

Received October 18, 2019, accepted November 12, 2019, date of publication November 18, 2019, date of current version December 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2953972

# The Optimal Wavelet Basis Function Selection in Feature Extraction of Motor Imagery Electroencephalogram Based on Wavelet Packet Transformation

LIWEI CHENG<sup>1</sup>, DUANLING LI<sup>1,2</sup>, XIANG LI<sup>3</sup>, AND SHUYUE YU<sup>3</sup>

<sup>1</sup>School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China

<sup>2</sup>College of Mechanical and Electrical Engineering, Shaanxi University of Science and Technology, Xi'an 712000, China

<sup>3</sup>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 (No. 51775052), and in part by the Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2019JM-181), and in part by the Beijing Key Laboratory of Space-ground Interconnection and Convergence.

**ABSTRACT** To solve the problem of optimal wavelet basis function selection in feature extraction of motor imagery electroencephalogram (MI-EEG) by wavelet packet transformation (WPT), based on the analysis of wavelet packet transformation and wavelet basis parameters, combine with the characteristics of MI-EEG, the characteristics of wavelet basis function suitable for feature extraction of MI-EEG are summarized. On the basis of processing and analyzing of two BCI competition data sets, signal to noise ratio (SNR), root mean squared error (RMSE), classification accuracy, and kappa value are introduced as evaluation criteria for feature extraction effect, it is concluded that the rbio2.2 wavelet basis function is the optimal wavelet basis function for feature extraction of MI-EEG. Finally, the MI-EEG collected in the laboratory is processed and analyzed, further proving that the rbio2.2 wavelet basis function is the optimal wavelet basis function for feature extraction of MI-EEG.


**INDEX TERMS** Motor imagery electroencephalogram, signal to noise ratio, root mean squared error, wavelet basis function, deep belief networks.

## I. INTRODUCTION

Wavelet packet transformation with good time-frequency localization property has a good application in feature extraction of MI-EEG [1]. The frequency band range of MI-EEG is generally 1~30Hz, and the magnitude is on the  $\mu\text{V}$  scale [2], [3], it is easily disturbed by external signals, such as electrooculogram (30~200hz), electrocardiogram (0.05~100hz), electromyogram (5~2000hz), power frequency interference (50Hz), etc [4]. Not all wavelet basis functions are suitable for feature extraction of MI-EEG, and different wavelet basis functions have different processing results for the same signal, so the feature extraction effect of wavelet packet transformation depends on the selection of wavelet basis function [5]. The comparison and selection of wavelet basis functions is always a difficult problem in the application of wavelet packet transformation. At present, there is no systematic method and theory to solve the problem

of selecting the optimal wavelet basis function in feature extraction of MI-EEG [6], [7].

Servín-Aguilar *et al.* applied Haar, Daubechies, and Coiflets to wavelet transform the EEG signals, compared of three different criteria: normalized mean square error (NMSE), percent root mean square percentage (PRD), and compression ratio (CR), Haar wavelet had better performance of EEG signal processing [8]. Yan *et al.* in the wavelet packet transformation extraction of EEG signal characteristics, the best basis function algorithm was used to automatically select the most suitable wavelet basis function, wavelet methods used include Daubechies, Coiflets, and Symlets. It was concluded that db4, sym6, and coif6 had better feature extraction performance [9]. Khatun *et al.* compared stationary wavelet transform (SWT) and discrete wavelet transform (DWT) with different wavelet basis functions (such as sym3, haar, coif3, and bior4.4). The effectiveness of each combination was measured by correlation coefficient (CC), normalized mean square error (NMSE), time-frequency analysis, and execution time. It was concluded that coif3 and

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

bior4.4 were two wavelet basis functions with relatively good performance [10]. Lema-Condo *et al.* applied 18 wavelet transforms (sym2~10 and db2~10) to process the EEG signals of asymptomatic volunteers, the similarity between each filter and the average value of Daubech and Symlet was compared. It was concluded that sym6 was the most suitable EEG signal processing, followed by db5 [11]. Eraldemir and Yildirim compared different wavelet types commonly used in EEG signal analysis and classification, the feature extraction and classification performance of BayesNet and J48 classifier were compared by using Symlet, Coiflet and Bior wavelet. It was concluded that the bior2.4 wavelet basis function was the optimal wavelet basis function [12]. M. I. Al-Kadi et al. studied 113 wavelet basis functions (Daubechies, Coiflets, Biorthogonal, Reverse Biorthogonal, Discrete Meyer and Symlets), looked for the function most similar to EEG signals. By determining the minimum mean square error (MSE) and high signal to noise ratio (SNR), 57 different signals were detected. The results showed that the denoising compatibility of master wavelet symbol (sym24) was the best [5]. Al-Qazzaz *et al.* selected Daubechies, Symlet, Coiflet wavelet basis function to test the similarity of EEG signals recorded in the working memory task. Four evaluation indexes (SNR, PSNR, MSE, xcoor) were used for verification, Sym9 was the best [6]. For various details on wavelet packets, one may see [13]–[19].

In this paper, based on the wavelet packet transformation and the parameters of wavelet basis function, combined with the characteristics of MI-EEG, the wavelet basis function suitable for the feature extraction of MI-EEG is analyzed theoretically. Two public data sets (BCI Competition II Data set III and BCI Competition IV dataset 2b) were processed. Firstly, the second-order moment was used for preprocessing. Secondly, different wavelet basis functions were used for feature extraction under wavelet packet transformation. Then, the results of feature extraction were imported into deep belief networks (DBN) for training [20]. Finally, softmax classifier was used to classify imagery tasks [21]. The experimental results are identical with the theoretical analysis, and it is concluded that the rbio2.2 wavelet basis function is the optimal wavelet basis function for feature extraction of MI-EEG. In the processed of BCI Competition II Data set III, replaced DBN and softmax with NN, CNN, LDA, SVM, got the same result as DBN and softmax. It is proved that the rbio2.2 wavelet basis is the optimal wavelet basis for feature extraction of MI-EEG, which is applicable to different classifiers. The MI-EEG collected in the laboratory was processed, it is further proved that the rbio2.2 wavelet basis function is the optimal wavelet basis function for feature extraction of MI-EEG.

## II. METHOD AND PRINCIPLE

### A. WAVELET PACKET TRANSFORMATION

As a time-frequency analysis method, wavelet packet analysis is suitable for the processing of non-stationary signals, the decomposition of low and high frequency information

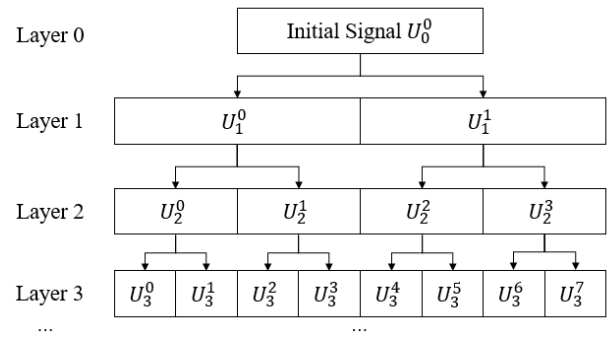


FIGURE 1. Spatial structure of wavelet packet decomposition.

at the same time, it is beneficial to extract more abundant features. Due to the zoom distance characteristic of wavelet packet transformation, it is easy to highlight the parts with the largest difference between categories, thus amplifying the differences between different classes, and help to improve the classification accuracy.

The wavelet packet uses the binary method to subdivide the wavelet subspace frequency to improve the frequency resolution. The decomposition of the wavelet packet spatial structure is shown in Fig. 1.

$U_l^r$  represents the  $r$ th wavelet packet subspace of layer  $l$ , where  $r = 1, 2, 3, \dots, 2^l - 1$ . The corresponding orthogonal basis function of the subspace is  $u_{l,k}^r(t) = 2^{-l/2} u^r(2^l t - k)$ , where  $k$  is the translation factor and satisfies the two-scale formula:

$$u_{l,0}^r = \sum_k g_0(k) u_{l-1,k}^r \quad (\text{even}) \quad (1)$$

$$u_{l,0}^r = \sum_k g_1(k) u_{l-1,k}^r \quad (\text{odd}) \quad (2)$$

In the formula,  $l, k \in Z, r = 1, 2, 3, \dots, 2^l - 1$ , low-pass filter  $g_0(k)$  and high-pass filter  $g_1(k)$  as a set of mutually orthogonal filters, both of them satisfy the condition  $g_1(k) = (-1)^{1-k} g_0(1 - k)$ .

Orthogonal wavelet packet transformation is used to decompose the signal  $x(t)$ . Wavelet packet decomposition coefficients of the  $l$  layer and the  $k$  point are obtained from the following formulas:

$$d_l^{2r}(k) = \sum_k g_0(m - 2k) d_{l-1}^r(m) \quad (3)$$

$$d_l^{2r+1}(k) = \sum_k g_1(m - 2k) d_{l-1}^r(m) \quad (4)$$

After wavelet packet decomposition, the original signal is divided into several wavelet packet subspaces according to frequency bands. The corresponding frequency bands of each subspace in the  $l$  layer are:

$$\left\{ \left[ 0, \frac{f_s}{2^{l+1}} \right]; \left[ \frac{f_s}{2^{l+1}}, \frac{2f_s}{2^{l+1}} \right]; \left[ \frac{2f_s}{2^{l+1}}, \frac{3f_s}{2^{l+1}} \right]; \dots; \left[ \frac{(2l-1)f_s}{2^{l+1}}, \frac{f_s}{2} \right] \right\}$$

$f_s$  is the signal sampling rate.

