in the verification of common set data [39]. Zhang *et al.* proposed an EEG classification method based on multi-kernel ELM (MKELM), which could effectively mine supplementary information in multi-nonlinear feature space and provided stronger robustness for the classification of EEG signals [40]. Jiao *et al.* proposed a new sparse group representation model (SGRM), which used intersubject information to improve the efficiency of MI based BCI and effectively reduced the training samples required by target objects [41]. Zhang *et al.* proposed a temporally constrained sparse group spatial pattern (TSGSP) to optimize both the filter band and the time window in the CSP to further improve the classification accuracy of MI-EEG [42].

In this paper, a method of MI-EEG feature extraction based on PCA and DBN is proposed. The remaining sections are organized as follows: Section 2 introduces the basic theory used in this paper, including second-order moment, PCA, DBN, and softmax classifier. Section 3 introduces the public data sets used in this article, the performance evaluation index, the data analysis method, and the algorithm flow of PCA-DBN. Section 4 introduces the comparison and analysis of the experimental verification results of different data sets. Section 5 introduces the algorithm in this paper and compared it with the previous research results, and the potential limitations and future research directions of this method are discussed. Section 6 summarizes the whole thesis and gives the conclusion.

II. PRIMARY METHODS

A. ONLINE RECURSIVE ESTIMATION OF THE NORMALIZED SECOND-ORDER MOMENT

For a random signal with zero average value $x_i^j(t)$, the normalized second-order moment is defined as:

$$m_2 = E_i^j[x^2(t)] = [x_i^j(t)]^2 \tag{1}$$

For a signal $x_i^j(t)$ of length N, the second-order moment defined by formula (1) is estimated by the following formula [43]:

$$m_2 = E_i^j[x^2(t)] \approx \frac{1}{N}\sum_{n=1}^{N}[x_i^j(n)]^2 \tag{2}$$

In the actual processing of MI-EEG, new samples are constantly collected and input. Therefore, in order to reflect the dynamic change of signal statistical characteristics in real-time, a second-order moment recursive algorithm is established. There are two second-order moment estimation algorithms, one is based on sliding window length, the other is based on variable window length.

Second-order moment estimation of sliding window: suppose the window length is N, the current moment is n, the data in the window is $x(n-N+1), \ldots, x(n-1), x(n)$, where the second-order moment at a time n is $m_2(n)$, and the newly arrived signal sample at a time $n+1$ is $x(n+1)$. $x(n-N+2), \ldots, x(n), x(n+1)$ is the data in the window at this time, easy to figure out:

$$m_2(n+1) = m_2(n) - \frac{1}{N}\{x^2(n-N) - x^2(n+1)\} \tag{3}$$

Estimation of second-order moment cumulant with variable window length: if the left end of the window is fixed, suppose the first sample point $x(1)$ of the data as the starting point, the second-order moment is updated with the coming of data, another recursion of second-order moments can be

obtained, the formula is:

$$m_2(n+1) = \frac{n}{n+1} m_2(n) + \frac{x^2(n+1)}{n+1} \tag{4}$$

In practical application, the two algorithms can be combined. At the beginning of data receiving, the data length cannot reach the specified length, and formula (4) can be used for online recursive estimation of the second-order moment. When the data length reaches the window length, the online recursion method of the second-order moment in formula (3) is used. It is worth noting that the above normalized second-moment recurrence formula is established on the premise that the average value of the input signal must be zero. Considering that the EEG data has been processed by band-pass filtering without the dc component, the recurrence formula is valid [44].

B. PRINCIPAL COMPONENT ANALYSIS

PCA is based on the variance maximization principle, representing the rows (or columns) of the original data matrix with a new set of linearly independent and orthogonal vectors, to compress the number of variables, eliminate redundant information and maximize effective information. The PCA is solved as follows [45]:

(1) Average value processing of the original data $x_i^j(t)$, and then take its covariance matrix C_{p*p}:

$$cov(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}),\ 1 \le x \le p,\ 1 \le y \le p \tag{5}$$

where n is the number of samples, p is the number of features, x and y is a certain feature.

(2) The characteristic root decomposition of the covariance matrix C is carried out, get characteristic root Λ_{p*p}, and the eigenvector U_{p*p}, the eigenvector is used as the coordinate axis of the main component to form the new vector space, the size of each characteristic root represents the information contained in each principal component.

$$C_{p*p} = U_{p*p} \Lambda_{p*p} U_{p*p} \tag{6}$$

(3) Find the projection F_{n*p} of the original data $x_i^j(t)$ in the new coordinate system:

$$F_{n*p} = X_{n*p} U_{n*p} \tag{7}$$

(4) Accumulation contribution. The characteristic root size of each main component represents how much information is contained, the cumulative contribution rate of the former k main component:

$$pre_k = \sum_{i=1}^{k} \lambda_i \Big/ \sum_{i=1}^{p} \lambda_i \tag{8}$$

λ_i is the *ith* characteristic root of the solution.

(5) Select the appropriate cumulative contribution *pre*, make the previous d main component F_{n*d} as a new data, alternative raw data $x_i^j(t)$ for pattern classification (In general: $d < p$).

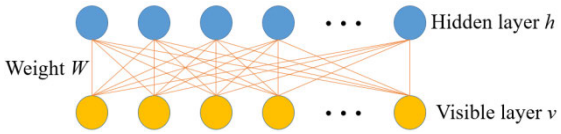


FIGURE 1. RBM structural representation.

C. DBN ALGORITHM

1) RBM MODEL

Restricted Boltzmann Machine (RBM) is an energy-based neural network model, and its undirected graph model is shown in Fig. 1. Where h is the hidden layer unit and v is the visible layer unit, w is the connection weight matrix between the hidden layer and the visible layer [46].

The energy of the joint configuration is:

$$E(v, h/\theta) = -\sum_{i=0}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j - \sum_{i=1}^{n} \sum_{j=1}^{m} v_i w_{ij} h_j \tag{9}$$

In the formula: m and n are respectively the number of neurons in the hidden layer and the visible layer, v and h are respectively the number of neurons in the visible layer and the hidden layer, v_i represents the state of the i visible neuron in the visible layer, and h_i represents the state of the j hidden unit. $\theta = \{w_{ij}; a_i; b_j\}$ is a parameter of RBM, they're all real numbers. w_{ij} represents the connection weight matrix between the visible element i and the hidden element j, a_i represents the bias of the visible element i, b_j represents the bias of the hidden element j.

2) GRBM MODEL

RBM does not work well with continuously valued data, but the Gaussian restricted Boltzmann machine (GRBM) model can well model the continuous values [47]. Both hidden and visible nodes in the RBM undirected graph model are fully connected, nodes within layers are independent of each other. The joint distribution described by the RBM graph model can be decomposed into the following form:

$$P(v, h) = P(h)P(v|h) \tag{10}$$

In the formula: $P(h)$ can be seen as the Gaussian mixing coefficient, if $P(v|h)$ is the Gaussian distribution, RBM under this condition is equivalent to the Gaussian mixture distribution. To assume that $P(v|h)$ is the Gaussian, the energy of GRBM is defined as follows:

$$E(v, h) = -\sum_{i=vis} \frac{(v_i - b_i)^2}{2\sigma_i^2} - \sum_{j=hid} b_j h_j - \sum_{i,j} \frac{v_i}{\sigma_i} h_j w_{ij} \tag{11}$$

Parameter requirements are: $\theta = \{W, b_i, b_j, \sigma_i, \sigma_j\}$, the learning between parameters is also realized by the contrastive divergence (CD) algorithm.

3) THE STRUCTURE MODEL AND TRAINING PROCESS OF DBN

Because of the complexity of the signal, single-layer RBM cannot extract the high-dimensional features of the signal.

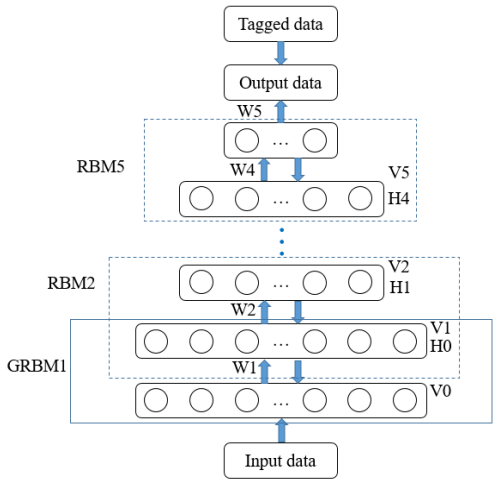


FIGURE 2. DBN structural representation.

DBN is composed of multiple RBM networks, each of which can act as a feature extractor. Therefore, the DBN model is formed by stacking multiple RBM to extract high-level features. The structure of the DBN is shown in Fig. 2.

The training process of the DBN model is divided into two steps:

Step 1. Use unsupervised learning to train each layer of RBM, the output of each RBM can be used as input to the next layer of RBM.

Step 2. Use a supervised way to train the entire network and fine-tune the entire network.

D. SOFTMAX CLASSIFIER

Softmax classifier is an extension of the logistic regression model, often used in conjunction with DBN [48]. The sample of m training sets is $\{(x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)})\}$, the label is $y^{(i)} \in \{1, 2, 3, \ldots, k\}$, probability $P(y = j|x)$ means the input is x, the probability that the sample is defined as j, the one has the highest probability is defined as the one. That is, for a K class classifier, the output is a vector of k dimensions (the sum of the elements of vectors is 1), the output is:

$$h_{\theta(x)} = p(y^{(i)} = k|x^{(i)}; \theta) = \frac{\exp(\theta_k^T x(i))}{\sum_{j=1}^{k} \exp(\theta_j^T x(i))} \quad (12)$$

In the formula: θ is the model parameter, $k = 1, 2, \ldots, K$, it is obtained by minimizing the cost function $J(\theta)$ shown in formula (13):

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m}\sum_{j=0}^{l} 1\{y^{(i)} = j\} \log p(y^{(i)} = 1|x^{(i)}; \theta)] \quad (13)$$

In the formula:

$$p(y^{(i)} = j|x^{(i)}; \theta) = \frac{\exp(\theta_j^T x(i))}{\sum_{j=1}^{k} \exp(\theta_j^T x(i))} \quad (14)$$

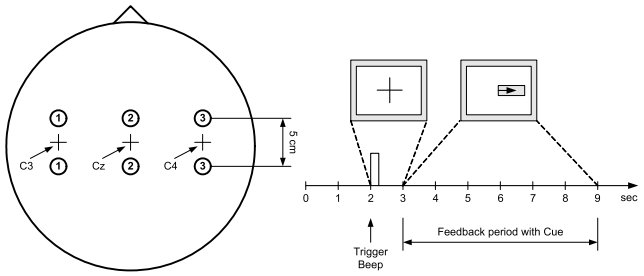


FIGURE 3. Electrode positions (left) and timing scheme (right).

By adding weight attenuating term to the cost function, penalize parameters that have too much weight, and make the parameters converge to the optimal.

