Biomedical Signal Processing and Control 60 (2020) 101991

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

A novel hybrid kernel function relevance vector machine for multi-task motor imagery EEG classification

Enzeng Dong [a,*], Kairui Zhou [a], Jigang Tong [a], Shengzhi Du [b]

[a] Tianjin Key Laboratory for Control Theory & Applications in Complicated Systems, Tianjin University of Technology, Tianjin 300384, China
[b] Department of Electrical Engineering, Tshwane University of Technology, Pretoria 0001, South Africa

ARTICLE INFO

Article history:
Received 15 July 2019
Received in revised form 6 March 2020
Accepted 18 April 2020

Keywords:
Motor imagery EEG
Phase space reconstruction
Common spatial pattern
Hybrid kernel function
Relevance vector machine

ABSTRACT

Relevance vector machine (RVM) is a sparse Bayesian probability model commonly utilized in classification problems. Kernel functions are critical for the classification capacity of RVM. The kernel functions of RVM are not limited by the Mercer theorem, which differs RVM from the traditional support vector machine (SVM). As a typical local kernel function, the Gaussian kernel function has strong interpolating capacity, while the polynomial kernel function, a representative of global kernel function, is good at extrapolation. By combining the Gaussian kernel function and the polynomial kernel function, this paper proposed a novel hybrid kernel function RVM which is effective for classification at both local and global feature levels. Multi-task motor imagery electroencephalogram (EEG) classification is considered to validate the proposed method. Firstly, the phase space reconstruction (PSR) is employed to project EEG data from the time domain into the high-dimensional phase space, where the phase space common spatial pattern (PSCSP) features are extracted by using the "one versus one" common spatial pattern (OVO-CSP) strategy. Then the obtained PSCSP features are utilized as the input feature vectors of the proposed hybrid kernel function RVM for classification. The experimental results show that the proposed method improves the accuracy and Kappa coefficient for the multi-task motor imagery EEG classification problem. The main contributions of the paper include the novel hybrid kernel function RVM and the PSCSP features extracted from EEG.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Brain-computer interface (BCI) provides the communication between the brain and computers or other external devices without connection of actual peripheral nerves or muscles. Due to its lower price and convenience on signal measurement, EEG based BCI is commonly considered to improve the living quality of disabled groups and providing more help for medical rehabilitation. However, because of the influences from hardware and external environment, EEG based BCI has the weakness of low signal-to-noise ratio (SNR), and significant individual differences, which leads to low control accuracy and poor real-time performance (for instance, the transmission rate). Towards higher performance, two main branches of research are developed in BCI technology. One is to obtain more effective feature algorithms, and the other focuses on more effective classification algorithms.

Many methods for EEG feature extraction were proposed. For instance, EEG signal was transformed from the time domain to the frequency domain using fast Fourier transform (FFT) [1,2], where the peak characteristics of different frequency bands were observed. The wavelet coefficient of EEG was extracted using wavelet transform (WT) [3,4], and the correspondence relationship between the signal in the time domain and the frequency domain was established. Autoregressive (AR) model was employed to analyze univariate time series and predict the model of epilepsy EEG time series [5,6]. To reduce the computation cost, the dimension of the feature space was reduced using principal component analysis (PCA) [7,8]. The independent component analysis (ICA) [9,10] were developed to remove the artifact of electrooculogram (EOG) components. Common spatial pattern (CSP) is an effective method to extract the variance characteristics of EEG with different motor imagery tasks [11]. A quadcopter control in three-dimensional space [12] was proposed based on motor imagery sensorimotor rhythms (SMRs). A new approach was proposed by combining Mu/Beta rhythm during motor imagery and P300 potential for the horizontal and the vertical movements of the 2-D cursor [13]. In reference [14], EEG is employed to record and decode sensorimo-

* Corresponding author.
E-mail address: dongenzeng@163.com (E. Dong).

https://doi.org/10.1016/j.bspc.2020.101991
1746-8094/© 2020 Elsevier Ltd. All rights reserved.

tor rhythms (SMRs) induced from motor imagination to achieve three-dimensional movement of a virtual helicopter. A sparse filter band common spatial pattern (SFBCSP) [15] was developed for optimizing the spatial patterns, which suggests that the SFBCSP is a potential method for improving the performance of MI-based BCI. The temporally constrained sparse group spatial pattern (TSGSP) [16] was developed for the simultaneous optimization of filter bands and time windows within CSP to further boost classification accuracy of MI-based BCI.