**TABLE 1.** Comparison of parameter characteristics of common wavelet basis functions.

Wavelet Basis function	Haar	Daubechies	Biorthogonal	Coiflets	Symlets	Dmeyer	ReverseBior
Abbreviation	haar	db	bior	coif	sym	dmey	rbio Nr.Nd
Representation	haar	db N	bior Nr.Nd	coif N	sym N	dmey	rbio Nr.Nd
Orthogonality	Yes	Yes	No	Yes	Yes	No	No
Biorthogonality	Yes	Yes	Yes	Yes	Yes	No	Yes
Compact Support	Yes	Yes	Yes	Yes	Yes	No	Yes
CWT	Yes	Yes	Yes	Yes	Yes	No	Yes
DWT	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Support Width	1	2N-1	Reconstitution:2Nr+1 Decomposition:2Nd+1	6N-1	2N-1	-	-
Filter Length	2	2N	Max(2Nr,2Nd)+2	6N	2N	-	-
Symmetry	Symmetry	Approximate Symmetry	Dissymmetry	Approximate Symmetry	Approximate Symmetry	Symmetry	Symmetry
Vanishing Moment of Wavelet Function	1	N	Nr-1	2N	N	-	-
Vanishing Moment of Scaling Function	-	-	-	2N-1	-	-	Nr-1

The reconstruction formula of wavelet packet coefficients at the  $l + 1$  layer and the  $k$  point is:

$$d_{l+1}^r(k) = \sum_m g_0(m-2k)d_l^{2r}(m) + \sum_m g_1(m-2k)d_l^{2r+1}(m) \quad (5)$$

According to the frequency domain characteristics of the extracted signal, the wavelet decomposition coefficient corresponding to the corresponding frequency band is selected. By using formula (5) for reconstruction, a specific frequency band signal can be obtained [3], [22]–[25].

### B. WAVELET BASIS FUNCTION PARAMETERS AND CHARACTERISTICS ANALYSIS

Wavelet basis function has five important properties: orthogonality (or biorthogonality), symmetry (or linear phase), regularity, vanishing moment, and compact support. Orthogonality reflects the perfection of wavelet basis function, describes the redundancy of data, and strict orthogonality is beneficial to the exact reconstruction of wavelet packet decomposition coefficients. Symmetry indicates that the wavelet basis function has a linear phase, and phase distortion can be avoided effectively in signal processing, prevent phase distortion during signal decomposition and reconstruction. Regularity is the embodiment of smoothness and continuous differentiability of wavelet basis function, the singularities in the signal can be found effectively. The vanishing moment determines the rate of convergence when wavelet approximates smooth function, it shows the concentration degree of energy after wavelet packet transformation. The compact support (support width) reflects the localization ability of wavelet basis function, the smaller the support width, the stronger the localization ability of wavelet basis function, and the lower the computational complexity of wavelet packet transformation [6], [8], [26]. For more details on wavelets and their applications, one may refer [27]–[32]. Tab. 1 shows the comparison of parameters of common wavelet basis functions.

(1) Haar wavelet basis function. It is one of the earliest orthogonal wavelet functions with compact support, it's also the simplest wavelet basis function, it is a single rectangular wave in the support field in the range  $t \in [0, 1]$ . Haar wavelet basis function is discontinuous in the time domain [33].

(2) Daubechies (dbN) wavelet basis function. The wavelet function  $\psi(t)$  and the scaling function  $\varphi(t)$  in the support area are  $2N - 1$ , the vanishing moment of  $\psi(t)$  is  $N$ . DbN wavelet basis function has good regularity, the smoothing error introduced by the wavelet as a sparse basis is not easily detected, the signal reconstruction process is relatively smooth. The characteristic of dbN wavelet basis function is that the vanishing moment order increases with the increase of order (sequence  $N$ ), the higher the vanishing moment, the better the smoothness, the stronger the localization ability of frequency domain, and the better the effect of frequency band division. However, it will weaken the compact support in the time-domain, increase the computation amount greatly, and make the real-time performance worse. In addition, dbN wavelet has no symmetry (nonlinear phase) except  $N = 1$ , some phase distortion will occur when the signal is analyzed and reconstructed. DbN has no definite expression (except  $N = 1$ , it is the Haar wavelet basis function when  $N = 1$ ) [34].

(3) Biorthogonal (bior Nr.Nd) wavelet basis function. Although this wavelet basis function is not an orthogonal wavelet, it is a biorthogonal wavelet, it has regularity and compact support, the reconstructed support width is  $2Nr + 1$ , and the decomposed support width is  $2Nd + 1$ . The main characteristic of bior Nr.Nd wavelet basis function is the linear phase. In general, in order to obtain the linear phase, the limitation of orthogonality should be reduced. Therefore, the biorthogonal wavelet reduces the requirement of orthogonality, some orthogonality of orthogonal wavelet is preserved, and the properties of the linear phase and short branch set are obtained [35].

(4) Coiflets (coifN) wavelet basis function. The  $2N$  moment of the wavelet function  $\psi(t)$  is zero, the  $2n - 1$  moment of the scaling function  $\varphi(t)$  is zero. The support width of the  $\psi(t)$  and  $\varphi(t)$  are  $6N - 1$ . It has better symmetry than dbN [36].

(5) Symlets (symN) wavelet basis function. It's an improvement on dbN wavelet basis function, the Symlets wavelet system is usually expressed as symN ( $N = 2, 3, \dots, 8$ ). The support width of the wavelet is  $2N - 1$ , the vanishing moment is  $N$ , and it also has good regularity. Compared with dbN wavelet basis function, this wavelet basis function is consistent with dbN wavelet basis function in terms of continuity, compact support width, and filter length, but symN wavelet has better symmetry, and it can reduce the phase distortion during signal analysis and reconstruction to some extent [37].

(6) Dmeyer wavelet basis function. The wavelet function and the scaling function of Meyer wavelet are defined in the frequency domain, it's not compact support, but it converges very quickly. Dmeyer is the discrete Meyer wavelet basis function, it is Meyer wavelet based FIR approximation, for the calculation of the fast discrete wavelet transform [38].

(7) ReverseBior wavelet basis function. It is derived from the Biorthogonal wavelet basis function. In order to solve the symmetry and the incompatibility of accurate signal reconstruction, biorthogonal wavelet is introduced, biorthogonal wavelet is a dual wavelet used for signal decomposition and reconstruction. Biorthogonal wavelet solves the contradiction between linear phase and orthogonality. Because it has a linear phase characteristic, it is mainly used in signal reconstruction [39].

The difference between biorthogonal wavelet and orthogonal wavelet is that: orthogonal wavelet satisfies  $\langle \Psi_{j,k}, \Psi_{l,m} \rangle = \delta_{j,k} \delta_{l,m}$ , that is the basis function formed by the scaling and translation of the wavelet function is completely orthogonal, but the orthogonality of biorthogonal wavelet is  $\langle \Psi_{j,k}, \Psi_{l,m} \rangle = \delta_{j,k} \delta_{l,m}$ , that is to say, it has orthogonality for wavelet functions with different scales, while there is no orthogonality of the wavelet function obtained by translation between the same scale. Therefore, the wavelet used for decomposition and reconstruction is not the same function, and the corresponding filter cannot be generated by the same wavelet.

### C. ONLINE RECURSIVE ESTIMATION OF THE NORMALIZED SECOND-ORDER MOMENT

For a random signal with zero average value  $x(n)$ , the normalized second-order moment is defined as:

$$m_2 = E[x^2(n)] \tag{6}$$

For a signal  $x(n)$  of length  $N$ , the second-order moment defined by formula (1) is estimated by the following formula [40]:

$$m_2 = E_i^j[x^2(t)] \approx \frac{1}{N} \sum_{n=1}^N [x_i^j(n)]^2 \tag{7}$$

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  $x(n - N + 1), \dots, x(n - 1), x(n)$ , 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  $x(n + 1)$ .  $x(n - N + 2), \dots, x(n), x(n + 1)$  is the data in the window at this time, easy to figure out:

$$m_2(n + 1) = m_2(n) - \frac{1}{N} \{x^2(n - N) - x^2(n + 1)\} \tag{8}$$

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 defined as:

$$m_2(n + 1) = \frac{n}{n + 1} m_2(n) + \frac{x^2(n + 1)}{n + 1} \tag{9}$$

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. Once 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, and does not contain the dc component, the recurrence formula is valid [41].

### D. DBN ALGORITHM

DBN is one of the most representative network structures in deep learning, it consists of several layers of unsupervised restricted Boltzmann machines (RBM) stacked on top of each other [42]. As the core of the DBN network, RBM consists of a visible layer and a hidden layer. The purpose of the RBM network is to binary the input  $v$  for the visual layer, it can generate a set of hidden layer characteristic signals  $h$ , and make the reconstruction of  $v'$  through the characteristic signal  $h$  of the hidden layer, the error between  $v'$  and  $v$  is minimal. Let  $I$  and  $J$  be the unit number of visible layer and hidden layer respectively, given input signal  $v$  and RBM network parameter  $w$ , the system energy of this RBM is calculated as follows:

$$E(v, h) = - \sum_{i=1}^I a_i v_i - \sum_{j=1}^J b_j h_j - \sum_{i=1}^I \sum_{j=1}^J v_i h_j w_{ij} \tag{10}$$

where  $v_i$  and  $h_j$  represent binary states,  $a_i$  and  $b_j$  represent the bias terms of visible layer neuron  $i$  and hidden layer neuron  $j$  respectively. Based on the above definition, the joint probability of visible layer and hidden layer elements in the RBM network is:

$$p(v, h) = e^{-E(v, h)} / \sum_{v, h} e^{-E(v, h)} \quad (11)$$

Therefore, for the neuron  $j$  in the hidden layer, the probability that it is equal to 1 is:

$$p(h_j = 1|v) = \sigma(b_j + \sum_{i=1}^I v_i w_{ij}) \quad (12)$$

where,  $\sigma()$  is the sigmoid function.