III. MATERIALS AND ALGORITHM FLOW

A. DATASET

1) BCI COMPETITION II DATA SET III

Dataset 1 is from BCI Competition II Data set III. This is an open dataset for BCI Competition and provided by the Department of Medical Informatics, Institute for Biomedical Engineering, Graz University of Technology. In the experiment, a female subject (25y) controlled a feedback bar by imagining the movement of her left and right hands. The experiment consisted of 7 runs with 40 trials each, a total of 280 trials. The training group and the testing group each had 140 trials. The samples were collected every 9s, when t = 0∼2s, the subject was in a ready state and did not make any movement. Started from t = 2s, voice prompt; during the period of t = 3∼9s, subjects exercised imagination. Data were collected from the 10∼20 pilot system of the international standard, and three bipolar EEG channels (anterior '+', posterior '−') were measured over C3, Cz, and C4. The sampling frequency was 128Hz, which was filtered by a 0.5∼30Hz bandpass filter. Fig. 3 is electrode positions and timing scheme [49]–[52]. For more information on the dataset, please refer to the website http://bbci.de/competition/ii/.

2) BCI COMPETITION IV DATA SETS 2b

Dataset 2 is from BCI Competition IV Data sets 2b. This is an open dataset for BCI Competition and provided by the Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology. The data were EEG data from the left/right hand motor imagery of 9 subjects. There were five groups of data, the first three sessions (01∼03T) were training data and the last two sessions (04∼05E) were test data. The first two sessions (01∼02T) contained 120 trails per session without feedback, and the last three sessions (03T, 04∼05E) contained 160 trails per session with smiley feedback. Data were collected from the 10∼20 pilot system of the international standard, and three bipolar EEG channels were measured over C3, Cz, and C4. The sampling frequency was 250Hz and filtered by a 0.5∼100Hz bandpass filter, and 50Hz power frequency

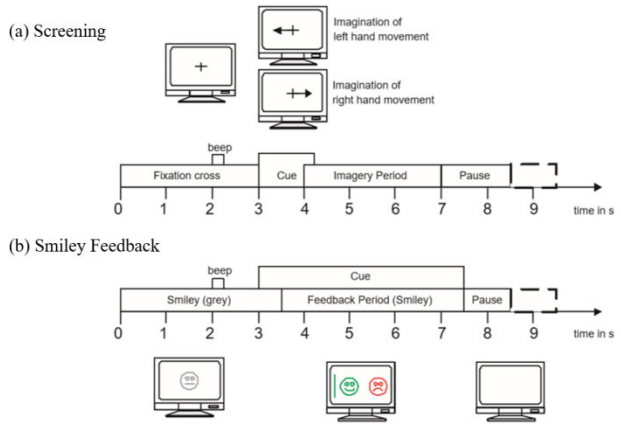


FIGURE 4. Timing scheme of the paradigm. (a) The first two sessions, (b) the last three sessions (BCI Competition IV Data sets 2b).

was eliminated. In the first two sessions. When t = 0∼3s, the subject was in a ready state, and a brief prompt (1kHz, 70ms) on 2s. Visualize left/right hand movement cues at random in 3∼4.25s. During 4∼7s, the subjects performed motor imagery Each experiment was followed by a short break of 1.5∼2.5s. In the last three sessions. When t = 0∼3s, the subject was in a ready state, and a brief prompt (1kHz, 70ms) on 2s. Visual cue time was followed within 3∼7.5s, the feedback period was within 3.5∼7.5s. Subjects imagined moving their left or right hands, the smiley face turned green when it moved correctly and red when it moved incorrectly. Each experiment was followed by a short break of 1.5∼2.5s. The timing scheme of the paradigm is shown in Fig. 4 [53]–[56]. For more information on the dataset, please refer to the website http://bbci.de/competition/iv/desc_2b.pdf.

B. PERFORMANCE EVALUATION INDEX

Different statistical performance indexes are used to objectively compare the performance of different algorithms in feature extraction, in this study, the accuracy of recognition classification after feature extraction and the length of training/test time are used as the performance evaluation index.

In the accuracy of recognition and classification, the first standard is the accuracy of classification, which is used to directly measure the accuracy of the classification of signals after feature extraction.

The second standard is the kappa value [57], which is also an index used to measure classification accuracy and is often used for the consistency test of EEG signal classification. The formula is:

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \tag{15}$$

P_0 is the total number of samples (correct classification) divided by the total number of samples, which is the classification accuracy. Suppose the actual number of samples for each category is $a_1, a_2, \ldots, a_c$, The predicted sample number of each category is $b_1, b_2, \ldots, b_c$, the total number of samples is n, there is:

$$P_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \cdots + a_c \times b_c}{n \times n} \tag{16}$$

In the length of the training/test time, the time consumption of the algorithm in the training/test process is calculated separately, which can directly reflect the time efficiency of the algorithm.

C. DATA ANALYSIS METHOD

In this study, a paired-sample t-test with FDR correction is used to analyze the experimental data.

1) PAIRED-SAMPLES T-TEST

It is used to test whether the data of two sets of correlated samples are derived from a normally distributed population with the same mean, that is to infer whether there is a significant difference in the mean values of two related sample populations. The original hypothesis is $H_0 : \mu_1 - \mu_2 = 0$, where μ_1 and μ_2 are the mean of the first population and the second population, respectively. Assume that x_{1i}, x_{2i} $(i = 1, \ldots, n)$ are paired samples, respectively, the sample difference is $d_i = x_{1i} - x_{2i}$, at this point, the test statistic is:

$$t = \frac{\overline{d} - (\mu_1 - \mu_2)}{s_d / \sqrt{n}} \tag{17}$$

where $\overline{d} = \frac{\sum_{i=1}^{n} d_i}{n}$ is the average of the difference between the paired samples, $s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \overline{d})^2}{n-1}}$ is the standard deviation of the paired sample difference, n is the number of paired samples. When $\mu_1 - \mu_2 = 0$, the t statistic obeys the t distribution of degree of freedom $n - 1$.

2) P-VALUE

In hypothesis testing, when the original hypothesis (H_0) is true, the probability of the resulting sample observations or more extreme results. If the p-value is small, it means that if the original hypothesis is true, the probability of this data being even more extreme is small. However, when the p-value is less than a preset value α (usually 0.05 in biological analysis), it is more reasonable to reject the original hypothesis (H_0) rather than believe in the occurrence of such a small probability.

3) FDR CORRECTION

The expected value of the proportion of false rejects (Reject the true (original) hypothesis) to all rejected original hypotheses. The algorithm steps are:

I. Enter the p-value obtained by the paired-sample t-test (suppose the number of p-values is m), sort all p-values, that is $p(1) \leq p(2) \leq \cdots \leq p(m)$, k is the ranking corresponding to the p-value of one of the test results.

II. Calculate in turn and find the maximum value of k that conforms to the original threshold value α, that's the maximum value of k that $p(k) \leq \alpha \times k/m$.

III. There are significant differences in all tests from 1 to k, calculate the corresponding q-value of each p-value, that is $q = p \times (m/k)$.

FDR can be controlled at $q(0 \leq q \leq 1)$ level by this algorithm. At this point, a new significance level value will be obtained. If the p-value is less than this new significance level value after the statistical test, the result will be considered statistically significant.

D. THE ALGORITHM FLOW BASED ON PCA-DBN

An MI-EEG feature extraction method is proposed by combining PCA with DBN, the specific steps of the algorithm are as follows:

Hypothesis MI-EEG is Expressed as:

$$x_i^j(t) = [x_1^1(t), x_2^2(t), \ldots, x_n^m(t)] \in R^{N \times n \times m} \tag{18}$$

where N is the total number of sample points, n is the number of EEG channels, m is the number of sampling points, $x_i^j(t)$ (the *jth* sampling point of *ith* channel) is a filtered MI-EEG signal:

$$t = \{1, 2, \ldots, N\},\ i = \{1, 2, \ldots, n\},\ j = \{1, 2, \ldots, m\} \tag{19}$$

(1) The time domain characteristics of MI-EEG are analyzed by the second-order moment method

The MI-EEG signal is collected through the electrode cap and stored in the form of voltage amplitude. Therefore, formula (20) is used to calculate the instantaneous energy.

$$E_i^j[x^2(t)] = [x_i^j(t)]^2 \tag{20}$$

In the formula: $E_i^j[x^2(t)]$ express the *jth* sampling point of the *ith* channel of the *tth* sample, point of MI-EEG signal of transient energy.

Suppose E_i^j is the average energy of the MI-EEG in the *jth* sampling point of the *ith* channel in the N experiments, and is expressed as:

$$E_i^j = \frac{1}{N} \sum_{n=1}^{N} [E_i^j(n)] \tag{21}$$

According to formula (21), the average energy of each lead MI-EEG signal is calculated, and the MI-EEG signals with distinct time periods are selected for feature extraction.

(2) The selected signals are analyzed by principal component analysis

For the MI-EEG signal of an obvious time period selected in step (1) $x_i^j(t)(i = \{1, 2, \ldots, n\}, j = \{1, 2, \ldots, m\})$ are analyzed by principal component analysis, compress the number of data variables and remove redundant features, to maximize the availability of information.

I. $x_i^j(t)$ is processed by means of mean processing. Then, the covariance matrix C_{p*p} is obtained through formula (5).

II. The eigenvalue decomposition C_{p*p} is carried out through formula (6), get the eigenvalue Λ_{p*p} and eigenvectors U_{p*p}.

III. The projection of $x_i^j(t)$ onto the new coordinate F_{n*p} is obtained by formula (7).

IV. Through formula (8), the cumulative contribution rate pre_k of the first k principal components can be obtained.

V. Select the appropriate *pre* and use F_{n*d} instead of $x_i^j(t)$ for pattern classification.

(3) DBN network is used for feature extraction of F_{n*d}.

Firstly, take F_{n*d} as the input to the DBN network. Then, unsupervised initialization of all RBMs in DBN from the bottom up [14].

I. Calculate $P(h_1|v)$ based on the visual layer v, so the hidden layer h_1 is $P(h_1|v)$.

II. Know the hidden layer h_1, calculate F_{n*d} for the first RBM visual layer v, that is v', and $v' = p(v'|h_1')$.

III. Repeat calculation step I and II, and use Gibbs sampling calculation v^n and h_1^n, then the weight W_1 is updated according to the CD algorithm.

IV. Repeat step I∼III, up to the maximum number of iterations, the first RBM pretraining is over.

V. Repeat step I∼IV, train the other RBMs in turn, get the weight value $W_2, W_3, \ldots, W_n$, n is the number of RBMs in DBN.

(4) The output error C of the DBN network is calculated, use the back-propagation (BP) algorithm to complete fine-tuning of weights of the entire network, implement DBN supervised training. Where, the gradient calculation formula is $\frac{\partial C}{\partial \theta}(\theta = (W, b))$, $\theta = \theta - \varepsilon \frac{\partial C}{\partial \theta}$, the weight matrix of each layer is updated. The DBN network output is the final MI-EEG feature.