Classification for the EEG BCI extracts wide attention, common classifiers include: linear discriminate analysis (LDA) based on statistical laws [17], support vector machine (SVM) based on structural risk minimization [18], and relevance vector machine (RVM) based on sparse Bayesian probability model [19]. With probabilistic inference, reference [20] introduced a sparse Bayesian ELM-based algorithm (SBELM) to improve the classification performance of MI. SBELM is able to automatically control the model complexity and exclude redundant hidden neurons by combining advantageous of both ELM and sparse Bayesian learning. Reference [21] proposed a sparse Bayesian method by exploiting Laplace priors, namely, SBLaplace, for EEG classification. All required model parameters are automatically estimated from training data without the need of cross validation. As a machine learning model, the RVM is featured with sparsity, strong generalization ability, automatic determination of parameters, and no limitation on Mercer theorem [35] in kernel function selection.

EEG is non-stationary signal with some random characteristics. The phase space reconstruction (PSR) method was proposed to reconstruct the low-dimensional signal into the high-dimensional phase space, to obtain rich dynamic characteristics [22–29], then the filter structure of PSR was developed to obtain more useful information [30].

The kernel functions commonly used by classifiers fall into two categories, local kernel functions and global kernel functions. The local kernel function has better interpolating characteristics, while the global kernel function has strong extrapolation characteristics. Single kernel functions have certain limitations when classifiers do data classification processing. This paper proposed a new hybrid kernel function to fit both local (Gaussian) and global (polynomial) kernel functions for RVM. Four-task motor imagery EEG classification is used to validate the proposed method. The ICA is used to obtain the independent components of each channel EEG, and removing the artifacts. Low dimensional EEG time series are projected into a high-dimensional phase space via PSR. Therefore, more hidden dynamic characteristics of the EEG can be observed in the phase space. The filter matrix constructed by CSP is used to calculate the variance characteristic as the input vector of the classifier. Finally, a new RVM classifier with hybrid kernel function is fitted to classify four-task motor imagery EEG. The simulation results shown that a good classification performance is achieved by the hybrid kernel function RVM based on extracted phase space CSP (PSCSP) features.

2. Dataset

The dataset used in this paper is from the BCI competition IV dataset 2a [31]. The dataset was performed under experimental paradigms of four motor imagery tasks (left hand, right hand, both feet and tongue). EEG signals were collected from 9 subjects. All subjects completed the data collection within different days, and each subject performed two-session experiments. In one session, the EEG data is measured as a training set, and the other as estimating (testing) set. The subjects in each session conducted 6 rounds of experiments, and 48 trials of data were collected for each round. Four-task motor imagery EEG were carried out in 12 groups. Each subject measured 288 trials of experimental data totally. The EEG is acquired through 22 electrodes, and the left ear lobe and the right ear lobe are used as reference electrodes and ground, respectively. The electrode montage corresponding to the international 10–20 system is shown in Fig. 1. The signal sampling frequency is 250 Hz, and the signal is subjected to bandpass filtering of 0.5–100 Hz.

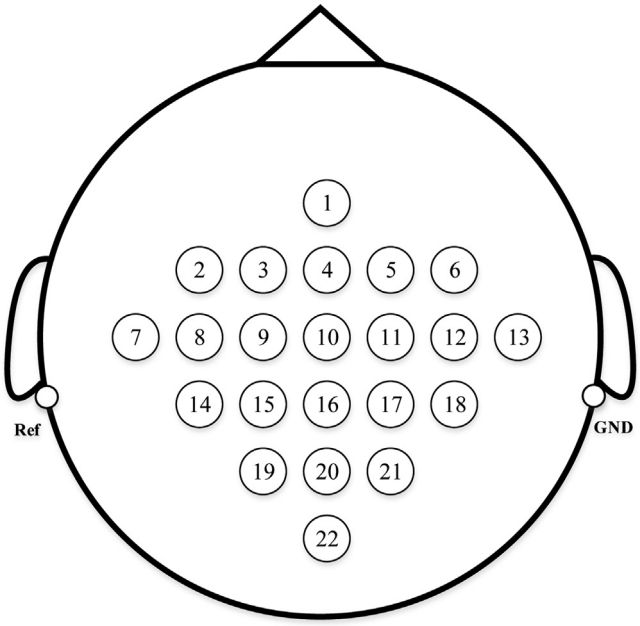

Fig. 1. Electrode montage corresponding to the international 10–20 system [24].