The likelihood function of the maximized training sample is a common method to solve such problems in pattern recognition. In the RBM structure, the likelihood function is defined as:

$$\ln(f) = \sum \log p(v, h) \quad (13)$$

Further, the solution of related parameters  $w$ ,  $a$ , and  $b$  can be obtained by comparing the gradient descent and the Gibbs sampling method.

### E. SOFTMAX CLASSIFIER

Softmax classifier is an extension of the logistic regression model, often used in conjunction with DBN [21]. The sample of  $m$  training sets is  $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ , the label is  $y^{(i)} \in \{1, 2, 3, \dots, k\}$ , 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:

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

In the formula:  $\theta$  is the model parameter,  $k = 1, 2, \dots, K$ , it is obtained by minimizing the cost function  $J(\theta)$  shown in formula (15):

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{j=0}^l 1\{y^{(i)} = j\} \log p(y^{(i)} = 1|x^{(i)}; \theta) \right] \quad (15)$$

In the formula:

$$p(y^{(i)} = j|x^{(i)}; \theta) = \frac{\exp(\theta_j^T x(i))}{\sum_{j=1}^k \exp(\theta_j^T x(i))} \quad (16)$$

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. THEORETICAL ANALYSIS OF WAVELET BASIS FUNCTION SELECTION

There are many environmental disturbances in the acquisition of MI-EEG, the signal has the characteristics of low SNR and singularity. MI-EEG has a wide frequency band and interference signals and useful signals have high frequency overlap, so in the feature extraction of the EEG signals, we should not only improve the SNR but also extract the singular information in the signals, to get better features. Because of the real-time characteristic of MI-EEG, the signal processing speed is required to be very fast. Therefore, the parameter characteristics of wavelet basis function and the characteristics of MI-EEG are considered comprehensively when selecting wavelet basis function, the wavelet basis function suitable for feature extraction of MI-EEG should meet the following requirements:

(1) Good symmetry. In order to ensure that the signal is not distorted, the symmetry of filter banks is required, approximately symmetric or symmetric wavelet basis function is required.

(2) Good orthogonality. In order to facilitate the accurate reconstruction of MI-EEG after wavelet packet decomposition, a wavelet basis function with better orthogonality should be chosen, but sometimes the biorthogonal wavelets are more effective when it comes to processing the MI-EEG. So we can sacrifice some symmetry and replace orthogonality with biorthogonality.

(3) High vanishing moment. When the vanishing moment is high, the wavelet coefficients in the signal smoothness decrease rapidly with the increase of decomposition scale. However, the wavelet coefficient at the singularity does not decrease rapidly, and it can quickly determine the singular point location of the signal. Therefore, the wavelet basis function with higher vanishing moments should be selected.

(4) Good regularity. The continuous differentiability of wavelet basis function is a necessary condition for effectively finding singularities in wavelet packet transformation, at the same time to ensure a good time-frequency resolution, wavelet basis function with better regularity should be chosen. And the higher the regularity, the higher the vanishing moment.

(5) Moderate compact support width. Because the regularity depends on the width of the compact support, the greater the compact support width, the better the regularity; meanwhile, the smaller the compact support width, the stronger the localization ability of wavelet basis function, the lower the computational complexity of wavelet packet transformation, the faster the implementation speed. Therefore, considering the requirement of computing speed and signal singularity, the wavelet basis function with the proper width of compact support is selected.

From the above requirements, none of the wavelet basis functions can meet the requirements completely, only choose the wavelet basis function with good comprehensive performance. It is found that rbio wavelet clusters have better

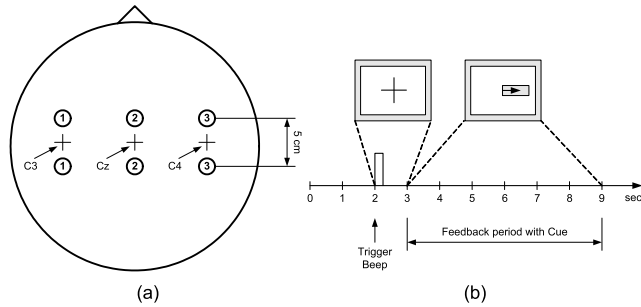


FIGURE 2. Electrode positions (a) and timing scheme (b).

performance than other wavelet clusters, more suitable for feature extraction of MI-EEG.

IV. EXPERIMENTAL VERIFICATION OF WAVELET BASIS FUNCTION SELECTION

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 consists 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 0.5~30Hz bandpass filter. Fig. 2 is electrode positions and timing scheme [43]–[46]. 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 0.5~100Hz bandpass filter, and 50Hz power frequency was

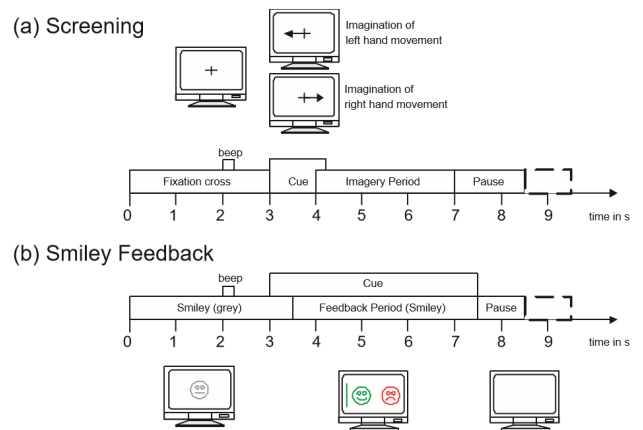


FIGURE 3. Timing scheme. (a) first two sessions, (b) last three sessions.

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. 3 [47]–[50]. For more information on the dataset, please refer to the website [http://bbci.de/competition/iv/desc\\_2b.pdf](http://bbci.de/competition/iv/desc_2b.pdf).

B. COMPREHENSIVE EVALUATION INDEX

Different statistical performance indexes are used to evaluate the performance of different wavelet basis functions in feature extraction. The first category is to measure the similarity between the recovered signal and the original signal, the second category is to measure the classification accuracy after feature extraction.

The first criterion in the first category is SNR [51], which is the ratio of signal to noise in the system. In feature extraction, the larger SNR is, the closer the signal is to the original signal after wavelet packet transformation, and the better the result of wavelet packet feature extraction.

In the window of feature extraction analysis, the data is denoted as:

$$D = [d_{ij}]_{M \times N} (i = 1, 2, \dots, M : j = 1, 2, \dots, N) \quad (17)$$

where,  $M$  is the number of time sampling points in the time window and  $N$  is the number of signal channels in the time window. That is  $d_{ij} = s_{ij} + n_{ij}$ ,  $s_{ij}$  is the signal after wavelet packet decomposition and reconstruction, and  $n_{ij}$  is the noise signal. The SNR of the whole time window is:

$$SNR = 10 \log_{10} \frac{E_s}{E_n} = 10 \log_{10} \frac{\sum d_{ji}^2}{\sum n_{ij}^2} \quad (18)$$

where  $E_s$  is the source signal energy,  $E_n$  is the signal energy after wavelet packet decomposition and reconstruction (In order to facilitate the graphic display, the SNR values in all figures in this paper are the actual SNR values divided by 100, that is  $Display\ SNR = \frac{Actual\ SNR}{100}$ ).

The second criterion in the first category is RMSE [51], which measures the deviation between the predicted value and the real value. In feature extraction, the smaller the RMSE is, the closer the signal is to the original signal after wavelet packet transformation, and the better the result of wavelet packet feature extraction.

The RMSE in the time window is:

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (d_{ij} - s_{ij})^2} \quad (19)$$

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

In the second category, the second standard is kappa value [52], which is also an index used to measure classification accuracy and is often used for the consistency test of EEG signal classification. The calculation formula of kappa value is:

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \quad (20)$$

$P_0$  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  $a_1, a_2, \dots, a_c$ , The predicted sample number of each category is  $b_1, b_2, \dots, b_c$ , the total number of samples is  $n$ , there is:

$$P_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \dots + a_c \times b_c}{n \times n} \quad (21)$$

### C. SIGNAL PROCESSING PROCESS

Hypothesis MI-EEG is expressed as:

$$x_i^j(t) = [x_1^j(t), x_2^j(t), \dots, x_n^m(t)] \in R^{N \times n \times m} \quad (22)$$

where  $N$  is the total number of sample points,  $n$  is EEG lead number,  $m$  is the number of sampling points,  $x_i^j(t)$  (the  $j$ th sampling point of lead  $i$ ) is a filtered MI-EEG signal :

$$x_i^j(t) = [x_1^j(t), x_2^j(t), \dots, x_n^m(t)] \in R^{N \times n \times m} \quad (23)$$

(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, equation (24) is used to calculate the instantaneous energy.

$$E_i^j[x^2(t)] = [x_i^j(t)]^2 \quad (24)$$

In the formula:  $E_i^j[x^2(t)]$  express the  $j$ th sampling point of the lead  $i$  of the  $t$  sample, point of MI-EEG signal of transient energy.