IV. RESULTS AND ANALYSIS

A. EXPERIMENTAL RESULTS AND ANALYSIS (EXPERIMENTAL DATA SET: BCI COMPETITION II DATA SET III)

Experimental verification methods adopt in this study are as follows:

I. In data selection and partitioning: The 10-fold cross-validation method was used to classify the experimental data [58]. The original train sets were divided into new train sets and validation sets for classifier training, then the trained classifier was used to classify the test sets. Finally, the average value of classification results was taken as the final classification accuracy.

II. In data processing: Firstly, we selected the time period suitable for feature extraction through time-domain analysis. Secondly, we extracted eigenvectors through the PCA feature extraction algorithm. Thirdly, the extracted eigenvalues are imported into the DBN network for the second extraction of eigenvalues. Finally, we imported the final eigenvalues into the softmax classifier to train the appropriate classifier, implement the classification of tasks.

III. In the comparison and analysis of data results: Classification accuracy, kappa value, and training/test time were used as evaluation criteria.

IV. In the comparison validation method: We proposed two different ways of contrast. One is the comparative analysis of six traditional feature extraction algorithms: AR, WPT, DWT,

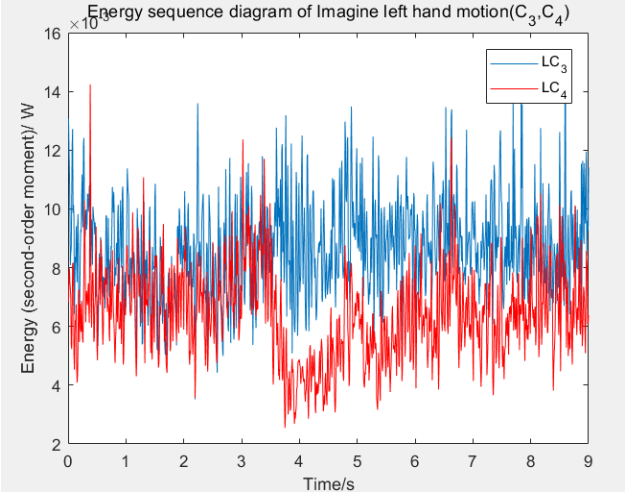


FIGURE 5. Energy sequence diagram of imagining left hand motion (C_3, C_4).

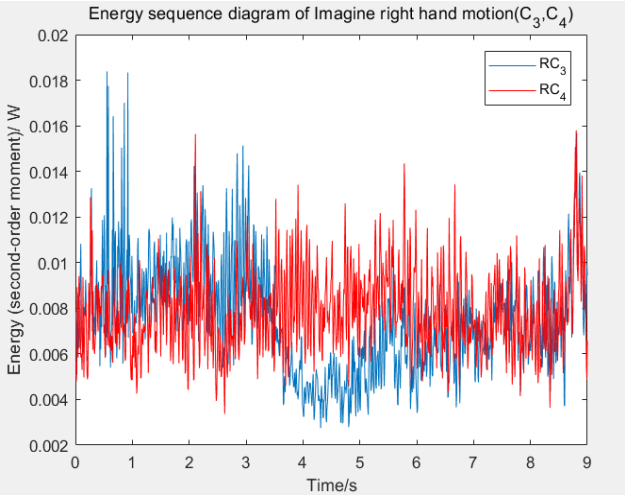


FIGURE 6. Energy sequence diagram of imagining right hand motion (C_3, C_4).

CSP, PSD, PCA. The other is the comparative analysis after the cross combination of six traditional feature extraction algorithms (AR, WPT, DWT, CSP, PSD, PCA) and five deep learning algorithms (CAE, CNN, NN, SAE, DBN). Finally, the validity of the proposed algorithm is verified.

1) TIME-DOMAIN ANALYSIS OF MI-EEG

According to formulas (3) and (4), calculate when $N = 140$, the average energy corresponding to each sampling point of C_3 and C_4 leads within 0∼9s when imagining the motion of the left hand, namely the LC_3/LC_4. And the average energy corresponding to each sampling point of C_3 and C_4 leads within 0∼9s when imagining the motion of the right hand, namely the RC_3/RC_4. As shown in Fig. 5 and Fig. 6, the average power of the left/right movement in the 3∼9s period is clearly different. Therefore, this paper selects the brain electrical signals in the 3∼9s section to extract the features.

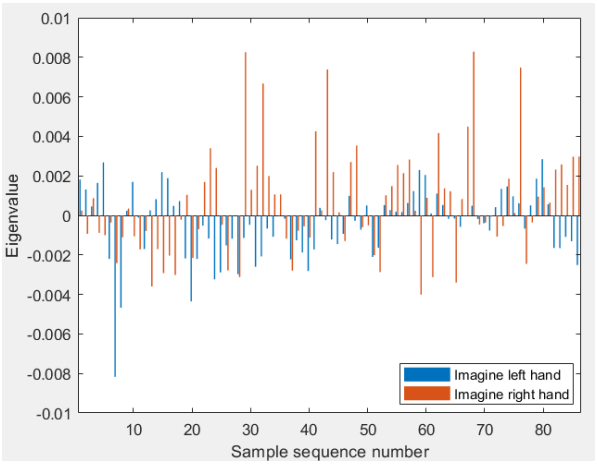


FIGURE 7. Principal component mean distribution map.

2) PCA FEATURE SELECTION

Through the time-domain analysis of MI-EEG mentioned above, 768 sampling points in the range of 3∼9s were selected, PCA was applied to 140 groups of training data. For 140 groups of training data, after PCA analyzed, the average value of 86 ($d_1, d_2, \ldots, d_{86}$) principal component samples of the two types of samples were obtained, as shown in Fig. 7. When the accumulative contribution rate is more than 90%, the principal component is retained, 18 principal components far apart from each other in Fig. 7 are selected as new feature points:

$$d_5, d_{13}, d_{15}, d_{23}, d_{24}, d_{29}, d_{32}, d_{41}, d_{48},$$
$$d_{59}, d_{61}, d_{62}, d_{67}, d_{68}, d_{76}, d_{82}, d_{83}, d_{86}$$

Get an 18-dimensional eigenvector. This eigenvector is the value imported into the DBN network.

3) DETERMINATION OF DBN PARAMETERS

Parameters of DBN have a great influence on the feature extraction and recognition results of signals. In this paper, the optimal DBN parameters are determined through experience and the test of the experimental data set (BCI Competition II Data set III).

The main parameters of DBN are network layer number, iteration times, learning rate, epochs, etc. The number of DBN layers is generally set as 3∼10 layers, and a softmax classifier is adopted. The DBN network parameters such as learning rate, iteration times and momentum are set as shown in Table 1, the experimental results are shown in Table 2 [59], [60].

Table 2 shows that, with the increase of DBN layers, the capacity of the model to accommodate information will be enhanced, and the classification effect will show an upward trend. When the number of network layers increases to 6, the classification accuracy reaches 95.01%. With the further increase in the number of network layers, the classification accuracy begins to decline. This may be due to the network underfitting when the training data amount remains

TABLE 1. DBN main parameter settings.

Parameters	Training process	
	Unsupervised	Supervised
Learning rate	0.01	0.01
Iterations	150	100
Momentum	0.5	0.1
Cost function weights	0.002	0.002

TABLE 2. Classification accuracy of different DBN layers.

DBN layers	3	4	5	6
Accuracy (%)	84.36	87.89	90.24	95.01
DBN layers	7	8	9	10
Accuracy (%)	93.45	91.68	88.21	85.66

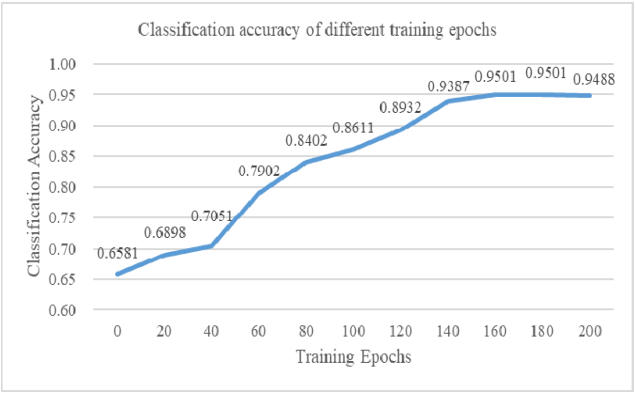


FIGURE 8. Classification accuracy of different training epochs.

unchanged and the number of network layers increases to a certain value. Therefore, the DBN layer number is set to 6 in the subsequent experimental study.

The 6-layer DBN network consists of 1 input layer, 4 hidden layers, and 1 output layer, the number of units of the 4 hidden layers is 16, 12, 8, 4 in order, the classification accuracy of different training epochs is shown in Fig. 8.

As can be seen from Fig. 8, when epochs are less than 40, the classification accuracy rate gradually increases from 65% to 70%. With the increase of epochs, the classification accuracy is significantly improved, reaching 93.87% at 140. After that, as epochs continued to increase, the classification accuracy increased slowly and reached a maximum value of 95.01% around 160. With the increase of epochs, the training time keeps increasing, and the increase of classification accuracy of epochs in the range of 150∼160 tends to be flat. Considering the time and classification accuracy, epochs are 150 in this paper.

4) COMPARISON OF EXPERIMENTAL RESULTS

a: COMPARATIVE ANALYSIS OF TRADITIONAL FEATURE EXTRACTION ALGORITHMS

Table 3 shows the classification accuracy, kappa value, and training/test time of signals processed by six different traditional feature extraction algorithms.

In Terms of Classification Accuracy and Kappa Value:

As a relatively basic feature extraction algorithm, AR has poor performance in processing such highly non-stationary signals as EEG, and its classification accuracy and kappa value are relatively low (64.26% and 0.2852). The classification accuracies of DWT and WPT are 68.35% and 68.29% respectively, and the kappa values are 0.367 and 0.3658 respectively. WPT and DWT are better than AR in the fine degree of decomposition and have a great advantage in the decomposition of low-frequency signals such as EEG. Because the EEG of left/right hand movement is mainly in the Mu rhythm and Beta rhythm, the two frequency bands are not continuous. Compared with AR, WPT and DWT can accurately extract the information of these two frequency bands. Therefore, compared with AR, WPT and DWT have certain advantages, which are conducive to extracting EEG features related to the imaginary task and improving the classification effect. As a spatial filtering feature extraction algorithm for two classification tasks, CSP can extract the spatial distribution component characteristics of each category from the BCI data of multiple channels, so its classification accuracy and kappa value (70.31% and 0.4062) are slightly higher than that of WPT and DWT. However, it is susceptible to noise interference, which makes it not very effective in feature extraction of EEG. PSD, as a method to reflect the energy change of EEG signals, has no obvious statistical characteristics when the data length of EEG signals is short. The classification accuracy and kappa value are 65% and 0.3, only slightly higher than AR and lower than WPT, and the feature extraction performance is relatively general. As a dimensionality reduction technology with good performance, PCA can achieve better dimensionality reduction for high-dimensional EEG signals and obtain better feature extraction accuracy and kappa value, its classification accuracy and kappa value are 72.85% and 0.457.