Each subject was seated in a comfortable chair facing to a computer screen. At t = 0 s, the experiment starts with a fixation cross appears on the screen, and a short beep stimulates the subject to start the experiment; at t = 2 s, a directional arrow appears on the screen (left, right, up or downwards) and will remain on the screen for 1.25 s, at the same time, the subject is reminded to start preparing for a motor imagery task that prompts the direction arrow. The EEG recording from t = 3 s, and the motor imagery continued until t = 6 s. The next experiment was performed after a short interval. The experimental timing diagram is shown in Fig. 2.

3. Method

3.1. Preprocessing

3.1.1. Independent component analysis

Independent component analysis (ICA) is a blind source separation algorithm that recovers statistically independent signals from linearly hybrid signals using high-order statistic measures [32]. The artifacts and noise in the EEG signal can be culled by the ICA, retaining useful EEG information. The original EEG signal contains an EOG signal, and the artifact relevant to EOG existing in the EEG channels can be removed by ICA.

For the original EEG signal $X_{N\times T}$, where N is the number of channels, T is the number of sampling points, A is the mixing matrix, it can be considered as the linear hybrid signal from the mutual independent source signal $Y_{N\times T}$, as shown in Eq. (1).

$$X = AY \tag{1}$$

where $W = A^{-1}$ is the de-mixing matrix. Y is unknown, but if W can be estimated, then one will get the mutual independent components Z from the original hybrid EEG signal X as shown in Eq. (2).