Suppose  $E_i^j$  is the average energy of the MI-EEG in the  $j$ th sampling point of the lead  $i$  in the  $N$  experiments, and is expressed as:

$$E_i^j = \frac{1}{N} \sum_{n=1}^N [x_i^j(n)]^2 \quad (25)$$

According to equation (25), 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 signal is analyzed by wavelet packet transformation

For the MI-EEG signal of an obvious time period selected in step (1)  $x_i^j(t)$  ( $i = \{1, 2, \dots, n\}, j = \{1, 2, \dots, m\}$ ) are analyzed by wavelet packet transformation, extract the characteristics of EEG signals. In the experimental comparison stage, different wavelet basis functions are selected.

(3) The features extracted by wavelet packet transformation are processed by DBN network

Firstly, take  $F$  as the input to the DBN network. Then, unsupervised initialization of all RBMs in DBN from the bottom up.

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  $F$  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  $W_2, W_3, \dots, W_n$ ,  $n$  is the number of RBMs in DBN.

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.

(4) Softmax classifier is used to classify tasks

The final MI-EEG feature is imported into a softmax classifier, realizes the classification of motor imagery tasks.

### D. EXPERIMENTAL VERIFICATION METHODS AND RESULTS (BCI COMPETITION II DATA SET III)

Experimental verification methods adopted 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. 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.

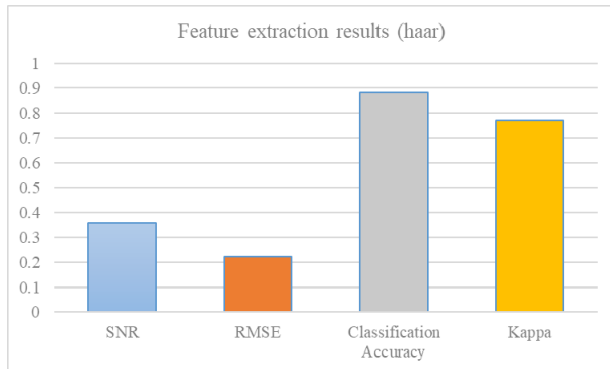


FIGURE 4. Feature extraction results (haar).

II. In data processing: Firstly, we selected the time period suitable for feature extraction through time-domain analysis. Then, we extracted feature vectors through the WPT feature extraction algorithm. Finally, we imported them into the DBN network and softmax classifier to train the appropriate classifier.

III. In the comparison and analysis of data results: SNR, RMSE, classification accuracy, and kappa value were used as evaluation criteria.

IV. In the comparison validation method: Wavelet basis functions of seven kinds of wavelet clusters are used to verify the data. Firstly, the optimal wavelet basis function in each cluster is selected. Then the optimal wavelet basis function of each cluster is compared horizontally. Finally, the optimal wavelet basis function is determined which is suitable for feature extraction of MI-EEG.

Experimental verification results are as follows:

#### (1) Haar wavelet basis function

Fig. 4 shows the results of the Haar wavelet basis function feature extraction. As can be seen from Fig. 4:

In terms of SNR and RMSE. The SNR and RMSE results obtained by Haar wavelet basis function processing are better, however, the Haar wavelet basis function is discontinuous in the time domain. Combine the time domain characteristic of MI-EEG itself, its performance as a wavelet basis function is not particularly good.

In terms of classification accuracy and kappa value, it can be seen that its classification accuracy is not up to 90%, kappa value is not up to 0.8, which is relatively average.

Therefore, it is not the choice of the optimal wavelet basis function.

#### (2) Daubechies wavelet basis function

Fig. 5 shows the results of Daubechies wavelet basis functions feature extraction. As can be seen from Fig. 5:

In terms of SNR and RMSE. From db5 to db45 wavelet basis functions, with the increase of filter length, the compact support width increases gradually. Although the smoothness of the wavelet is guaranteed, the width of the compact support increases, which leads to the locality decrease and the RMSE presents a steady rise. Db1~db4 wavelet basis functions are the four with higher SNR and lower RMSE in db wavelet

clusters. Where, when  $N=1$ , dbN is the Haar wavelet basis function, because of the Haar wavelet basis function in (1), db1 is not selected as the optimal wavelet basis function. Compare db4 with db3, SNR of db4 is slightly lower, RMSE of db4 is slightly higher, so db4 is not considered as the best choice. Compare db2 with db3, although SNR and RMSE of db2 are better, db2 is not considered as the best choice due to its smaller support width, less smoothness, smaller vanishing moment order, and less concentrated reconstruction energy.

In terms of classification accuracy and kappa value, the classification accuracy and kappa value of db1~db3 gradually increase, reaching a maximum value of 90.71% and 0.8142 at db3. After that, classification accuracy and kappa value begin to decline but keep at about 90% and 0.8, reaching the highest value of 90.71% and 0.8142 at some wavelet basis functions. Therefore, considering the perspective of classification accuracy and kappa value, the wavelet basis function up to 90.71% and 0.8142 can be considered as the alternative of the best choice.

Therefore, combine with SNR, RMSE, classification accuracy, and kappa value, in db wavelet clusters, db3, which can better take into account the compact support width and smoothness of wavelet, and has better SNR, RMSE, classification accuracy, and kappa value results, is selected as the optimal wavelet basis function.

#### (3) Biorthogonal wavelet basis function

Fig. 6 shows the results of Biorthogonal wavelet basis functions feature extraction. As can be seen from Fig. 6:

In terms of SNR and RMSE. From bior2.4 to bior6.8 wavelet basis functions, with the increase of filter length, the compact support width increases gradually. Although the smoothness of the wavelet is guaranteed, the RMSE is large, which indicates that the width of compact support increases, leading to the locality decrease. Bior1.1~bior2.2 wavelet basis functions are the four with higher SNR and lower RMSE in bior wavelet clusters. Bior1.1 is not selected as the best because of its small vanishing moment order and insufficient concentration of reconstruction energy. Compare bior1.5 with bior2.2, SNR of bior1.5 is slightly lower, RMSE of bior1.5 is slightly higher, so bior1.5 is not considered as the best choice. Compare bior1.3 with bior2.2, SNR and RMSE of bior1.3 are better, but its compact support width and smoothness are not enough, so bior1.3 is not considered as the best choice.

In terms of classification accuracy and kappa value, the classification accuracy and kappa value of bior1.1~bior2.2 gradually increase, reaching a maximum value of 90.71% and 0.8142 at bior2.2. After that, classification accuracy and kappa value begin to decline, except for bior3.5, all other classification accuracies are below 90%, and kappa values are below 0.8. Therefore, considering the classification accuracy and kappa value, the bior2.2 wavelet basis function is the best choice.

Therefore, combine with SNR, RMSE, classification accuracy, and kappa value, in bior wavelet clusters, bior2.2, which can better take into account the compact support width and



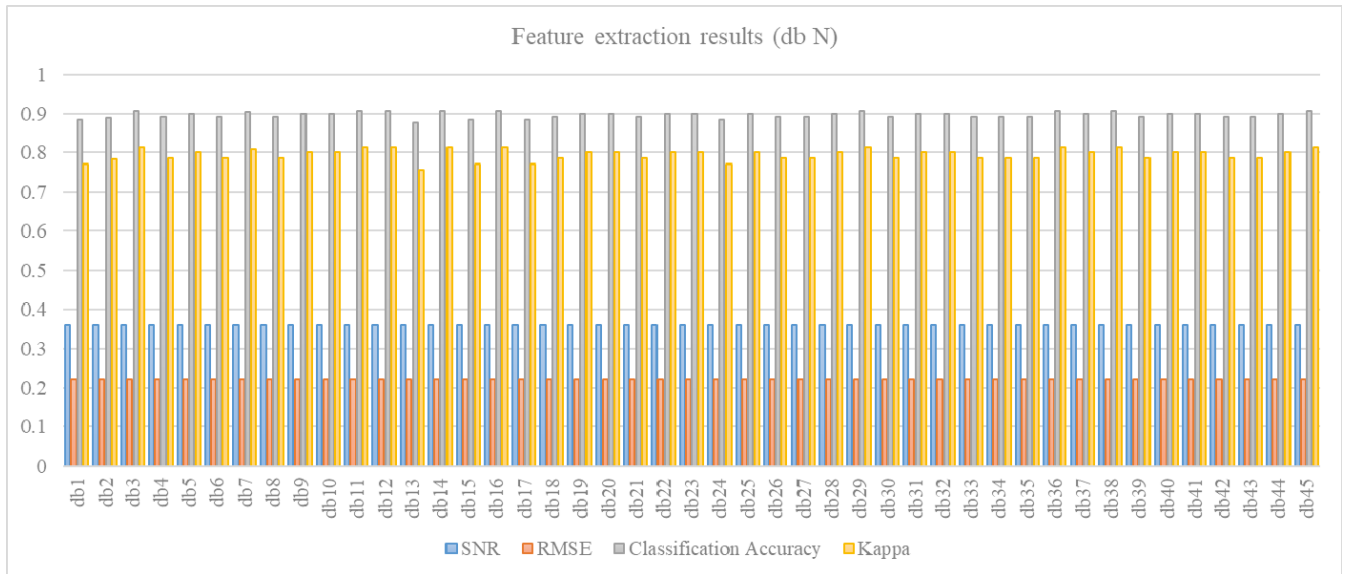


FIGURE 5. Feature extraction results (db N).

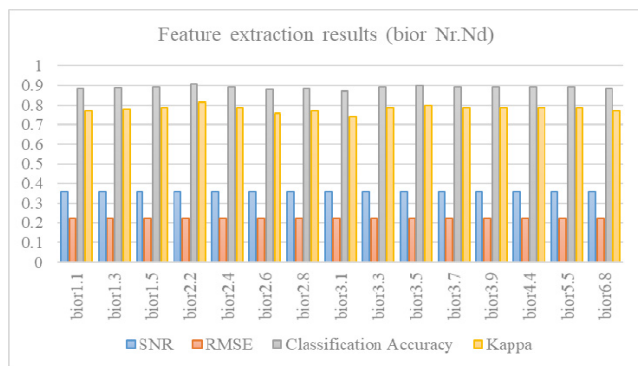


FIGURE 6. Feature extraction results (bior Nr.Nd).