In Terms of Training/Test Time:

CSP training/test time are minimal (1.4449s and 0.0098s), WPT has the most training/test time (3.8991s and 0.0373s). The training/test time of the other four feature extraction algorithms is within this range, with little time difference and relatively average time.

On the whole, using these feature extraction algorithms alone, the feature extraction effect is relatively average in the case of little difference in training/test time, where the accuracy is below 75% and the kappa value is below 0.46. It indicates that the extracted eigenvalue is not good enough to reflect the signal, and there is still a long way from the good feature extraction performance, and the feature extraction performance of the signal needs to be further improved.

b: A COMPARATIVE ANALYSIS OF THE CROSS COMBINATION OF TRADITIONAL FEATURE EXTRACTION ALGORITHMS AND DEEP LEARNING ALGORITHMS

Table 4 shows the classification accuracy, kappa value, and training/test time of signals processed by the cross

TABLE 3. Comparison of traditional feature extraction algorithms (BCI Competition II Data Set III).

	AR	WPT	DWT	CSP	PSD	PCA
Accuracy (%)	64.26	68.29	68.35	70.31	65.00	72.85
κ	0.2852	0.3658	0.367	0.4062	0.3	0.457
Training time (s)	1.4937	3.8991	1.7997	1.4449	2.5921	2.002
Test time (s)	0.0107	0.0373	0.0113	0.0098	0.0219	0.0152

TABLE 4. Results of the traditional feature extraction algorithm combined with a deep learning algorithm (BCI Competition II Data Set III).

Deep Learning Algorithm	Evaluation Criteria	Feature Extraction Algorithm					
		AR	WPT	DWT	CSP	PSD	PCA
CAE	Accuracy (%)	83.75	85.63	86.74	82.86	81.42	85.79
	κ	0.675	0.7126	0.7348	0.6572	0.6284	0.7158
	Training time (s)	2.6698	5.0752	2.9758	2.621	3.7682	3.1781
	Test time (s)	0.0191	0.0486	0.0187	0.0178	0.0318	0.0241
CNN	Accuracy (%)	85.26	88.76	88.92	86.31	83.06	93.16
	κ	0.7052	0.7752	0.7784	0.7262	0.6612	0.8632
	Training time (s)	6.6032	9.0086	6.9092	6.5544	7.7016	7.1115
	Test time (s)	0.0473	0.0862	0.0434	0.0445	0.0651	0.054
NN	Accuracy (%)	80.32	85.89	87.61	84.75	83.32	89.32
	κ	0.6064	0.7178	0.7522	0.695	0.6664	0.7864
	Training time (s)	3.6199	6.0253	3.9259	3.5711	4.7183	4.1282
	Test time (s)	0.0259	0.0576	0.0247	0.0242	0.0399	0.0313
SAE	Accuracy (%)	87.58	86.21	83.25	86.29	84.28	86.51
	κ	0.7516	0.7242	0.665	0.7258	0.6856	0.7302
	Training time (s)	3.2781	5.6835	3.5841	3.2293	4.3765	3.7864
	Test time (s)	0.0235	0.0544	0.0225	0.0219	0.037	0.0287
DBN	Accuracy (%)	87.24	92.01	91.43	87.94	86.74	96.25
	κ	0.7448	0.8402	0.8286	0.7588	0.7348	0.925
	Training time (s)	3.8527	6.2581	4.1587	3.8039	4.9511	4.361
	Test time (s)	0.0276	0.0599	0.0261	0.0258	0.0418	0.0331

combination of traditional feature extraction algorithm and deep learning algorithm.

In Terms of Classification Accuracy and Kappa Value:

In horizontal contrast. The classification accuracy and kappa value of CAE combined with WPT, DWT and PCA are higher, the classification accuracy is higher than 85.5%, and kappa value is higher than 0.71. DWT+CAE has the highest classification accuracy and kappa value, which are 86.74% and 0.7348. The classification accuracy and kappa value of CNN combined with WPT, DWT, and PCA are higher, the classification accuracy is higher than 88.5%, and kappa value is higher than 0.77. CNN+PCA has the highest classification accuracy and kappa value, which are 93.16% and 0.8632. The classification accuracy and kappa value of NN combined with WPT, DWT, and PCA are higher, the classification accuracy is higher than 85.5%, and kappa value is higher than 0.71. NN+PCA has the highest classification accuracy and kappa value, which are 89.32% and 0.7864. The classification accuracy and kappa value of SAE combined with AR and PCA are higher, the classification accuracy is higher than 86.5%, and kappa value is higher than 0.73. SAE+AR has the highest classification accuracy and kappa value, which are 87.58% and 0.7516. The classification accuracy and kappa value of DBN combined with WPT, DWT, and PCA are higher, the classification accuracy is higher than 91%, and kappa value is higher than 0.82. DBN+PCA has the highest classification accuracy and kappa value, which are 96.25% and 0.925.

In vertical contrast. The classification accuracy and kappa value of AR combined with SAE and DBN are higher, the classification accuracy is higher than 87%, and kappa value is higher than 0.74. AR+SAE has the highest classification accuracy and kappa value, which are 87.58% and 0.7516. The classification accuracy and kappa value of WPT combined with CNN and DBN are higher, the classification accuracy is higher than 88.5%, and kappa value is higher than 0.77. WPT+DBN has the highest classification accuracy and kappa value, which are 92.01% and 0.8402. The classification

accuracy and kappa value of DWT combined with CNN and DBN are higher, the classification accuracy is higher than 88.5%, and kappa value is higher than 0.77. DWT+DBN has the highest classification accuracy and kappa value, which are 91.43% and 0.8286. The classification accuracy and kappa value of CSP combined with CNN, SAE, and DBN are higher, the classification accuracy is higher than 86%, and kappa value is higher than 0.72. CSP+DBN has the highest classification accuracy and kappa value, which are 87.94% and 0.7588. The classification accuracy and kappa value of PSD combined with SAE and DBN are higher, the classification accuracy is higher than 84%, and kappa value is higher than 0.68. PSD+DBN has the highest classification accuracy and kappa value, which are 86.74% and 0.7348. The classification accuracy and kappa value of PCA combined with CNN and DBN are higher, the classification accuracy is higher than 93%, and kappa value is higher than 0.86. PCA+DBN has the highest classification accuracy (96.25%) and kappa value (0.925).

From the comparison of the above two aspects. In horizontal contrast, in addition to CAE and SAE, PCA combined with other algorithms has the best results. Because CAE and SAE belong to the autoencoder class, it is an unsupervised machine learning algorithm similar to PCA. When the method is applied to the second feature extraction, there are still redundant features, so the results are not optimal. In vertical contrast, except for the combination with AR, the results of DBN combined with other algorithms are the best. Because AR is a method with poor processing effect for non-stationary signals, the result after processing is unstable. Therefore, the result of the second feature extraction is also unstable.

In Terms of Training/Test Time:

In horizontal contrast. The training/test time of CAE combined with AR, DWT, CSP, PSD, and PCA are less, the training/test time are between 2.621∼3.769s and 0.017∼0.032s, respectively. WPT+CAE has the highest training/test time, which is 5.0752s and 0.0486s. The training/test time of CNN combined with AR, DWT, CSP, PSD, and PCA are less, the training/test time are between 6.603∼7.702s and 0.043∼0.066s, respectively. WPT+CNN has the highest training/test time, which is 9.0086s and 0.0862s. The training/test time of NN combined with AR, DWT, CSP, PSD, and PCA are less, the training/test time are between 3.571∼4.719s and 0.024∼0.040s, respectively. WPT+NN has the highest training/test time, which is 6.0253s and 0.0576s. The training/test time of SAE combined with AR, DWT, CSP, PSD, and PCA are less, the training/test time are between 3.229∼4.377s and 0.022∼0.037s, respectively. WPT+SAE has the highest training/test time, which is 5.6835s and 0.0544s. The training/test time of DBN combined with AR, DWT, CSP, PSD, and PCA are less, the training/test time are between 3.803∼4.952s and 0.025∼0.042s, respectively. WPT+DBN has the highest training/test time, which is 6.2581s and 0.0599s.

In vertical contrast. The training/test time of AR combined with CAE, NN, SAE, and DBN are less, the training/test time are between 2.669∼3.853s and 0.019∼0.028s, respectively. AR+CNN has the highest training/test time, which is 6.6032s and 0.0473s. The training/test time of WPT combined with CAE, NN, SAE, and DBN are less, the training/test time are between 5.075∼6.259s and 0.048∼0.060s, respectively. WPT+CNN has the highest training/test time, which is 9.0086s and 0.0862s. The training/test time of DWT combined with CAE, NN, SAE, and DBN are less, the training/test time are between 2.975∼4.159s and 0.018∼0.027s, respectively. DWT+CNN has the highest training/test time, which is 6.9092s and 0.0434s. The training/test time of CSP combined with CAE, NN, SAE, and DBN are less, the training/test time are between 2.621∼3.804s and 0.017∼0.026s, respectively. CSP+CNN has the highest training/test time, which is 6.5544s and 0.0445s. The training/test time of PSD combined with CAE, NN, SAE, and DBN are less, the training/test time are between 3.768∼4.952s and 0.031∼0.042s, respectively. PSD+CNN has the highest training/test time, which is 7.7016s and 0.0651s. The training/test time of PCA combined with CAE, NN, SAE, and DBN are less, the training/test time are between 3.178∼4.361s and 0.024∼0.034s, respectively. PCA+CNN has the highest training/test time, which is 7.1115s and 0.054s.

From the comparison of the above two aspects. In horizontal contrast, the training/test time consumption of CAE, CNN, NN, SAE, DBN combined with WPT is the highest. In vertical contrast, the training/test time consumption of AR, WPT, DWT, CSP, PSD, PCA combined with CNN are the highest. The training/testing time of other combinations is not different, and the time is relatively average.

On the whole, among several combinations with little difference in training/test time, taking classification accuracy and kappa value into consideration, it can be concluded that the PCA-DBN algorithm has better processing results.

It can be seen from the comparison between Table 4 and Table 3, after the second feature extraction, in terms of classification accuracy and kappa value, the results are significantly improved. The classification accuracies are higher than 80%, and the kappa values are higher than 0.6. The classification accuracies and the kappa values of WPT+DBN, DWT+DBN, PCA+CNN, and PCA+DBN are reached over 90% and 0.8. PCA+DBN has the highest performance, its classification accuracy is 96.25% and the kappa value is 0.925. In terms of training/test time, the increase in training time is larger, while the increase in test time is smaller. Although BCI systems are strict about the time required to process a single sample signal in real time, the offline training model is usually applied to the online mode. Therefore, it is not sensitive to the training time of the model and focuses more on the signal processing time in the test process. Although the second feature extraction of the results of the traditional feature extraction method will increase the test time, the overall time is less than 0.1s, which can meet the requirements of the control system.