$$Z = WX = WAY \tag{2}$$

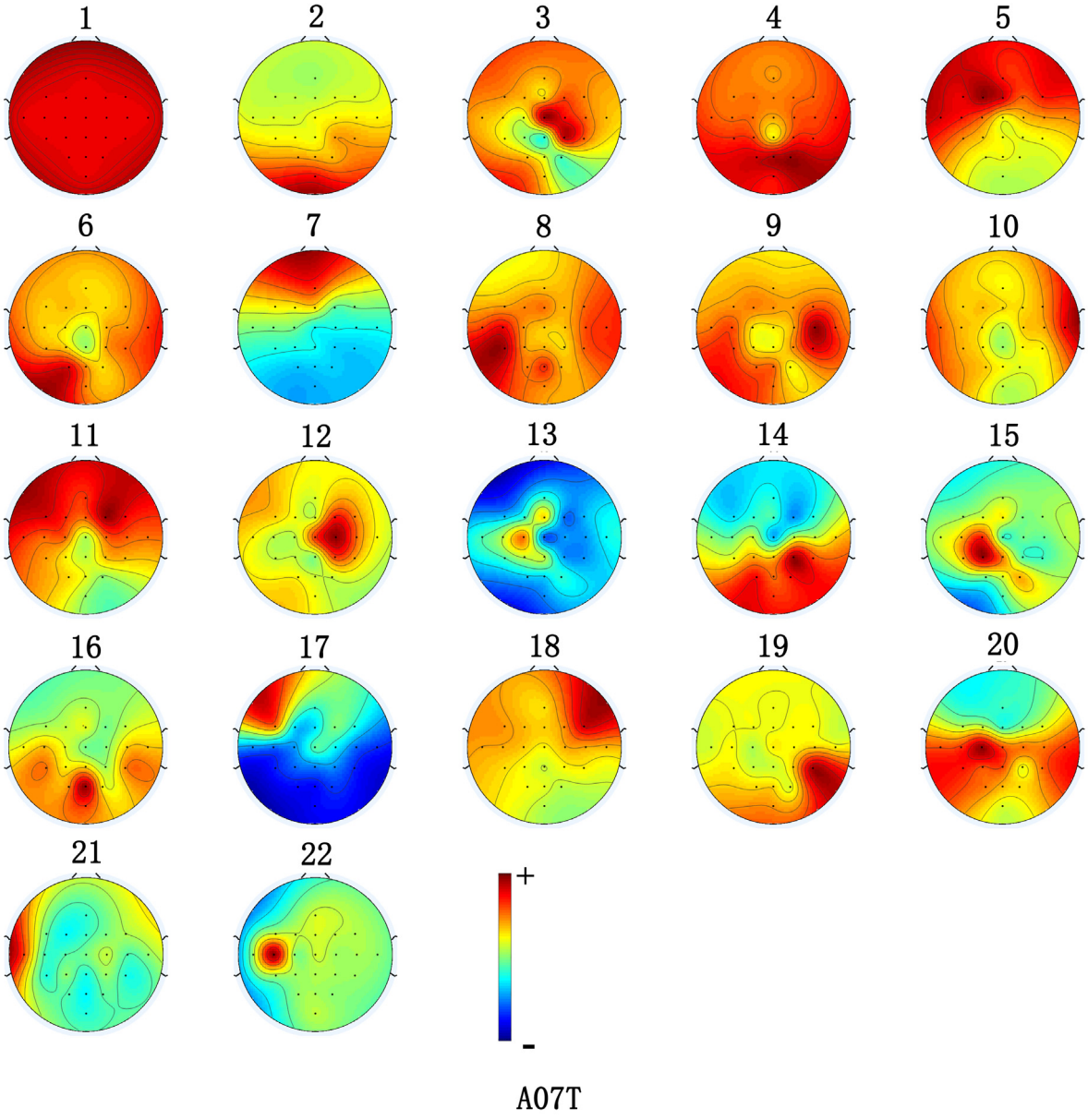

Fig. 2. Timing diagram of the motor imagery experiment [24].

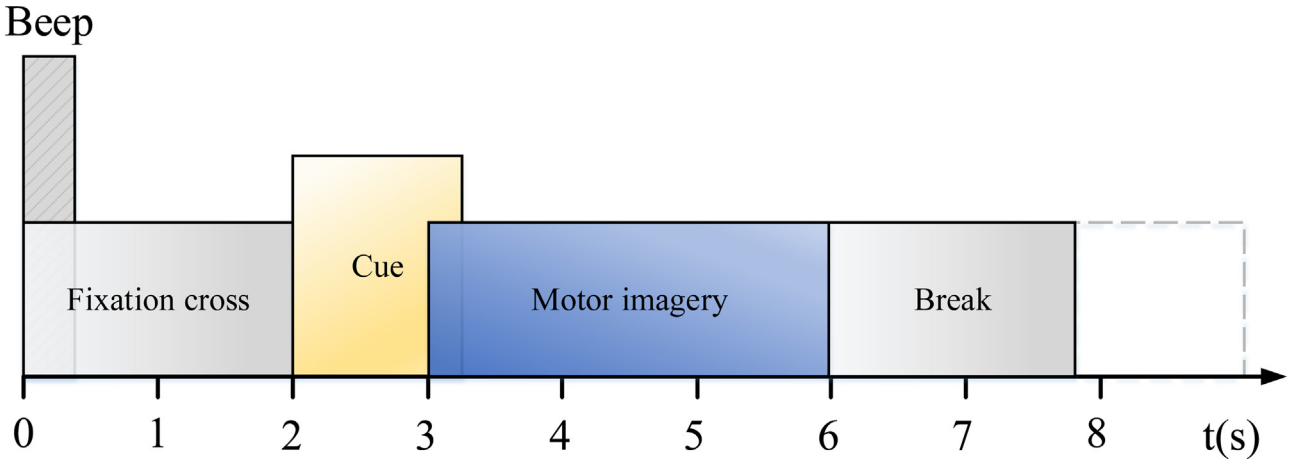

Fig. 3. Brain topographic map of 22 channels independent components.

In BCI competition IV dataset 2a, 22 channels of EEG signals were acquired from each subject, and the hybrid matrix $A_{22\times22}$ obtained by the ICA algorithm generates 22 independent components for each sample. The EEG signals after the ICA de-mixing get the EOG artifacts and noise signals removed, which more accurately describe the electrophysiological activities of different brain regions. Fig. 3 is a brain topographic map of 22 independent components. For instance, the seventh independent component of the topographic map in the forehead position can be clearly observed as the presence of EOG artifacts.

3.1.2. Phase space reconstruction

When the nonlinearity of EEG time series is concerned, the phase space reconstruction (PSR) [33] is commonly employed. Using PSR, one or more time series generated by a singular attractor can be reconstructed into a new attractor, which was proved that the new attracting subsystem has the same dynamics as the original system [34]. A d-dimensional attractor can be reconstructed into the m-dimensional phase space by the delay theorem, where $m > 2d$.

According to the delay theorem, a one-dimensional EEG time series $x(t) = \{x_1, x_2, x_3, \cdots, x_N\}$ with the length of N can be reconstructed into m-dimensional phase space. The concrete expression of the reconstructed phase space trajectory is as shown Eq. (3):

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} = \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(m-1)\tau} \\ x_2 & x_{2+\tau} & \cdots & x_{2+(m-1)\tau} \\ \vdots & \vdots & \vdots & \vdots \\ x_M & x_{M+\tau} & \cdots & x_{M+(m-1)\tau} \end{bmatrix} \tag{3}$$

where $M = N - (m-1)\tau$, τ is the delay time, and m is the embedding dimension. The correlation integral from the C-C method [35] is used to obtain the values of the delay time and the embedding dimension.