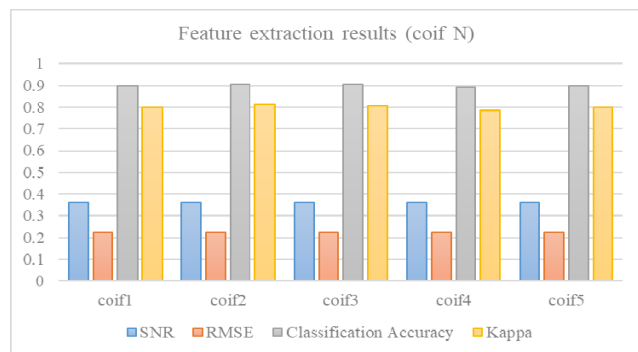


FIGURE 7. Feature extraction results (coif N).

smoothness of wavelet, and has better SNR, RMSE, classification accuracy, and kappa value results, is selected as the optimal wavelet basis function.

(4) Coiflets wavelet basis function

Fig. 7 shows the results of Coiflets wavelet basis functions feature extraction. As can be seen from Fig. 7:

In terms of SNR and RMSE. In coif1~coif5, SNR gradually decreases and RMSE gradually increases. Therefore, the optimal wavelet basis function is selected from the two wavelet basis functions with better results, coif1 and coif2. Coif1 is not selected as the best because of its small vanishing moment order and insufficient concentration of reconstruction energy.

In terms of classification accuracy and kappa value, the classification accuracy and kappa value of coif1~coif2 gradually increase, reaching a maximum value of 90.66% and 0.8132 at coif2. Classification accuracy and kappa value of coif3~coif5 is about 90% and 0.8, both are lower than coif2. Therefore, considering the classification accuracy and kappa value, the coif2 wavelet basis function is the best choice.

Therefore, combine with SNR, RMSE classification accuracy, and kappa value, in coif wavelet clusters, coif2, which can better take into account the compact support width and smoothness of wavelet, and has better SNR, RMSE, classification accuracy, and kappa value results, is selected as the optimal wavelet basis function.

(5) Symlets wavelet basis function

Fig. 8 shows the results of Symlets wavelet basis functions feature extraction. As can be seen from Fig. 8:

In terms of SNR and RMSE. From sym5 to sym30 wavelet basis functions, with the increase of filter length, the compact support width increases gradually. Although the smoothness of the wavelet is guaranteed, the RMSE is large, it indicates that the local decrease is caused by the increase of the width of compact support. Sym2~sym4 wavelet basis functions are the three with higher SNR and lower RMSE in sym wavelet clusters. Sym2 is not selected as the best because of its small vanishing moment order and insufficient concentration of reconstruction energy. Compare sym4 with sym3, SNR of

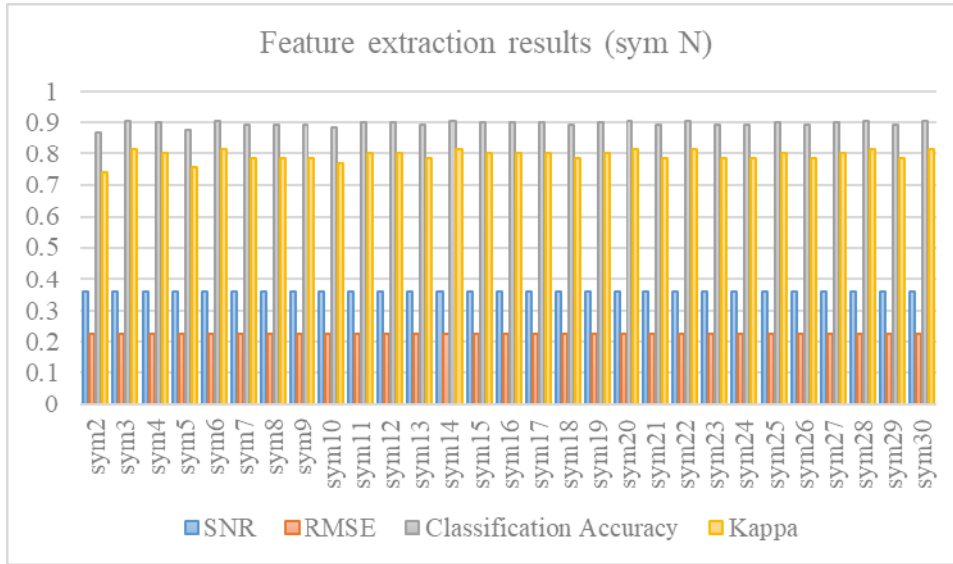


FIGURE 8. Feature extraction results (sym N).

sym4 is slightly lower, RMSE of sym4 is slightly higher, so sym4 is not considered as the best choice.

In terms of classification accuracy and kappa value, the classification accuracy and kappa value of sym2~sym3 gradually increase, reaching a maximum value of 90.71% and 0.8142 at sym3. After that, classification accuracy and kappa value begin to decline but keep at about 90% and 0.8, reaching the highest value of 90.71% and 0.8142 at some wavelet basis functions. Therefore, considering the perspective of classification accuracy and kappa value, the wavelet basis function up to 90.71% and 0.8142 can be considered as the alternative of the best choice.

Therefore, combine with SNR, RMSE, classification accuracy, and kappa value, in sym wavelet clusters, sym3, which can better take into account the compact support width and smoothness of wavelet, and has better SNR, RMSE, classification accuracy, and kappa value results, is selected as the optimal wavelet basis function.

(6) Dmeyer wavelet basis function

Fig. 9 shows the results of Dmeyer wavelet basis function feature extraction. As can be seen from Fig. 9:

In terms of SNR and RMSE. The SNR and RMSE result of dmeY is better, although it is not compact support, it converges quickly. It is an approximation of the Meyer wavelet basis function based on FIR.

In terms of classification accuracy and kappa value, its classification accuracy is 90.71%, and its kappa value is 0.8142.

Combine with SNR, RMSE, classification accuracy, and kappa value, dmeY can be used as the optimal wavelet basis function.

(7) ReverseBior wavelet basis function

Fig. 10 shows the results of ReverseBior wavelet basis functions feature extraction. As can be seen from Fig. 10:

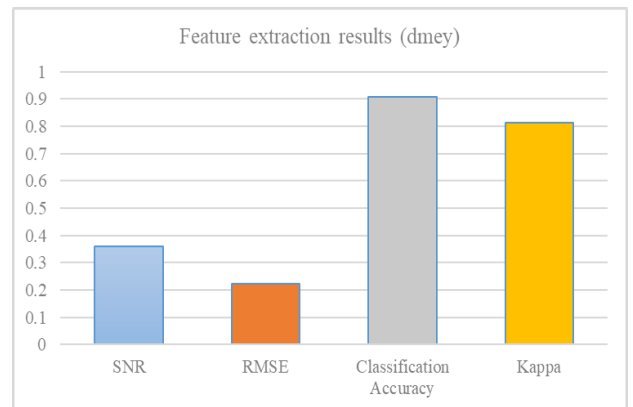


FIGURE 9. Feature extraction results (dmeY).

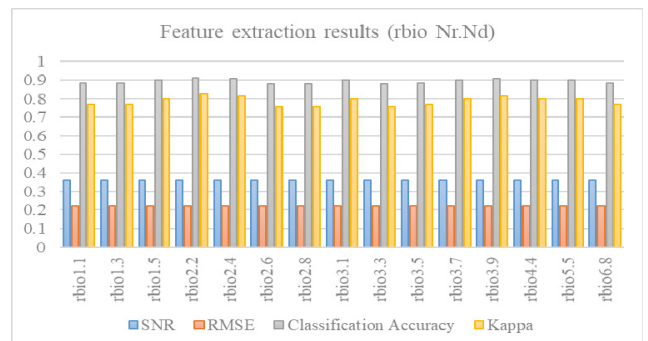


FIGURE 10. Feature extraction results (rbio Nr.Nd).

In terms of SNR and RMSE. From rbio2.4 to rbio6.8 wavelet basis functions, with the increase of filter length, the compact support width increases gradually. Although the smoothness of the wavelet is guaranteed, the RMSE is large, it indicates that the locality decrease

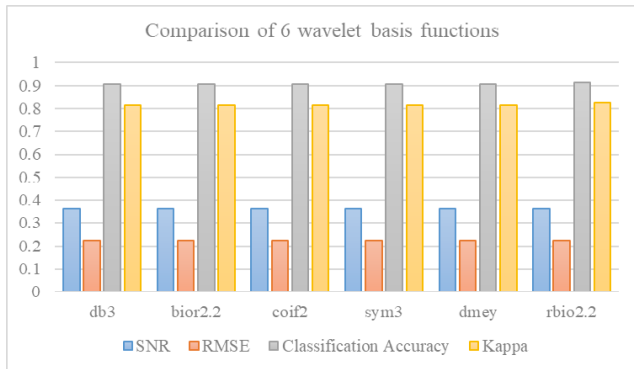


FIGURE 11. Comparison table of SNR, RMSE, classification accuracy, and kappa value of 6 wavelet basis functions (BCI competition II data set III).

is caused by the increase of the width of compact support. Rbio1.1~rbio2.2 wavelet basis functions are the four with higher SNR and lower RMSE in rbio wavelet clusters. Rbio1.1 is not selected as the best because of its small vanishing moment order and insufficient concentration of reconstruction energy. Compare rbio1.5 with rbio2.2, SNR of rbio1.5 is slightly lower, RMSE of rbio1.5 is slightly higher, so rbio1.5 is not considered as the best choice. Compare rbio1.3 with rbio2.2, although SNR and RMSE of rbio1.3 are better, its compact support width and smoothness are not enough, so rbio1.3 is not considered as the best choice.