TABLE 5. The result of the paired-sample t-test with FDR correction (BCI Competition II Data set III).

Feature Extraction Algorithm	Average Value (%)			T-test Results (P-value)	T-test Results (Q-value)
	Without Deep Learning Algorithm	With Deep Learning Algorithm			
AR	64.26	CAE	83.75	2.4641E-209	3.0801E-209
		CNN	85.26	3.7715E-213	6.2858E-213
		NN	80.32	6.8082E-195	6.8082E-195
		SAE	87.58	7.7190E-232	3.8595E-231
		DBN	87.24	6.7942E-223	1.6986E-222
WPT	68.29	CAE	85.63	1.5872E-205	1.9840E-205
		CNN	88.76	8.9260E-216	2.2315E-215
		NN	85.89	3.1479E-203	3.1479E-203
		SAE	86.21	6.4056E-210	1.0676E-209
		DBN	92.01	3.0423E-234	1.5216E-233
DWT	68.35	CAE	86.74	3.3264E-210	4.1580E-210
		CNN	88.92	1.0317E-224	2.5793E-223
		NN	87.61	7.0774E-213	1.1796E-212
		SAE	83.25	9.1415E-196	9.1415E-196
		DBN	91.43	4.0965E-229	2.0483E-228
CSP	70.31	CAE	82.86	2.1522E-185	2.1522E-185
		CNN	86.31	3.5511E-202	8.8778E-202
		NN	84.75	3.7249E-192	4.6561E-192
		SAE	86.29	1.3343E-199	2.2238E-199
		DBN	87.94	1.7721E-212	8.8605E-212
PSD	65.00	CAE	81.42	1.2389E-227	2.0648E-227
		CNN	83.06	2.8196E-229	7.0490E-229
		NN	83.32	1.2221E-221	1.2221E-221
		SAE	84.28	5.7103E-222	7.1379E-222
		DBN	86.74	6.6708E-233	3.3354E-232
PCA	72.85	CAE	85.79	9.7940E-189	9.7940E-189
		CNN	93.16	3.7229E-216	9.3073E-216
		NN	89.32	4.0854E-201	6.8090E-201
		SAE	86.51	7.0209E-197	8.7761E-197
		DBN	96.25	3.7286E-219	1.8643E-218

c: STATISTICAL ANALYSIS AND HYPOTHESIS TESTING

In order to verify the significance of the performance differences between the combine algorithms, the classification accuracy (mean value) of the traditional feature extraction algorithm and the combine algorithm is tested by paired-sample t-test with FDR correction. The results are shown in Table 5.

Table 5 shows that. From the results of the t-test, the p-value is far less than 0.05, and the results of the paired-sample t-test (q-value) correct by FDR are also far less than the results calculated by $\alpha = 0.05$, the results show that the combined algorithm has a significant influence on the feature extraction of EEG signals. By analyzing and comparing the classification accuracy, p-value and q-value of the traditional feature extraction algorithm and the combine algorithm, it can be seen that the deep learning algorithm is integrated into the traditional feature extraction algorithm, which can effectively improve the feature extraction performance of EEG signals and achieve better analysis results.

d: COMPARISON BETWEEN PCA-DBN ALGORITHM AND EXISTING FEATURE EXTRACTION ALGORITHMS

Table 6 shows the comparison of the PCA-DBN feature extraction algorithm proposed in this study, the feature extraction algorithms used in the top three of BCI Competition II Data set III, and some recent algorithms with high feature extraction performance.

TABLE 6. Comparison between PCA-DBN and existing feature extraction algorithms (BCI Competition II Data Set III).

Proposed by	Feature extraction	Classifier	Accuracy (%)	κ
BCIC II 1st winner [25], [26]	Morlet-Wavelets	MND+ Bayes error	89.29	0.7858
BCIC II 2nd winner [26]	AR	LDA	84.29	0.6858
BCIC II 3rd winner [26]	AAR	NN	82.86	0.6572
M. Y Chen et al. [19]	AFAPS	LDA	90.71	0.8142
Wanzhong Chen et al. [20]	MEMD	LDA	88.35	0.767
Ö. Aydemir [21]	SBPF+LOOCV	LDA	92.9	0.858
Y. Wang et al. [22]	SC+WPT	BPNN	91	0.82
Rajdeep Chatterjee et al. [23]	FDM	SVM	81.43	0.6286
Phuoc Nguyen et al. [24]	AR	SVDD	89.29	0.7858
Kübra Saka et al. [27]	FWHT	ANN	88.87	0.7774
Jie Zhou et al. [28]	HT+DWT	LSTM	91.43	0.8286
Xiaojing Guo et al. [29]	ICA	——	87.14	0.7428
Our work	PCA+DBN	Softmax	96.25	0.925

As can be seen from the table, the classification accuracy of the traditional feature extraction algorithm is less than 90%. For example, Morlet-wavelets, AR and AAR are used in BCIC II 1st, BCIC II 2nd and BCIC II 3rd in reference [26]; AF is used in reference [19]; MEMD is used in reference [20]; FDM is used in reference [23]; AR is used in reference [24]; FWHT is used in reference [27]; ICA is used in reference [29]. Although the algorithm used in reference [26], [19], and [23] is an optimization of the traditional feature extraction algorithm, the classification accuracy rate still does not exceed 90%. The classification accuracy of the composite feature extraction algorithm is above 91%. For example, reference [21] uses an algorithm combining SBPF and LOOCV; reference [22] uses an algorithm combining SC and WPT; reference [28] uses an algorithm combining HT and DWT. The results of traditional feature extraction algorithms are lower than that of composite feature extraction algorithms, it shows that the second extraction of feature can get better results.

The accuracy and kappa value of the algorithm proposed in this study are higher than those used by the BCI Competition champion and some recent researches. It achieves better feature dimension reduction and achieves better classification accuracy.

B. EXPERIMENTAL RESULTS AND ANALYSIS (EXPERIMENTAL DATA SET: BCI COMPETITION IV DATA SET 2b)

1) EXPERIMENTAL VERIFICATION METHODS

Experimental verification methods adopt in this part are the same as A in IV:

I. In data selection and partitioning: The 10-fold cross-validation method was used to classify the experimental data. The original train sets (01∼03T) were divided into new train sets and validation sets for classifier training, then the trained classifier was used to classify the test sets (04∼05E). Finally, the average value of classification results was taken as the final classification accuracy.

II∼IV are the same as A.

2) COMPARATIVE STUDY OF VARIOUS FEATURE EXTRACTION ALGORITHMS

a: COMPARATIVE ANALYSIS OF TRADITIONAL FEATURE EXTRACTION ALGORITHMS

Table 7 shows the classification accuracy, kappa value, and training/test time of signals processed by six different traditional feature extraction algorithms.

In Terms of Classification Accuracy and Kappa Value:

As a relatively basic feature extraction algorithm, AR has a relatively low classification accuracy and kappa value, the highest values are 75.81% and 0.5162 of subject 5, and the average values are 64.57% and 0.2914. The classification accuracies and kappa values of WPT, DWT and CSP algorithms are close. The average of classification accuracies and kappa values are 69.16%, 69.63%, 69.89%, and 0.3832, 0.3926, 0.3978. The classification accuracy and kappa value of PSD are 68.28% and 0.3656, which are higher than AR, but lower than WPT, DWT, and CSP, its feature extraction performance is relatively general. The average classification accuracy and kappa value of PCA are 70.24% and 0.4048.

In Terms of Training/Test Time:

CSP training/test time are minimal (1.8344s and 0.0109s), WPT has the most training/test time (4.2136s and 0.0384s). The training/test time of the other four feature extraction algorithms is within this range.

On the whole, the results are consistent with Table 3. Using these feature extraction algorithms alone, the feature extraction effect is relatively average in the case of little difference in training/test time, where the accuracy is below 71% and the kappa value is below 0.42. It is still a long way from the good feature extraction performance.

TABLE 7. Comparison of traditional feature extraction algorithms (BCI Competition IV Data Sets 2b).

Subject	Evaluation Criteria	AR	WPT	DWT	CSP	PSD	PCA
1	Accuracy (%)	60.17	66.43	62.97	65.42	63.58	67.61
	κ	0.2034	0.3286	0.2594	0.3084	0.2716	0.3522
2	Accuracy (%)	62.21	59.46	60.36	55.61	59.44	60.27
	κ	0.2442	0.1892	0.2072	0.1122	0.1888	0.2054
3	Accuracy (%)	52.18	67.86	62.65	61.33	62.13	65.39
	κ	0.0436	0.3572	0.253	0.2266	0.2426	0.3078
4	Accuracy (%)	69.32	64.51	77.17	79.13	72.83	75.27
	κ	0.3864	0.2902	0.5434	0.5826	0.4566	0.5054
5	Accuracy (%)	75.81	75.22	74.21	75.64	76.13	78.55
	κ	0.5162	0.5044	0.4842	0.5128	0.5226	0.571
6	Accuracy (%)	70.94	72.37	75.57	72.31	71.86	72.13
	κ	0.4188	0.4474	0.5114	0.4462	0.4372	0.4426
7	Accuracy (%)	66.28	70.45	71.23	70.75	68.28	72.36
	κ	0.3256	0.409	0.4246	0.415	0.3656	0.4472
8	Accuracy (%)	63.21	78.88	68.96	75.72	71.71	69.92
	κ	0.2642	0.5776	0.3792	0.5144	0.4342	0.3984
9	Accuracy (%)	61.02	67.26	73.51	73.08	68.58	70.63
	κ	0.2204	0.3452	0.4702	0.4616	0.3716	0.4126
Average	Accuracy (%)	64.57	69.16	69.63	69.89	68.28	70.24
	κ	0.2914	0.3832	0.3926	0.3978	0.3656	0.4048
	Training time (s)	1.8837	4.3137	2.1928	1.8344	2.9933	2.3972
	Test time (s)	0.0118	0.0385	0.0124	0.0109	0.023	0.0163

b: A COMPARATIVE ANALYSIS OF THE CROSS COMBINATION OF TRADITIONAL FEATURE EXTRACTION ALGORITHMS AND DEEP LEARNING ALGORITHMS

Table 8 shows the average value (classification accuracy, kappa value, training/test time) of Subject 1-9 processed after the cross-combination of six traditional feature extraction algorithms and five deep learning algorithms.

From the comparison of classification accuracy, kappa value, and training/test time horizontal and vertical, it can be seen that the results are consistent with Table 4, and the same conclusion can be drawn.