3.1.3. C-C method

The correlation integral, for a one-dimensional EEG time series $x(t) = \{x_1, x_2, x_3, \cdots, x_N\}$, can be presented as Eq. (4),

$$C(m, N, r, t) = \frac{2}{M(M-1)} \sum_{1\le i\le j\le M} \Theta(r - \left\|x_i - x_j\right\|),\ r > 0 \tag{4}$$

where $\Theta(a) = \begin{cases} 0, & if\ a < 0 \\ 1, & if\ a \ge 0 \end{cases}$, N is the size of the data set, t is the index lag, $M = N - (m-1)t$ is the number of embedded points in m-dimensional space, and $\|\cdots\|$ denotes the sup-norm.

Thus, the embedding dimension and delay time can be acquired with the properties of

$$S(m, N, r, t) = C(m, N, r, t) - C^m(1, N, r, t) \tag{5}$$

For general time series, it consists of t disjoint time series.

$$S(m, N, r, t) = \frac{1}{t}\sum_{s=1}^{t} C_s\left(m, \frac{N}{t}, r, t\right) - C_s^m\left(1, \frac{N}{t}, r, t\right) \tag{6}$$

Finally, as $N \to \infty$, one gets

$$S(m, r, t) = \frac{1}{t}\sum_{s=1}^{t} C_s(m, r, t) - C_s^m(1, r, t),\ m = 2, 3, \cdots. \tag{7}$$

From Eq. (7), for fixed m and τ, $S(m, r, t)$ will be equal to 0 for all r values, if the data are independent and identically distributed (IID). However, real data sets are usually finite, and the data may be serially correlated. Therefore, generally we will have $S(m, r, t)$ not equal to 0. Thus, the local optimal occurs either on the zero crossing points of $S(m, r, t)$ or the points when $S(m, r, t)$ has the least variation with r, which indicates a nearly uniform distribution of data. Hence, one can select several representative values r_j, and define the quantity

$$\Delta S(m, t) = max\left\{S\left(m, r_j, t\right)\right\} - min\left\{S\left(m, r_j, t\right)\right\} \tag{8}$$

which is a measure of the variation of $S(m, r, t)$ with respect to r. The local optimum occurs on the zero crossings of $S(m, r, t)$ and the minima of $\Delta S(m, t)$. The zero crossings of $S(m, r, t)$ should be nearly the same for all m and r, and the minima of $\Delta S(m, t)$ should be nearly the same for all m (otherwise, the time is not locally optimal). The delay time τ_d is the first local optimal time.

For the one-dimensional EEG signal, it can be reconstructed into the high-dimensional phase space by phase space reconstruction, the characteristic dimension of the signal is increased, more useful dynamic characteristic information can be obtained, and the signal separability can be increased.

3.2. Feature extraction

Common spatial pattern (CSP) is a feature extraction method that uses a spatial filter to project signal into space to maximize the differentiation of variance features, which is commonly used in processing motor imagery EEG signals [11]. By constructing a set of spatial filter matrices, the CSP analysis projects two classes of signals onto the planes, in the way of maximizing the variance of one class of signal and minimizing the variance of the other class of signal. Therefore, the variance characteristics (CSP feature) increase the differences between the two classes of signals. The CSP feature vectors can be considered as the input of classifiers for higher classification accuracy.

For two classes of EEG signals X_i, $i \in \{1, 2\}$, the dimension of the signal is $N \times T$, where N is the number of channels and T is the number of samples in the time series. The normalized covariance of X_i can be represented as:

$$C_i = \frac{x_i x_i^T}{trace(x_i x_i^T)} \tag{9}$$

The mean of the covariance C_i is denoted as $\bar{C}_i$, then one has

$$C = \bar{C}_1 + \bar{C}_2 = U_0 \Sigma U_0^T \tag{10}$$

where U_0 is the matrix of eigenvector and Σ is the diagonal matrix of eigenvalues. The whitening transformation matrix is

$$P = \Sigma^{-1/2} U_0^T \tag{11}$$

Transforms the average covariance matrix, one gets

$$S_i = P\bar{C}_i P^T \quad i \in \{1, 2\} \tag{12}$$

S_i shares common eigenvectors and the sum of corresponding eigenvalues for the two matrices will always be 1,

$$S_1 = U\Sigma_1 U^T \tag{13}$$

$$S_2 = U\Sigma_2 U^T \tag{14}$$

$$\Sigma_1 + \Sigma_2 = I \tag{15}$$

The projection matrix W is

$$W = U^T P \tag{16}$$

With the projection matrix W, the original EEG can be transformed into uncorrelated components

$$Z = WX \tag{17}$$

where Z can be seen as EEG source components including common and specific of different tasks.