In terms of classification accuracy and kappa value, the classification accuracy and kappa value of rbio1.1~rbio2.2 gradually increase, reaching a maximum value of 91.29% and 0.8258 at rbio2.2. After that, classification

accuracy and kappa value begin to decline but keep at 88.57%~90.71% and 0.7714~0.8142. Therefore, considering the classification accuracy and kappa value, the rbio2.2 wavelet basis function is the best choice.

Therefore, combine with SNR, RMSE, classification accuracy, and kappa value, in rbio wavelet clusters, rbio2.2, which can better take into account the compact support width and smoothness of wavelet, and has better SNR, RMSE, and classification accuracy results, is selected as the optimal wavelet basis function.

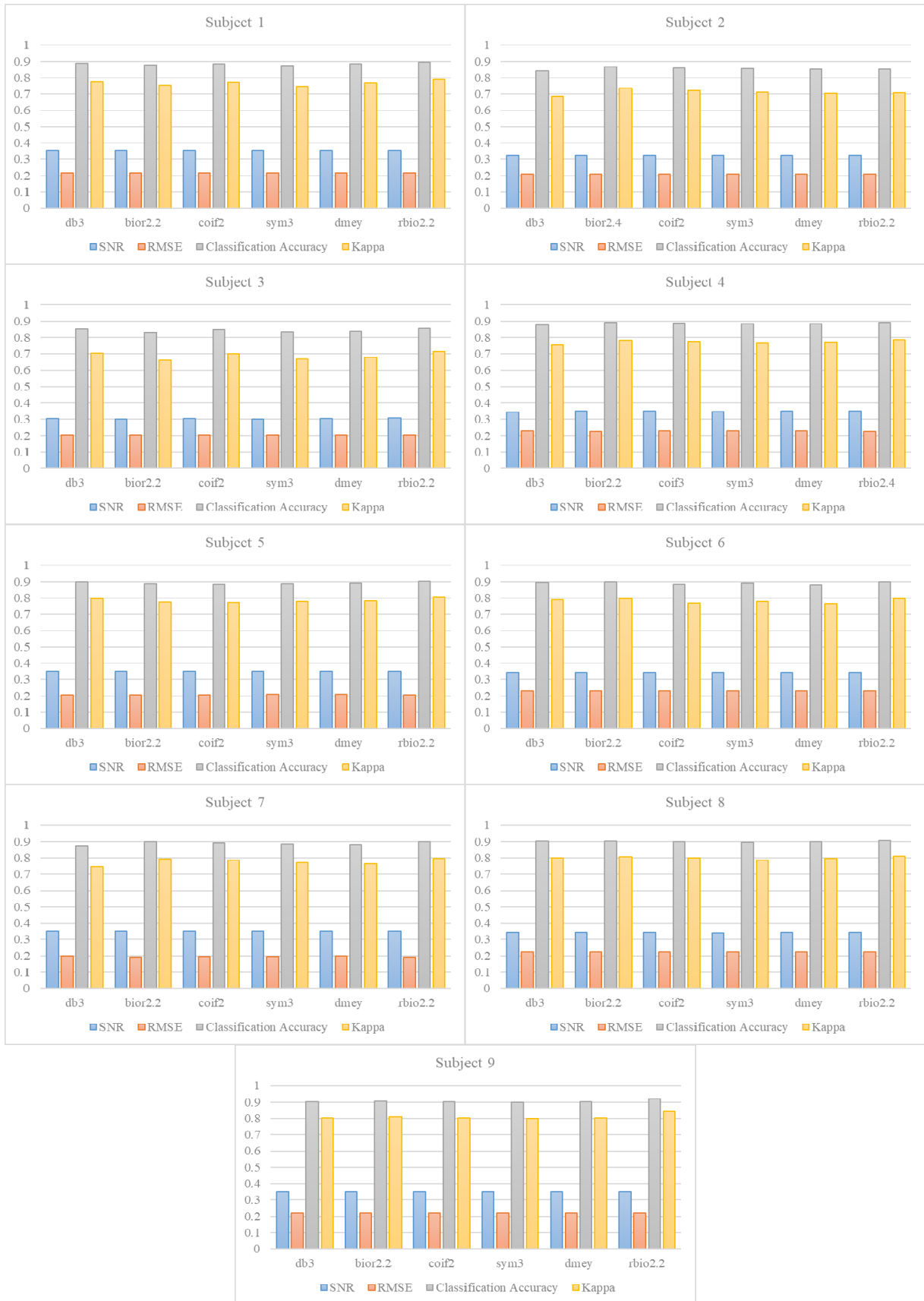
Through the analysis of seven kinds of wavelet clusters, the optimal wavelet basis function of each cluster is obtained, Fig. 11 is the comparison table of SNR, RMSE, classification accuracy, and kappa value of 6 wavelet basis functions. It can be seen that the rbio2.2 wavelet basis function has the best SNR, RMSE classification accuracy, and kappa value, and it is the optimal wavelet basis function for feature extraction of MI-EEG.

**E. FURTHER VERIFICATION OF OPTIMAL WAVELET BASIS FUNCTION SELECTION (BCI COMPETITION II DATA SET III)**

In order to further verify that the rbio2.2 wavelet basis function is the optimal wavelet basis for feature extraction of MI-EEG. In the experimental methods, only DBN and softmax are replaced by NN, CNN, LDA, and SVM respectively, other items remain unchanged, the optimal wavelet basis function of each cluster is obtained, the Fig. 12 is the comparison table of the SNR, RMSE, classification accuracy, and kappa value of the optimal wavelet basis in each wavelet cluster.



FIGURE 12. Comparison table of SNR, RMSE, classification accuracy, and kappa value of different classification methods (BCI competition II data set III).



**FIGURE 13.** Comparison table of SNR, RMSE, classification accuracy, and kappa value of 6 wavelet basis functions (BCI competition IV data sets 2b).

As can be seen from Fig. 12:

In NN, CNN, LDA, and SVM classification methods, the optimal wavelet basis function in the six wavelet clusters are db3, bior2.2, coif2, sym3, dmey, and rbio2.2. The optimal wavelet basis function of each cluster are compared horizontally, it can be seen that rbio2.2 wavelet basis function is optimal in SNR, RMSE, classification accuracy, and kappa value. It can be seen that in different classification methods, the rbio2.2 wavelet basis function has the best values. It is further proved that the rbio2.2 wavelet basis is the optimal wavelet basis for feature extraction of MI-EEG.

#### F. EXPERIMENTAL VERIFICATION METHODS AND RESULTS (BCI COMPETITION IV DATA SETS 2b)

Experimental verification methods adopted in this part are the same as D:

I. In data selection and partitioning: a 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 in D.

Experimental verification results are shown in Fig. 13:

I. In Subject 2, the optimal wavelet basis function in the six wavelet basis clusters are db3, bior2.4, coif2, sym3, dmey, and rbio2.2. The optimal wavelet basis function of each cluster are compared horizontally, it can be seen that bior2.4 wavelet basis function is optimal in SNR, RMSE, classification accuracy, and kappa value.

II. In Subject 4, the optimal wavelet basis function in the six wavelet clusters are db3, bior2.2, coif3, sym3, dmey, and rbio2.4. The optimal wavelet basis function of each cluster are compared horizontally, it can be seen that rbio2.4 wavelet basis function is optimal in SNR, RMSE, classification accuracy, and kappa value.

III. In Subject 1, 3, 5~9, the optimal wavelet basis function in the six wavelet clusters are db3, bior2.2, coif2, sym3, dmey, and rbio2.2. The optimal wavelet basis function of each cluster are compared horizontally, it can be seen that rbio2.2 wavelet basis function is optimal in SNR, RMSE, classification accuracy, and kappa value.

As can be seen from the above results. In Subject 2, the optimal wavelet basis function is in the bior wavelet clusters (bior2.4). In Subject 4, the optimal wavelet basis function is in the rbio wavelet clusters (rbio2.4). Among the other 7 subjects, the optimal wavelet basis function is rbio2.2. This further proves that the rbio2.2 wavelet basis function is the optimal wavelet basis function suitable for feature extraction of MI-EEG.

#### G. EXPERIMENTAL VERIFICATION METHODS AND RESULTS (LABORATORY DATA)

In order to further verify that the rbio2.2 wavelet basis function is the optimal wavelet basis for feature extraction of

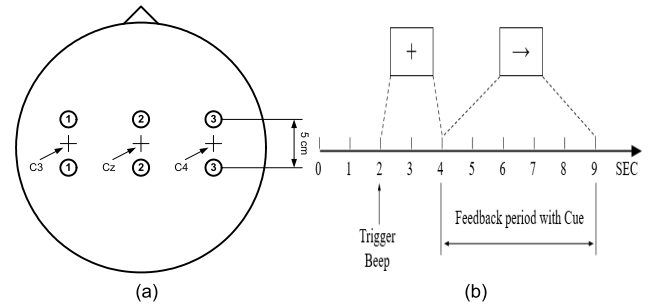


FIGURE 14. Electrode positions (a) and timing scheme (b) of laboratory data.