It can be seen from the comparison between Table 8 and Table 7, integrating deep learning into traditional feature extraction algorithms can improve the feature extraction performance of signals. In terms of classification accuracy and kappa value, classification accuracies are higher than 78% and kappa values are higher than 0.57. The classification accuracy and kappa value of AR+DBN are relatively high in AR class, the classification accuracy and kappa value of WPT+DBN are relatively high in WPT class, the classification accuracy and kappa value of DWT+DBN are relatively high in DWT class, the classification accuracy and kappa value of CSP+DBN are relatively high in CSP class, the classification accuracy and the kappa value of PSD+DBN are relatively high in PSD class, and the classification accuracy and the kappa value of PCA+DBN are relatively high in PCA class. The classification accuracies and the kappa values of WPT+DBN, DWT+DBN, and PCA+DBN are reached over 88% and 0.77. PCA+DBN has the highest performance, its classification accuracy is 91.71% and the kappa value is 0.8342. In terms of training/test time, although the integration of deep learning into the traditional feature extraction algorithm increases the test time, the overall time is less than 0.1s, which meets the requirements of the control system.

The comparison results in Table 8 and Table 7 are consistent with those in Table 4 and Table 3, which further verifies the effectiveness of the combined feature extraction algorithm.

c: STATISTICAL ANALYSIS AND HYPOTHESIS TESTING

Same as in A, in order to verify the significance of the performance differences between the combine algorithms, the classification accuracy (mean value) of the traditional feature extraction algorithm and the combine algorithm is tested by paired-sample t-test with FDR correction. The results are shown in Table 9.

It can be seen from Table 9 that the analysis results are consistent with those in Table 5, which further proves that the fusion of traditional feature extraction algorithms and deep learning algorithms can effectively improve the feature extraction effect of EEG signals.

d: COMPARISON BETWEEN PCA-DBN ALGORITHM AND EXISTING FEATURE EXTRACTION ALGORITHMS

Table 10 shows the comparison of the PCA-DBN feature extraction algorithm proposed in this study, the feature extraction algorithms used in the top three of BCI Competition IV

TABLE 8. Results of the traditional feature extraction algorithm combined with a deep learning algorithm (BCI Competition IV Data Sets 2b).

Deep Learning Algorithm	Evaluation Criteria	Feature Extraction Algorithm					
		AR	WPT	DWT	CSP	PSD	PCA
CAE	Accuracy (%)	82.25	83.88	85.17	82.27	79.85	84.42
	κ	0.645	0.6776	0.7034	0.6454	0.597	0.6884
	Training time (s)	3.3303	5.7602	3.6394	3.281	4.4399	3.8438
	Test time (s)	0.0209	0.0514	0.0206	0.0195	0.0342	0.0262
CNN	Accuracy (%)	83.79	87.26	86.39	85.43	83.25	86.91
	κ	0.6758	0.7452	0.7278	0.7086	0.665	0.7382
	Training time (s)	7.2979	9.7279	7.607	7.2486	8.4075	7.8114
	Test time (s)	0.0479	0.0868	0.0441	0.0452	0.0657	0.0542
NN	Accuracy (%)	78.51	83.61	85.21	83.36	81.62	85.87
	κ	0.5702	0.6722	0.7042	0.6672	0.6324	0.7174
	Training time (s)	4.2887	6.7186	4.5978	4.2394	5.3983	4.8021
	Test time (s)	0.0269	0.0599	0.0261	0.0252	0.0416	0.0327
SAE	Accuracy (%)	83.86	83.43	81.14	84.19	82.78	85.06
	κ	0.6772	0.6686	0.6228	0.6838	0.6556	0.7012
	Training time (s)	3.9439	6.3738	4.253	3.8946	5.0535	4.4574
	Test time (s)	0.0248	0.0569	0.0241	0.0232	0.0389	0.0304
DBN	Accuracy (%)	86.41	88.92	89.09	87.87	86.63	91.71
	κ	0.7282	0.7784	0.7818	0.7574	0.7326	0.8342
	Training time (s)	4.5235	6.9534	4.8326	4.4742	5.6331	5.037
	Test time (s)	0.0284	0.062	0.0274	0.0266	0.0434	0.0343

Data sets 2b, and some recent algorithms with high feature extraction performance.

As can be seen from the table, the classification accuracy of the traditional feature extraction algorithm is less than 80.3%. For example, FBCSP, CSSD and CSP are used in BCIC IV 1st, BCIC IV 2nd and BCIC IV 3rd in reference [34]; STFT are used in reference [30] and [31]; CSP are used in reference [33], [39], [40], and [41]; FBCSP are used in reference [34] and [36]. The algorithm used in reference [37] [42] is an optimization of the traditional feature extraction algorithm and achieved a high classification accuracy. The classification accuracy of the composite feature extraction algorithm is above 81%. For example, reference [32] uses an algorithm combining FFT and FDCSP; reference [35] uses an algorithm combining CSP, SFFS, and RES. The composite feature extraction algorithm is more stable than the traditional feature extraction algorithm, which indicates that the second feature extraction can achieve better performance.

The accuracy and kappa value of the algorithm proposed in this study are higher than those used by the BCI Competition champion and some recent researches. It achieves better feature dimension reduction and achieves better classification accuracy.

C. FURTHER VALIDATION OF THE PCA-DBN ALGORITHM (EXPERIMENTAL DATA SET: LABORATORY DATASET)

In order to further verify the accuracy of the PCA-DBN algorithm, the experimental data set of Beijing Aerospace Measurement & Control Technology Co. Ltd. R&D center was validated by the algorithm.

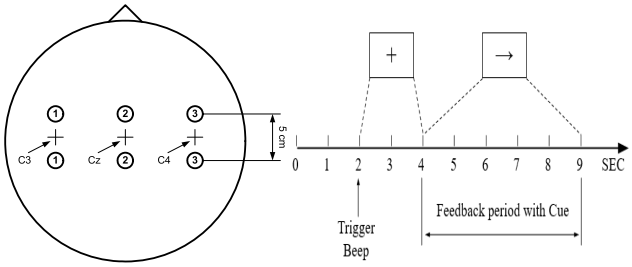


FIGURE 9. Electrode positions (left) and timing scheme (right) of laboratory data.

1) DATASET

The experimental data set was collected from 10 subjects, subjects were 22∼43 years old, each of whom conducted 20 experiments, a total of 200 motor imagery experiments, the left/right hand motor imagery experiments were conducted 100 times each. The samples were collected every 9s, when t = 0∼2s, the subject was in a ready state and did not make any movement. Started from t = 2s, voice prompt; during the period of t = 4∼9s, subjects exercised imagination. Experimental data were collected through a 64-channel Neuroscan, the sampling frequency is 128Hz, filtered by a bandpass filter of 0.5∼30Hz, C3, Cz, and C4 were selected. The electrode positions and timing scheme are shown in Fig. 9. Compared with the BCI Competition II Data set III, the prompt stage is extended to 2s, and the effective signal acquisition stage is 4∼9s.

2) EXPERIMENTAL VERIFICATION METHODS

Experimental verification methods adopt in this study are the same as A in IV:

TABLE 9. The result of the paired-sample t-test with FDR correction (BCI Competition IV Data Sets 2b).

Feature Extraction Algorithm	Average Value (%)			T-test Results (P-value)	T-test Results (Q-value)
	Without Deep Learning Algorithm	With Deep Learning Algorithm			
AR	64.57	CAE	82.25	5.9599E-05	7.4499E-05
		CNN	83.79	4.3083E-05	7.1805E-05
		NN	78.51	3.7732E-04	3.7732E-04
		SAE	83.86	2.9884E-05	7.4710E-05
		DBN	86.41	1.1798E-05	5.8990E-05
WPT	69.16	CAE	83.88	5.1498E-05	8.5830E-05
		CNN	86.91	1.8727E-05	4.6818E-05
		NN	83.61	7.6176E-05	9.5220E-05
		SAE	83.43	7.8334E-05	7.8334E-05
		DBN	88.92	8.3517E-06	4.1759E-05
DWT	69.63	CAE	85.17	7.3239E-05	9.1549E-05
		CNN	86.39	3.6630E-05	9.1575E-05
		NN	85.21	6.4260E-05	1.0710E-04
		SAE	81.14	4.5365E-04	4.5365E-04
		DBN	89.09	1.8068E-05	9.0340E-05
CSP	69.89	CAE	82.27	1.2570E-03	1.2570E-03
		CNN	85.43	3.1400E-04	7.8500E-04
		NN	83.36	6.8674E-04	8.5843E-04
		SAE	84.19	4.9134E-04	8.1890E-04
		DBN	87.87	1.1800E-04	5.9000E-04
PSD	68.28	CAE	79.85	2.9686E-04	2.9686E-04
		CNN	83.25	2.8084E-05	7.0210E-05
		NN	81.62	1.1680E-04	1.4600E-04
		SAE	82.78	7.2461E-05	1.2077E-04
		DBN	86.63	8.5328E-06	4.2664E-05
PCA	70.24	CAE	84.42	5.2703E-05	5.2703E-05
		CNN	87.26	1.5506E-05	3.8765E-05
		NN	85.87	2.5113E-05	4.1855E-05
		SAE	85.06	4.4785E-05	5.5981E-05
		DBN	91.71	1.9496E-06	9.7480E-06

I. In data selection and partitioning: The 10-fold cross-validation method was used to classify the experimental data. The original train sets were divided into new train sets and validation sets (validation sets were also test sets) for classifier training, then the trained classifier was used to classify the test sets. Finally, the average value of classification results was taken as the final classification accuracy.

II∼IV are the same as A.

3) EXPERIMENTAL VERIMENTAL RESULTS

Table 11 shows the MI-EEG data of 10 subjects are processed, the classification accuracy and kappa value (highest and average), and the training/test time (fastest and average) of each subject were obtained respectively. In terms of classification accuracy and kappa value, only the highest classification accuracy and the highest kappa value of the 5th subject are lower, which are 89.33% and 0.7866. The average classification accuracies and the average kappa values of subjects 4th and 5th are lower, which are 89.92%, 87.51% and 0.7984, 0.7502. The classification accuracies of other subjects are higher than 90%, and the kappa values of other subjects are higher than 0.8. Among them, subject 3 has the highest value (classification accuracy and kappa value). In terms of training/test time, the fastest training time of the 10 subjects is between 4.13s and 4.29s, and the average training time is between 4.34s and 4.38s. The fastest test time is between 0.0323s and 0.0334s, and the average test time is between 0.0327s and 0.0337s.

TABLE 10. Comparison between PCA-DBN and existing feature extraction algorithms (BCI Competition IV Data Sets 2b).