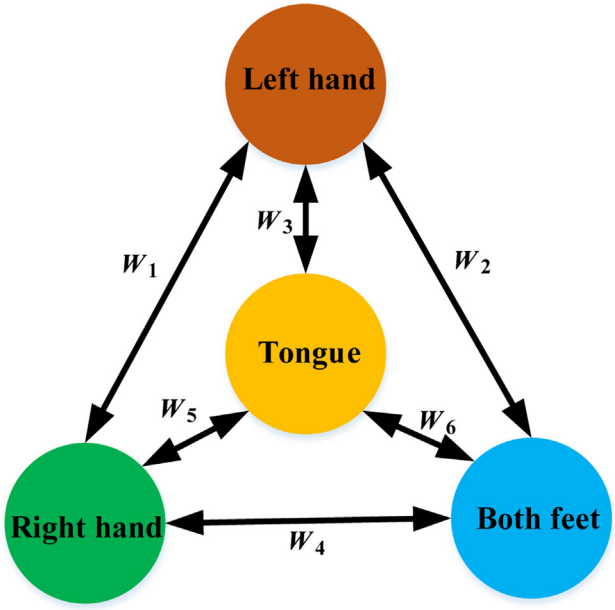

Fig. 4. "One versus one" CSP feature extraction strategy for four-task motor imagery EEG.

The feature of the *k-th* row in the projection matrix Z is calculated as

$$f_k = log\left(\frac{var(Z)}{\sum_{k=1}^{2m} var(Z)}\right) \tag{18}$$

where $\{f_k : k = 1, 2, \cdots, 2m\}$ is the EEG feature of different tasks as the input signal of the classifier.

For four-task motor imagery EEG signal, the "one versus one" CSP feature extraction algorithm was used to extract four-task motor imagery EEG features [19]. Six sets of spatial filters are constructed for six sets of CSP features (i.e. left hand vs right hand, left hand vs both feet, left hand vs tongue, right hand vs both feet, right hand vs tongue, both feet vs tongue) as input vectors to the classifier. Fig. 4 is the "one versus one" CSP feature extraction strategy.

3.3. Classification

Relevance vector machine (RVM) is a machine learning model based on sparse Bayesian probability model [36]. Differing from the support vector machine (SVM) based on the principle of structural risk minimization, the RVM does not have the limitation of Mercer theorem [37] in the selection of kernel functions, so one can arbitrarily construct kernel functions. The sparseness of RVM reduces the computation load in offline training and improves the real-time performance of online applications. Moreover, the RVM has higher generalization ability than the SVM.

Given a set of input data $\{x_n\}_{n=1}^N$ and the target vector $\{t_n\}_{n=1}^N$, the purpose of the RVM machine learning is to predict t according to the input data x. For two-class classification, it is desired to predict the posterior probability of membership of one of the classes given the input x. The classification function is based on the form as:

$$y(x) = \sum_{i=1}^{N} w_i K(x, x_i) + w_0 \tag{19}$$

where w_i denotes the weight vector, and w_0 is the bias, and $K(x, x_i)$ is the kernel function. We follow statistical convention and generalize the linear model by applying the logistic sigmoid link function $\sigma(y) = 1/(1 + e^{-y})$ to $y(x)$ and, adopting the Bernoulli distribution for $P(t|x)$, we write the likelihood as:

$$P(tw) = \prod_{n=1}^{N} \sigma\left\{y(x_n; w)\right\}^{t_n} \left[1 - \sigma\left\{y(x_n; w)\right\}\right]^{1-t_n} \tag{20}$$

where, following from the probabilistic specification, $w = (w_0, w_2, \cdots, w_N)$, the targets is $t_n \in \{0, 1\}$. Note that there is no 'noise' variance here. The maximum likelihood estimation of w in Eq. (20) often results in overfitting. Therefore, it was recommended that the imposition of some prior constrains on the parameters w should be considered by adding a complexity to the likelihood or error function [36]. This priori information controls the generalization ability of the learning process. Typically, higher-level parameters are used to constrain an explicit zero-mean Gaussian prior probability distribution over the weights, as shown in Eq. (21).

$$p(w\alpha) = \prod_{i=0}^{N} N(w_i|0, \alpha_i^{-1}) \tag{21}$$

where α is a vector of $N + 1$ hyperparameters that controls how far from zero each weight is allowed to deviate.