MI-EEG, the experimental data set of Beijing Aerospace Measurement & Control Technology Co. Ltd. R&D center.

#### 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  $t=0\sim 2s$ , the subject was in a ready state and did not make any movement. Started from  $t=2s$ , voice prompt; during the period of  $t=4\sim 9s$ , 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 sequence diagram of the experiment as shown in Fig. 14. 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 adopted in this study are the same as D:

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 are 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 in D.

#### 3) EXPERIMENTAL VERIFICATION RESULTS

Experimental verification results are shown in Fig. 15:

I. In Subject 8, the optimal wavelet basis function in the six wavelet basis clusters are db3, bior2.4, coif2, sym3, dmey, and rbio2.2. The optimal wavelet basis function of each cluster are compared horizontally, it can be seen that sym3 wavelet basis function is optimal in SNR, RMSE, classification accuracy, and kappa value.

II. In Subject 1~7,9,10, the optimal wavelet basis function in the six wavelet clusters are db3, bior2.2, coif2, sym3,

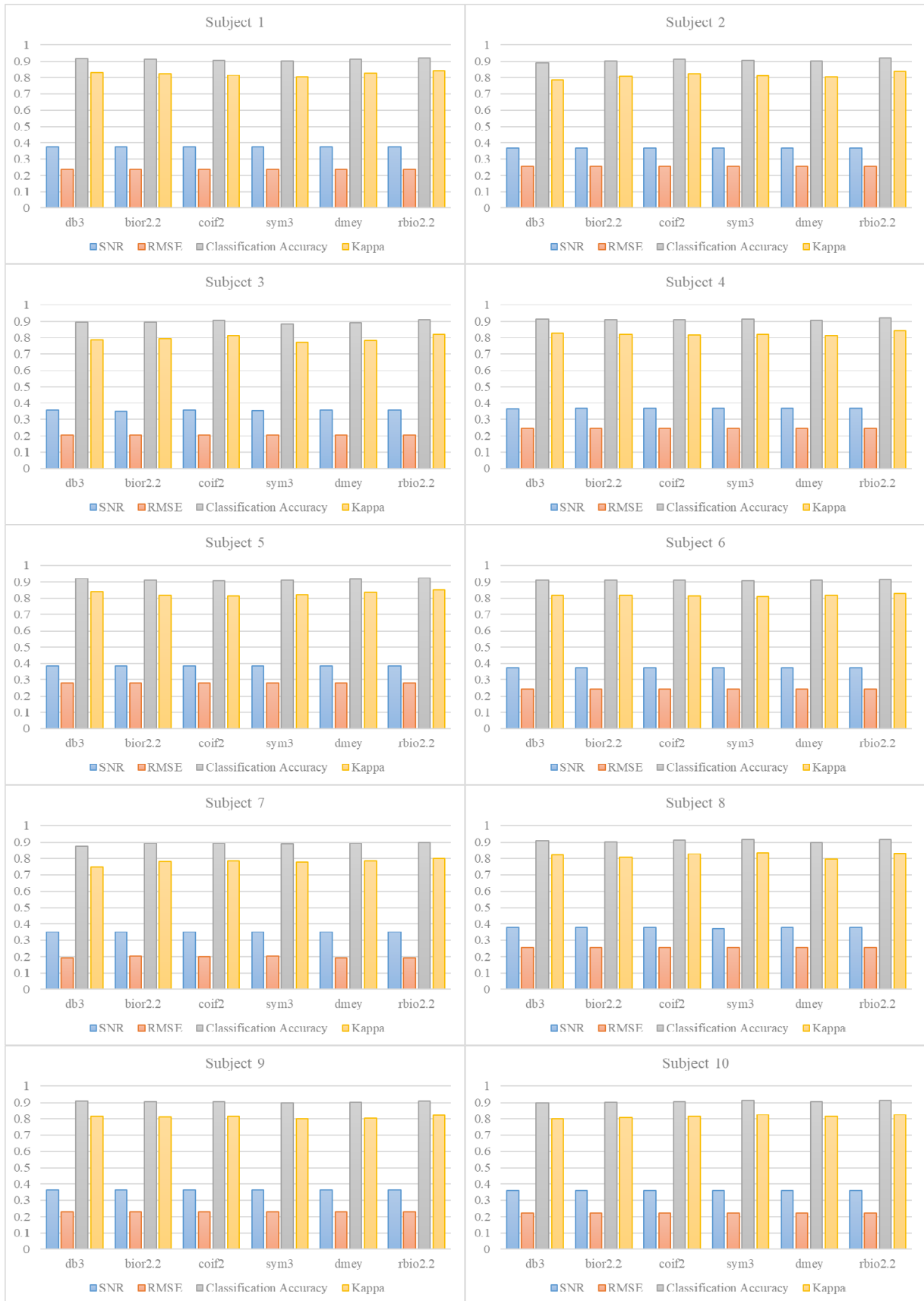


FIGURE 15. Comparison table of SNR, RMSE, classification accuracy, and kappa value of 6 wavelet basis functions (laboratory data).

dmey, and rbio2.2. The optimal wavelet basis function of each cluster are compared horizontally, it can be seen that rbio2.2 wavelet basis function is optimal in SNR, RMSE, classification accuracy, and kappa value.

As can be seen from the above results. In Subject 8, the optimal wavelet basis function is in the sym wavelet clusters (sym3). Among the other 9 subjects, the optimal wavelet basis function is rbio2.2. This further proves that the rbio2.2 wavelet basis function is the optimal wavelet basis function suitable for feature extraction of MI-EEG.

## V. DISCUSSION

As a relatively mature algorithm, wavelet packet transformation has been widely used in the analysis of biomedical signals such as ECG, EEG, and EMG. It overcomes the limitation of Fourier transformation and becomes a good time-frequency local signal analysis method due to multi-resolution. Compared with Fourier transformation, the disadvantage of wavelet packet transformation is that the wavelet basis function is not unique. Therefore, one of the difficulties in the practical application of wavelet packet analysis is the selection of the best wavelet basis function.

There is still no uniform standard for choosing a wavelet basis function. In the past, researchers usually choose the best wavelet basis function suitable for MI-EEG by experiment or experience. In the class of wavelet basis function: use several common ones, such as db, coif, and sym wavelet basis function. In the experimental data: use standard competition data sets or data sets from the researcher's lab. In the evaluation criteria: SNR, RMSE, and other indicators are usually used to measure the signal similarity. Compared with previous studies, this study has made corresponding improvements in the above three aspects. In the class of wavelet basis function: a total of 111 commonly used wavelet basis functions of 7 classes were compared and analyzed. In the experimental data: use standard competition data sets and data sets from the researcher's lab. In the evaluation criteria: include SNR, RMSE, classification accuracy, and kappa value, this set of indicators not only makes a comparative analysis on the similarity of signals but also reflects the difference of wavelet basis function from the result of the task.

Our next step of research faces several important questions: how to better relate the wavelet basis parameters and characteristic theory with the experimental results, and summarize the wavelet basis function theoretically. How to establish a suitable and generalized wavelet basis function to process more extensive data. How to design a kind of wavelet base function according to own demand, and realize better signal processing. These important questions are our next research objectives.

## VI. CONCLUSION

Aimed at the problem of optimal wavelet basis function selection in feature extraction of MI-EEG by wavelet packet transformation, based on the analysis of the basic parameters and characteristics of wavelet basis function, according to the

characteristics of the MI-EEG, the requirements of wavelet basis function suitable for the feature extraction of MI-EEG are summarized. It is concluded that "The rbio wavelet basis function of a certain order is suitable for feature extraction of MI-EEG". Then, SNR, RMSE, classification accuracy, and kappa value are introduced as evaluation criteria for wavelet basis function selection, 7 kinds of wavelet clusters are used to process and analyze the data (BCI Competition II Data set III, BCI Competition IV Data sets 2b, and Laboratory data). Finally, the rbio2.2 wavelet basis function is the optimal wavelet basis function in line with the feature extraction of MI-EEG. In the future, we hope to test more data sets, it is verified that this wavelet basis is the optimal wavelet basis for feature extraction of MI-EEG by wavelet packet transformation, more comprehensively verify the validity of the conclusion.