Proposed by	Feature extraction	Classifier	Accuracy (%)	κ
BCIC IV 1st winner [34], [61]	FBCSP	NBPW	80.00	0.6
BCIC IV 2nd winner [34]	CSSD	LDA	79.00	0.58
BCIC IV 3rd winner [34], [62]	CSP	LDA+SVM	73.0	0.46
Gaowei Xu et al. [30]	STFT	CNN	74.20	0.484
Kwon-Woo Ha et al. [31]	STFT	CapsNet	77.00	0.54
Jie Wang et al. [32]	FFT+FDCSP	SVM	81.29	0.6258
Hai-Jun Rong et al. [33]	CSP	ESAFIS	73.60	0.472
Maouia Bentlemsan et al. [34]	FBCSP	RF	79.78	0.5956
Paul Bustios et al. [35]	CSP+SFFS+RES	SVM	82.15	0.643
Sitthiphong Muthong et al. [36]	FBCSP	Regularized LDA	80.26	0.6052
Geliang Tian et al. [37]	STFT	CNN	89.07	0.7814
Jun Cai et al. [38]	—	CNN+GRU+BN	78.09	0.5618
Zhichao Jin et al. [39]	CSP	SBELM	78.50	0.57
Yu Zhang et al. [40]	CSP	MKELM	78.90	0.578
Yong Jiao et al. [41]	CSP	SGRM	78.20	0.57
Yu Zhang et al. [42]	TSGSP	SVM	84.30	0.686
Our work	PCA+DBN	Softmax	91.73	0.8342

It can be seen that the PCA-DBN algorithm can extract features of MI-EEG well, achieve high classification accuracy and maintain stability in training/test time.

V. DISCUSSION

BCI technology has developed rapidly in recent years, the classification tasks of MI-EEG range from simple two kinds of motor recognition to complex two kinds of motor recognition, and then to multiple kinds of motor recognition. It makes feature extraction more and more difficult, a better algorithm is needed to extract the features of signals so as to obtain better classification results. At present, some traditional feature extraction methods, such as AR, WPT, DWT, CSP, PSD, PCA, etc, they can perform feature extraction for simple tasks, but their efficiency is low in complex tasks. Therefore, this study proposes to combine the deep learning model with the traditional feature extraction method. Deep learning, which combines the bottom features to form more abstract high-level features, is used to extract second features from the results processed by the traditional feature extraction algorithm. This method of second feature extraction can get feature signals with more features and fewer dimensions, which can effectively improve the efficiency of feature extraction in complex tasks.

The previous researches are generally the comparison between the traditional feature extraction methods and have not systematically combined the deep learning algorithm with the traditional feature extraction methods. In terms of the types of traditional feature extraction algorithms: Several commonly used ones are used, such as AR, WPT, CSP, and PSD. In terms of experimental data: Standard competition data sets or researcher's laboratory data sets are usually used. In terms of evaluation criteria: Classification accuracy and kappa value are usually used. In terms of the comparison verification method: The one-dimensional comparison method is usually adopted. Compared with previous studies, this study has made corresponding improvements in the above four aspects. In terms of kind of algorithmic comparisons: Six traditional feature extraction methods and the cross combination of six traditional feature extraction methods and five deep learning algorithms are compared and analyzed respectively, it basically covers the existing traditional feature extraction algorithm and deep learning model. In terms of experimental data: Use standard competition data sets and data sets from the researcher's lab. In terms of evaluation criteria: Classification accuracy, kappa value and training/test time are adopted to better reflect the performance of classification. In terms of comparison verification method: Firstly, the results of traditional feature extraction algorithm are compared; then, the two-dimensional results of the cross combination of traditional feature extraction algorithm and deep learning are compared, the results are analyzed by paired-sample t-test with FDR correction. The verification results show the effectiveness of quadratic feature extraction. Finally, the efficiency of the PCA-DBN algorithm is verified.

Deep learning has brought new breakthroughs to the study of BCI, combined with the traditional feature extraction algorithms, it can better improve the accuracy of task classification and classify complex tasks. However, due to the complexity of EEG signals and the increasing requirements on task classification, the main potential limitation of the PCA-DBN algorithm is that the training time may increase rapidly in the face of more complex tasks. In view of this major problem, our next research direction mainly includes:

TABLE 11. Classification accuracy of different subjects under the PCA-DBN feature extraction algorithm (Laboratory Dataset).

		Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
Accuracy(%)	Highest	93.23	94.58	98.94	94.05	89.33
	Average	90.16	92.77	97.69	89.92	87.51
κ	Highest	0.8646	0.8916	0.9788	0.881	0.7866
	Average	0.8023	0.8554	0.9538	0.7984	0.7502
Training time (s)	Fastest	4.1521	4.1876	4.2637	4.1315	4.1998
	Average	4.3632	4.3715	4.3543	4.3671	4.3736
Test time (s)	Fastest	0.0329	0.0327	0.0331	0.0326	0.0327
	Average	0.0332	0.0331	0.0336	0.0329	0.033
		Subject 6	Subject 7	Subject 8	Subject 9	Subject 10
Accuracy(%)	Highest	96.25	97.33	92.81	93.20	96.56
	Average	94.06	96.54	90.72	90.58	95.87
κ	Highest	0.925	0.9466	0.8562	0.864	0.9312
	Average	0.8812	0.9308	0.8144	0.8116	0.9174
Training time (s)	Fastest	4.1473	4.2052	4.2874	4.1641	4.1786
	Average	4.3495	4.3537	4.3651	4.3518	4.3569
Test time (s)	Fastest	0.0332	0.0325	0.0334	0.0328	0.0323
	Average	0.0335	0.0328	0.0337	0.0331	0.0327

Find a general pattern combining traditional feature extraction algorithm with deep learning; Establish a suitable and generalized deep learning network that can process different data; Improve classification accuracy while maintaining classification time efficiency. Through the improvement of the above research directions, the algorithm is constantly optimized to improve the efficiency of the algorithm more comprehensively.

VI. CONCLUSION

To solve the problem of strong temporal variability, large individual differences, and large redundancy of eigenvalues extracted by traditional feature extraction algorithms of MI-EEG. In this paper, deep learning algorithms are combined with traditional feature extraction algorithms, and an MI-EEG feature extraction method based on principal component analysis and deep belief networks is proposed, namely PCA-DBN. This method well integrates the characteristics of PCA and DBN algorithms, not only achieve feature dimensionality reduction of EEG signals in a short time and low algorithm complexity, but also process the redundant information that still exists after feature dimensionality reduction, and extract higher signal features in MI-EEG. Verification of two types of public data sets and laboratory data sets shows that the method is effective, the comparison with other feature extraction algorithms shows that this method achieves good results in feature extraction of MI-EEG, and the classification accuracy and kappa value are obviously improved. This is a specific application of deep learning in EEG data processing and provides a new idea for expanding the application of deep learning in MI-EEG.

REFERENCES

[1] T. Vaughan, "Guest editorial brain-computer interface technology: A review of the second international meeting," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 94–109, Jun. 2003.

[2] M. Fu, J. Daly, and M. Cavusoglu, "Assessment of EEG event-related desynchronization in stroke survivors performing shoulder-elbow movements," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Jul. 2006, pp. 3158–3164.

[3] Z. Mu and J. Hu, "Research of EEG identification computing based on AR model," in *Proc. Int. Conf. Future Biomed. Inf. Eng. (FBIE)*, Dec. 2009, pp. 366–368.

[4] I. T. Hettiarachchi, T. T. Nguyen, and S. Nahavandi, "Motor imagery data classification for BCI application using wavelet packet feature extraction," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2014, pp. 519–526.

[5] N. K. Verma, L. S. V. S. Rao, and S. K. Sharma, "Motor imagery EEG signal classification on DWT and crosscorrelated signal features," in *Proc. 9th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Dec. 2014, pp. 1–6.

[6] B. Nasihatkon, R. Boostani, and M. Z. Jahromi, "An efficient hybrid linear and kernel CSP approach for EEG feature extraction," *Neurocomputing*, vol. 73, nos. 1–3, pp. 432–437, Dec. 2009.

[7] O. Kocak, F. Beytar, H. Firat, Z. Telatar, and O. Erogul, "Comparison of non-parametric PSD detection methods in the anaylsis of EEG signals in sleep apnea," in *Proc. Med. Technol. Nat. Congr. (TIPTEKNO)*, Oct. 2016, pp. 1–4.

[8] A. Chavan and M. Kolte, "Improved EEG signal processing with wavelet based multiscale PCA algorithm," in *Proc. Int. Conf. Ind. Instrum. Control (ICIC)*, May 2015, pp. 1056–1059.

[9] A. Z. Al-Marridi, A. Mohamed, and A. Erbad, "Convolutional autoencoder approach for EEG compression and reconstruction in m-health systems," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2018, pp. 370–375.

[10] Q. Liu, J. Cai, S.-Z. Fan, M. F. Abbod, J.-S. Shieh, Y. Kung, and L. Lin, "Spectrum analysis of EEG signals using CNN to model patient's consciousness level based on anesthesiologists' experience," *IEEE Access*, vol. 7, pp. 53731–53742, 2019.

[11] C. Oh, M.-S. Kim, and J.-J. Lee, "EEG signal classification based on PCA and NN," in *Proc. SICE-ICASE Int. Joint Conf.*, 2006, pp. 1848–1851.

[12] B. Yan, Y. Wang, Y. Li, Y. Gong, L. Guan, and S. Yu, "An EEG signal classification method based on sparse auto-encoders and support vector machine," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Jul. 2016, pp. 1–6.

[13] H. Cai, X. Sha, X. Han, S. Wei, and B. Hu, "Pervasive EEG diagnosis of depression using deep belief network withto be published three-electrodes EEG collector," in *Proc. IEEE Int. Conf. Bioinformatics Biomed. (BIBM)*, Dec. 2016, pp. 1239–1246.

[14] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "EEG-based emotion classification using deep belief networks," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2014, pp. 1–6.

[15] K. Li, X. Li, Y. Zhang, and A. Zhang, "Affective state recognition from EEG with deep belief networks," in *Proc. IEEE Int. Conf. Bioinformatics Biomed.*, Dec. 2013, pp. 305–310.

[16] P. Li, W. Jiang, and F. Su, "Single-channel EEG-based mental fatigue detection based on deep belief network," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 367–370.

[17] D. Wulsin, J. Blanco, R. Mani, and B. Litt, "Semi-supervised anomaly detection for eeg waveforms using deep belief nets," in *Proc. 9th Int. Conf. Mach. Learn. Appl.*, Dec. 2010, pp. 436–441.

[18] G. Altan, Y. Kutlu, and N. Allahverdi, "Deep belief networks based brain activity classification using eeg from slow cortical potentials in stroke," *Int. J. Appl. Math., Electron. Comput.*, vol. 4, p. 205, Dec. 2016.

[19] M. Chen, Y. Fang, and X. Zheng, "Phase space reconstruction for improving the classification of single trial EEG," *Biomed. Signal Process. Control*, vol. 11, pp. 10–16, May 2014.

[20] W. Chen and Y. You, "Masking empirical mode decomposition-based hybrid features for recognition of motor imagery in EEG," in *Proc. 3rd IEEE Int. Conf. Control Sci. Syst. Eng. (ICCSSE)*, Aug. 2017, pp. 548–551.

[21] O. Aydemir, "Combining sub-band power features extracted from different time segments of EEG trials," in *Proc. 40th Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2017, pp. 383–386.