However, we cannot integrate out the weights analytically. We thus choose to utilize the following approximation procedure, which is based on Laplace's method:

(1) For the current, fixed, values of α, the 'most probable' weights w_{MP} are found, giving the location of the mode of the posterior distribution. Since $p(w|t, \alpha) \propto P(t|w)p(w|\alpha)$, this is equivalent to finding the maximum, over w, of

$$\log\{P(t|w)p(w|\alpha)\} = \sum_{n=1}^{N} [t_n \log y_n + (1 - t_n)\log(1 - y_n)] - \frac{1}{2} w^T A w \tag{22}$$

with $y_n = \sigma\left\{y(x_n; w)\right\}$, $A = diag(\alpha_0, \alpha_2, \cdots, \alpha_N)$. This is a standard procedure, since Eq. (22) is a penalized (regularized) logistic log-likelihood function, and necessitates iterative maximization.

(2) Laplace's method is simply a quadratic approximation to the log-posterior around its mode. The quantity Eq. (22) is differentiated twice to give:

$$\nabla_w \nabla_w \log p(w|t, \alpha)|_{w_{MP}} = -(\Phi^T B \Phi + A) \tag{23}$$

where $B = diag\left(\beta_1, \beta_2, \cdots, \beta_N\right)$ is a diagonal matrix with $\beta_n = \sigma\left\{y(x_n)\right\}\left[1 - \sigma\left\{y(x_n)\right\}\right]$, and Φ is the $N \times (N + 1)$ 'design' matrix with $\Phi_{nm} = K(x_n, x_{m-1})$ and $\Phi_{n1} = 1$. This is then negated and inverted to give the covariance Σ for a Gaussian approximation to the posterior over weights centred at w_{MP}.

(3) Using the statistics Σ and w_{MP} of the Gaussian approximation, the hyperparameters α are updated until get suitable convergence [36] as is shown in Fig. 5.

$$\alpha_i^{new} = \frac{\gamma_i}{w_{MP_i}^2} \tag{24}$$

where $\gamma_i \equiv 1 - \alpha_i \Sigma_{ii}$, with Σ_{ii} the i-th diagonal element of the posterior weight covariance. For α_i large, where w_i is highly constrained by the prior, $\Sigma_{ii} \approx \alpha_i^{-1}$ and it follows that $\gamma_i \approx 0$. Conversely, when α_i is small and w_i fits the data, $\gamma_i \approx 1$.

$$\Sigma = \left(\Phi^T B \Phi + A\right)^{-1} \tag{25}$$

$$w_{MP} = \Sigma \Phi^T B t \tag{26}$$

The RVM does not have the limitation of the Mercer theorem [37], so any arbitrary kernel function can be constructed. In this paper, the local kernel function Gaussian kernel function and the global kernel function polynomial kernel function are selected [38].