## REFERENCES

- [1] Y. Li, L. Zhang, B. Li, G. Yan, X. Geng, Z. Jin, Y. Xu, H. Wang, X. Liu, R. Lin, X. Wei, and Q. Wang, "The comparison study of wavelet and wavelet packet analysis in the de-noising of EEG test Signal," in *Proc. 7th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, Nov. 2015, pp. 274–277.
- [2] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, I. Zunaidi, and D. Hazry, "EEG feature extraction for classifying emotions using FCM and FKM," *Int. J. Comput. Commun.*, vol. 1, no. 2, pp. 21–25, 2007.
- [3] G. Inuso, F. La Foresta, N. Mammone, and F. C. Morabito, "Wavelet-ICA methodology for efficient artifact removal from Electroencephalographic recordings," in *Proc. Int. Joint Conf. Neural Netw.*, Aug. 2007, pp. 1524–1529.
- [4] P. S. Kumar, R. Arumuganathan, K. Sivakumar, and C. Vimal, "A wavelet based statistical method for de-noising of ocular artifacts in EEG signals," *Int. J. Comput. Sci. Netw. Secur.*, vol. 8, no. 9, pp. 87–92, 2008.
- [5] M. I. Al-Kadi, M. B. I. Reaz, and M. A. M. Ali, "Compatibility of mother wavelet functions with the electroencephalographic signal," in *Proc. IEEE-EMBS Conf. Biomed. Eng. Sci.*, Dec. 2012, pp. 113–117.
- [6] N. K. Al-Qazzaz, S. Ali, S. A. Ahmad, M. S. Islam, and M. I. Ariff, "Selection of mother wavelets thresholding methods in denoising multi-channel EEG signals during working memory task," in *Proc. IEEE Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2014, pp. 214–219.
- [7] H. S. Hussain, M. B. I. Reaz, F. Mohd-Yasin, and M. I. Ibrahimy, "Electromyography signal analysis using wavelet transform and higher order statistics to determine muscle contraction," *Expert Syst.*, vol. 26, no. 1, pp. 35–48, Feb. 2009.
- [8] J. G. Servín-Aguilar, L. Rizo-Dominguez, and J. A. Pardiñas-Mir, "A comparison between wavelet families to compress an EEG signal," in *Proc. IEEE ANDESCON*, Oct. 2016, pp. 1–4.
- [9] G.-Z. Yan, B.-H. Yang, and S. Chen, "Automated and adaptive feature extraction for brain-computer interfaces by using wavelet packet," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Aug. 2006, pp. 4248–4251.
- [10] S. Khatun, R. Mahajan, and B. I. Morshed, "Comparative study of wavelet-based unsupervised ocular artifact removal techniques for single-channel EEG data," *IEEE J. Transl. Eng. Health Med.*, vol. 4, , 2016, Art. no. 2000108.
- [11] E. L. Lema-Condo, F. L. Bueno-Palomeque, S. E. Castro-Villalobos, E. F. Ordoñez-Morales, and L. J. Serpa-Andrade, "Comparison of wavelet transform symlets (2-10) and daubechies (2-10) for an electroencephalographic signal analysis," in *Proc. IEEE 24th Int. Conf. Electron., Elect. Eng. Comput. (INTERCON)*, Aug. 2017, pp. 1–4.
- [12] S. G. Eraldemir and E. Yildirim, "Comparison of wavelets for classification of cognitive EEG signals," in *Proc. 23rd Signal Process. Commun. Appl. Conf. (SIU)*, May 2015, pp. 1381–1384.
- [13] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, "Signal processing and compression with wavelet packets," in *Wavelets and Their Applications*. Dordrecht, The Netherlands: Springer, 1994, pp. 363–379.
- [14] E. Hernández and G. Weiss, *A First Course on Wavelets*. Boca Raton, FL, USA: CRC Press, 1996.

- [15] N. Khanna, V. Kumar, and S. K. Kaushik, "Wavelet packet approximation," *Integral Transforms Special Functions*, vol. 27, no. 9, pp. 698–714, 2016.
- [16] N. Khanna, V. Kumar, and S. K. Kaushik, "Vanishing moments of wavelet packets and wavelets associated with Riesz projectors," in *Proc. Int. Conf. Sampling Theory Appl. (SampTA)*, Jul. 2017, pp. 222–226.
- [17] N. Khanna, V. Kumar, and S. Kaushik, "Wavelet packets and their moments," *Poincare J. Anal. Appl.*, vol. 2017, no. 2, pp. 95–105, 2017.
- [18] A. M. Jarrah and N. Khanna, "Some results on vanishing moments of wavelet packets, convolution and cross-correlation of wavelets," *Arab J. Math. Sci.*, vol. 25, no. 2, pp. 169–179, 2019.
- [19] N. Khanna and S. Kaushik, "Wavelet packet approximation theorem for  $H^r$  type norm," *Integral Transforms Special Functions*, vol. 30, no. 3, pp. 231–239, 2019.
- [20] H. Cai, X. Sha, X. Han, S. Wei, and B. Hu, "Pervasive EEG diagnosis of depression using deep belief Network with three-electrodes EEG collector," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 1239–1246.
- [21] 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. Methodol. Commun. (ICCMC)*, Jul. 2017, pp. 985–988.
- [22] A. Graps, "An introduction to wavelets," *IEEE Comput. Sci. Eng.*, vol. 2, no. 2, pp. 50–61, Jun. 1995.
- [23] J. Rafiee, M. A. Rafiee, N. Prause, and M. P. Schoen, "Wavelet basis functions in biomedical signal processing," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 6190–6201, 2011.
- [24] S. Janjarasjitt, "Classification of the epileptic EEGs using the wavelet-based scale variance feature," *Int. J. Appl. Biomed. Eng.*, vol. 3, no. 1, pp. 19–25, 2010.
- [25] P. Jahankhani, V. Kodogiannis, and K. Revett, "EEG signal classification using wavelet feature extraction and neural networks," in *Proc. IEEE John Vincent Atanasoff Int. Symp. Mod. Comput.*, Oct. 2006, pp. 120–124.
- [26] N. Deng and C.-S. Jiang, "Selection of optimal wavelet basis for signal denoising," in *Proc. 9th Int. Conf. Fuzzy Syst. Knowl. Discovery*, May 2012, pp. 1939–1943.
- [27] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: SIAM, 1992.
- [28] N. Khanna and L. Kathuria, "On convolution of boas transform of wavelets," in *Proc. 8TH Int. Eurasian Conf.*, 2019, p. 83.
- [29] N. Khanna, S. Kaushik, and A. Jarrah, "Some remarks on Boas transforms of wavelets," *Integral Transforms Special Functions*, to be published.
- [30] N. Khanna, V. Kumar, and S. Kaushik, "Vanishing moments of Hilbert transform of wavelets," *Poincare J. Anal. Appl.*, vol. 2015, no. 2, pp. 115–127, 2015.
- [31] N. Khanna, V. Kumar, and S. Kaushik, "Approximations using Hilbert transform of wavelets," *J. Classical Anal.*, vol. 7, no. 2, pp. 83–91, 2015.
- [32] D. F. Walnut, *An Introduction to Wavelet Analysis*. New York, NY, USA: Springer, 2013.
- [33] N. Kalpakam and S. Venkataramanan, "Haar wavelet decomposition of EEG signal for ocular artifact de-noising: A mathematical analysis," in *Proc. 2nd Annu. IEEE Northeast Workshop Circuits Syst.*, Jun. 2004, pp. 141–144.
- [34] S. Shete and R. Shriram, "Comparison of sub-band decomposition and reconstruction of EEG signal by daubechies9 and symlet9 wavelet," in *Proc. 4th Int. Conf. Commun. Syst. Netw. Technol.*, Apr. 2014, pp. 856–861.
- [35] P. Zeman, S. Mahajan, P. Sorensen, and N. Livingston, "Biorthogonal 3.1 wavelet enhancement of brain activities," in *Proc. Can. Conf. Elect. Comput. Eng.*, Apr. 2007, pp. 1010–1013.
- [36] P. Swami, M. Bhatia, S. Anand, B. K. Panigrahi, and J. Santhosh, "SVM based automated EEG seizure detection using 'Coiflets' wavelet packets," in *Proc. Int. Conf. Recent Develop. Control, Automat. Power Eng. (RDCAPE)*, Mar. 2015, pp. 238–242.
- [37] Z. Mahmoodin, N. Jalaludin, W. Mansor, K. Y. Lee, and N. Mohamad, "Selection of Symlets wavelet function order for EEG signal feature extraction in children with dyslexia," in *Proc. IEEE Student Symp. Biomed. Eng. Sci. (ISSBES)*, Nov. 2015, pp. 113–117.
- [38] S. S. Patil and M. K. Pawar, "Quality advancement of EEG by wavelet denoising for biomedical analysis," in *Proc. Int. Conf. Commun., Inf. Comput. Technol. (ICCICT)*, Oct. 2012, pp. 1–6.
- [39] H. Abbasi, A. J. Gunn, L. Bennet, and C. P. Unsworth, "Reverse Bi-orthogonal wavelets & fuzzy classifiers for the automatic detection of spike waves in the EEG of the hypoxic ischemic pre-term fetal sheep," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 5404–5407.
- [40] 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.*, Feb. 2009, pp. 1015–1018.
- [41] 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, 2017.
- [42] X. An, D. Kuang, X. Guo, Y. Zhao, and L. He, "A deep learning method for classification of EEG data based on motor imagery," in *Proc. Int. Conf. Intell. Comput. Cham, Switzerland: Springer*, 2014, pp. 203–210.
- [43] A. Schlögl, 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, Oct./Nov. 1997, pp. 1533–1535.
- [44] 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, 1999.
- [45] G. Pfurtscheller, C. Neuper, A. Schlögl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 3, pp. 316–325, Sep. 1998.
- [46] 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.
- [47] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain-computer communication: Motivation, aim, and impact of exploring a virtual apartment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 4, pp. 473–482, Dec. 2007.
- [48] M. Fatourehchi, 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, 2007.
- [49] A. Schlögl, J. Kronegg, J. Huggins, and S. Mason, "19 evaluation criteria for BCI research," in *Toward Brain-Computer Interfacing*. Cambridge, MA, USA: MIT Press, 2007.
- [50] 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. Neurophys.*, vol. 118, no. 1, pp. 98–104, 2007.
- [51] M. S. Choudhry, R. Kapoor, A. Gupta, and B. Bharat, "A survey on different discrete wavelet transforms and thresholding techniques for EEG denoising," in *Proc. Int. Conf. Comput., Commun. Automat. (ICCCA)*, Apr. 2016, pp. 1048–1053.
- [52] 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.



**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 current research interests include machine learning, EEG signal processing, BCI, and robotics.





**DUANLING LI** received the M.S. degree from the Shaanxi University of Science and Technology, Shaanxi, China, in 1999, and the Ph.D. degree from Beihang University, Beijing, China, in 2003, all in mechanical engineering. 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.



**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 current research interests include robotics and BCI.

...



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