[22] Y. Wang, X. Shen, and Z. Peng, "Research of EEG recognition algorithm based on motor imagery," in *Proc. 2nd Int. Conf. Robot. Autom. Sci. (ICRAS)*, Jun. 2018, pp. 1–5.

[23] R. Chatterjee, T. Bandyopadhyay, D. K. Sanyal, and D. Guha, "Dimensionality reduction of EEG signal using fuzzy discernibility matrix," in *Proc. 10th Int. Conf. Hum. Syst. Interact. (HSI)*, Jul. 2017, pp. 131–136.

[24] P. Nguyen, D. Tran, T. Le, T. Hoang, and D. Sharma, "Multi-sphere support vector data description for brain-computer interface," in *Proc. 4th Int. Conf. Commun. Electron. (ICCE)*, Aug. 2012, pp. 318–321.

[25] S. Lemm, C. Schafer, and G. Curio, "BCI competition 2003-data set III: Probabilistic modeling of sensorimotor μ rhythms for classification of imaginary hand movements," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1077–1080, Jun. 2004.

[26] J. G. Servin-Aguilar, L. Rizo-Dominguez, and J. A. Pardinas-Mir, "A comparison between wavelet families to compress an EEG signal," in *Proc. IEEE ANDESCON*, Oct. 2016, pp. 1–4.

[27] O. Aydemir and T. Kayikçioğlu, "Classification of EEG signals recorded during right/left hand movement imagery using fourier transform based features," in *Proc. IEEE 19th Signal Process. Commun. Appl. Conf. (SIU)*, Apr. 2011, pp. 415–418.

[28] J. Zhou, M. Meng, Y. Gao, Y. Ma, and Q. Zhang, "Classification of motor imagery EEG using wavelet envelope analysis and LSTM networks," in *Proc. Chin. Control And Decis. Conf. (CCDC)*, Jun. 2018, pp. 5600–5605.

[29] X. Guo and X. Wu, "Motor imagery EEG classification based on dynamic ICA mixing matrix," in *Proc. 4th Int. Conf. Bioinf. Biomed. Eng.*, Jun. 2010, pp. 1–4.

[30] G. Xu, X. Shen, S. Chen, Y. Zong, C. Zhang, H. Yue, M. Liu, F. Chen, and W. Che, "A deep transfer convolutional neural network framework for EEG signal classification," *IEEE Access*, vol. 7, pp. 112767–112776, 2019.

[31] K.-W. Ha and J.-W. Jeong, "Decoding two-class motor imagery EEG with capsule networks," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Feb. 2019, pp. 1–4.

[32] J. Wang, Z. Feng, and N. Lu, "Feature extraction by common spatial pattern in frequency domain for motor imagery tasks classification," in *Proc. 29th Chin. Control And Decis. Conf. (CCDC)*, May 2017, pp. 5883–5888.

[33] H.-J. Rong, C. Li, R.-J. Bao, and B. Chen, "Incremental adaptive EEG classification of motor imagery-based BCI," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–7.

[34] M. Bentlemsan, E.-T. Zemouri, D. Bouchaffra, B. Yahya-Zoubir, and K. Ferroudji, "Random forest and filter bank common spatial patterns for EEG-based motor imagery classification," in *Proc. 5th Int. Conf. Intell. Syst., Modelling Simulation*, Jan. 2014, pp. 235–238.

[35] P. Bustios and J. L. Rosa, "Restricted exhaustive search for frequency band selection in motor imagery classification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 4208–4213.

[36] S. Muthong, P. Vateekul, and M. Sriyudthsak, "Stacked probabilistic regularized LDA on partitioning non-stationary EEG data for left/right hand imagery classification," in *Proc. IEEE EMBS Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2016, pp. 301–306.

[37] G. Tian and Y. Liu, "Study on classification of left-right hands motor imagery EEG signals based on CNN," in *Proc. IEEE 17th Int. Conf. Cognit. Informat. Cognit. Comput. (ICCI*CC)*, Jul. 2018, pp. 324–329.

[38] J. Cai, C. Wei, X.-L. Tang, C. Xue, and Q. Chang, "The motor imagination eeg recognition combined with convolution neural network and gated recurrent unit," in *Proc. 37th Chin. Control Conf. (CCC)*, Jul. 2018, pp. 9598–9602.

[39] Z. Jin, G. Zhou, D. Gao, and Y. Zhang, "EEG classification using sparse Bayesian extreme learning machine for brain–computer interface," *Neural Comput. Appl.*, vol. 10, pp. 1–9, Oct. 2018.

[40] Y. Zhang, Y. Wang, G. Zhou, J. Jin, B. Wang, X. Wang, and A. Cichocki, "Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces," *Expert Syst. Appl.*, vol. 96, pp. 302–310, Apr. 2018.

[41] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, and A. Cichocki, "Sparse group representation model for motor imagery EEG classification," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 631–641, Mar. 2019.

[42] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3322–3332, Sep. 2019.

[43] J. Choi, S. Yan, and J. Hong, "An automated method based on second order moment for defect extraction in photomask images," in *Proc. 11th Int. Conf. Adv. Commun. Technol.*, 2009, pp. 1015–1018.

[44] M.-A. Li, W. Zhu, H.-N. Liu, and J.-F. Yang, "Adaptive feature extraction of motor imagery EEG with optimal wavelet packets and SE-Isomap," *Appl. Sci.*, vol. 7, no. 4, p. 390, Apr. 2017.

[45] A. Subasi and M. Ismail Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8659–8666, Dec. 2010.

[46] V. A. Shim, K. C. Tan, C. Y. Cheong, and J. Y. Chia, "Enhancing the scalability of multi-objective optimization via restricted Boltzmann machine-based estimation of distribution algorithm," *Inf. Sci.*, vol. 248, pp. 191–213, Nov. 2013.

[47] J. Li, Z. L. Yu, Z. Gu, W. Wu, Y. Li, and L. Jin, "A hybrid network for ERP detection and analysis based on restricted Boltzmann machine," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 3, pp. 563–572, Mar. 2018.

[48] H. Rajaguru and S. K. Prabhakar, "Logistic regression Gaussian mixture model and softmax discriminant classifier for epilepsy classification from EEG signals," in *Proc. Int. Conf. Comput. Methodologies Commun. (ICCMC)*, Jul. 2017, pp. 985–988.

[49] A. Schloegl, K. Lugger, and G. Pfurtscheller, "Using adaptive autoregressive parameters for a brain-computer-interface experiment," in *Proc. 19th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., Magnificent Milestones Emerg. Opportunities Med. Eng.*, vol. 4, Nov. 2002,pp. 1533–1535.

[50] C. Neuper, A. Schlögl, and G. Pfurtscheller, "Enhancement of left-right sensorimotor EEG differences during feedback-regulated motor imagery," *J. Clin. Neurophysiol.*, vol. 16, no. 4, pp. 373–382, Jul. 1999.

[51] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Trans. Rehab. Eng.*, vol. 6, no. 3, pp. 316–325, Sep. 1998.

[52] A. Schlögl, C. Neuper, and G. Pfurtscheller, "Estimating the mutual information of an EEG-based brain-computer interface," *Biomedizinische Technik/Biomed. Eng.*, vol. 47, nos. 1–2, pp. 3–8, 2002.

[53] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Correction to 'Brain–computer communication: Motivation, aim, and impact of exploring a virtual apartment,'" *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 1, p. 119, Feb. 2008.

[54] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey," *Clin. Neurophysiol.*, vol. 118, no. 3, pp. 480–494, Mar. 2007.

[55] A. Schlögl, J. Kronegg, J. Huggins, and S. G. Mason, "Evaluation criteria for BCI research," in *Towards Brain-Computer Interfacing*. Cambridge, MA, USA: MIT Press, 2007, pp. 297–312.

[56] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clin. Neurophysiol.*, vol. 118, no. 1, pp. 98–104, Jan. 2007.

[57] S. Shahtalebi and A. Mohammadi, "A Bayesian framework to optimize double band spectra spatial filters for motor imagery classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 871–875.

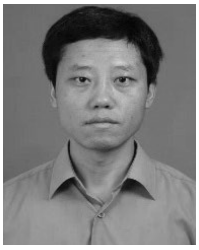
[58] K. Polat and S. Günes, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Appl. Math. Comput.*, vol. 187, no. 2, pp. 1017–1026, Apr. 2007.

[59] J. P. Kulasingham, V. Vibujithan, and A. C. De Silva, "Deep belief networks and stacked autoencoders for the P300 guilty knowledge test," in *Proc. IEEE EMBS Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2016, pp. 127–132.
[60] Z. Lu, N. Gao, Y. Liu, and Q. Li, "The detection of P300 potential based on deep belief network," in *Proc. 11th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2018, pp. 1–5.
[61] K. Keng Ang, Z. Yang Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in Brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 2390–2397.
[62] D. Coyle, A. Satti, G. Prasad, and T. M. Mcginnity, "Neural time-series prediction preprocessing meets common spatial patterns in a brain-computer interface," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 2626–2629.

LIWEI CHENG received the B.S. degree in electrical engineering and automation from Beijing Jiaotong University, Beijing, China, in 2012, and the M.S. degree in mechanical engineering from the Beijing University of Posts and Telecommunications, Beijing, in 2017, where he is currently pursuing the Ph.D. degree in mechatronic engineering. His research interests include machine learning, EEG signal processing, BCI, and robotics.

DUANLING LI received the M.S. degree in mechanical engineering from the Shaanxi University of Science and Technology, Shaanxi, China, in 1999, and the Ph.D. degree in mechanical engineering from Beihang University, Beijing, China, in 2003. She is currently a Professor with the School of Automation, Beijing University of Posts and Telecommunications. She has published many articles in international journals. Her current research interests include robotics and mechanisms.

GONGJING YU received the M.S. degree in navigation guidance and control from Beihang University, Beijing, China, in 1991. He is currently a Professor with Beijing Aerospace Measurement and Control Technology Company, Ltd. His main research interests include measurement and control technology, BCI, intelligent robot, prognostic and health management.



ZHONGHAI ZHANG received the Ph.D. degree in mechanical engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2014. He is currently a Professor with Beijing Aerospace Measurement and Control Technology Company, Ltd. He has published many articles in international journals. His current research interests include robotics and mechanisms.

XIANG LI received the B.S. degree in mechanical engineering from Liaoning Technical University, Liaoning, China, in 2015, and the M.S. degree in mechanical engineering from the Hebei University of Technology, Tianjin, China, in 2018. He is currently an Engineer with Beijing Aerospace Measurement and Control Technology Company, Ltd. His main research interests include robotics and BCI.

SHUYUE YU received the B.S. degree in measurement, control technology, and instrument and the M.S. degree in control science and engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2016 and 2019, respectively. She is currently an Engineer with Beijing Aerospace Measurement and Control Technology Company, Ltd. Her main research interests include robotics and BCI.