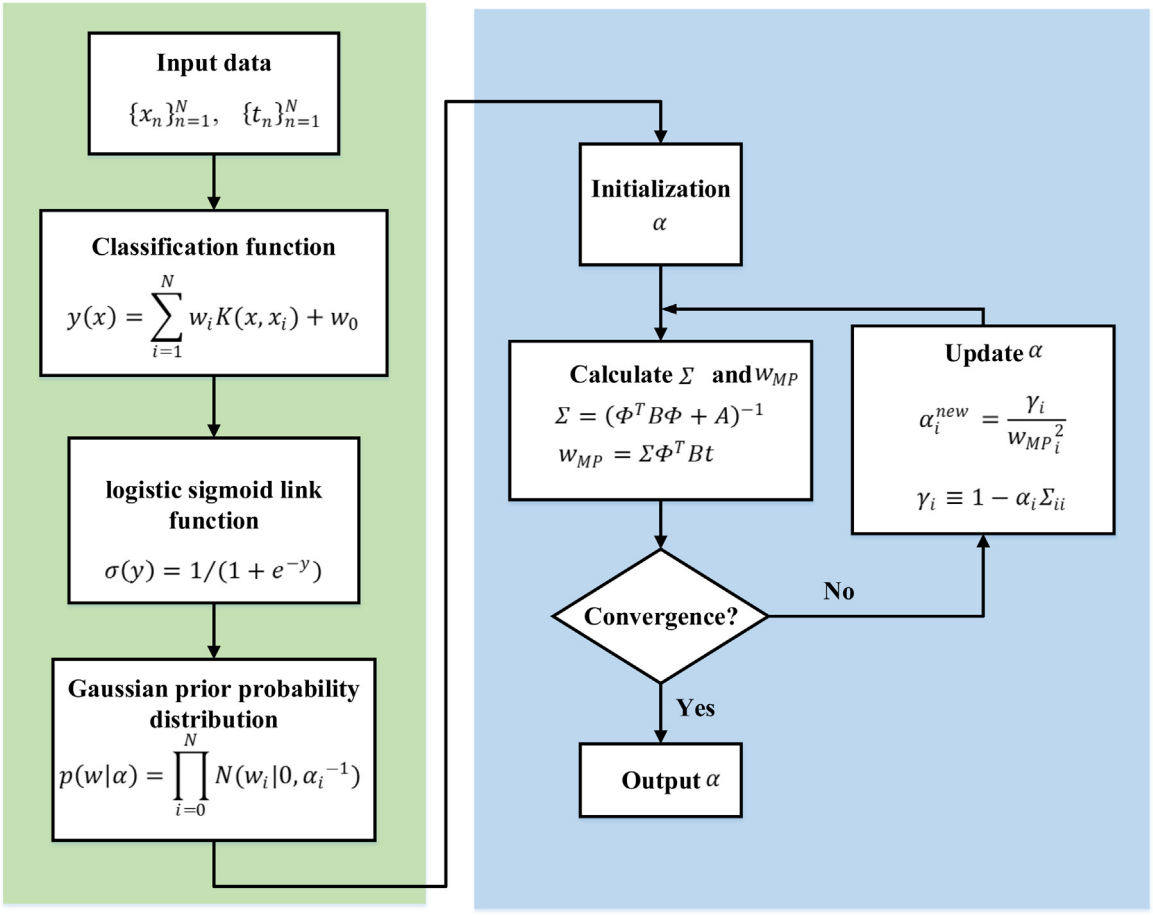

Fig. 5. Calculating the hyperparameters of RVM.

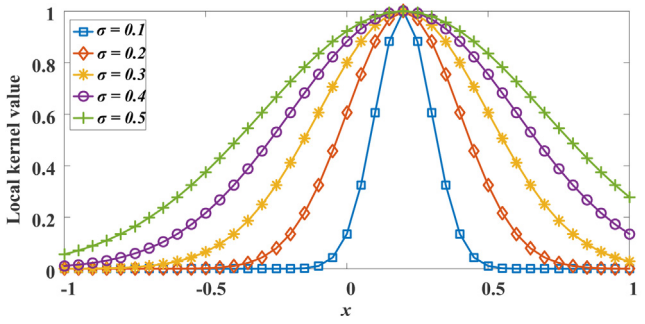

Fig. 6. Plot of Gaussian kernel function.

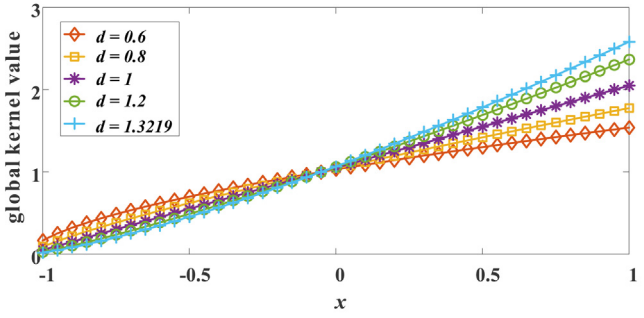

Fig. 7. Plot of polynomial kernel function.

By analyzing and comparing the characteristics of the two kernel functions, a new hybrid kernel function is fitted.

3.3.1. Kernel function

(1) Gaussian kernel function

The Gaussian kernel function with different σ parameters are shown in Eq. (27), and depicted in Fig. 6.

$$K_{gauss}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{27}$$

In Fig. 6, $x = 0.2$ is selected as the test point, one finds that around $x = 0.2$ the Gaussian kernel function has a greater amplitude gain, which means larger influence on the data near the test point, but smaller influence on the data far from the test point. As the σ increases, the local characteristics of the Gaussian kernel function decrease. The local characteristics of the gauss kernel function are relevant to the interpolating ability. The more significant local characteristics usually result in higher interpolating ability.

(2) Polynomial kernel function

The polynomial kernel function with different d parameters are shown in Eq. (28), and depicted in Fig. 7.

$$K_{poly}\left(x_i, x_j\right) = \left(x_i^T x_j + 1\right)^d \tag{28}$$

The $x = 0.2$ test point is considered to compare with the Gaussian kernel function. As shown in Fig. 7, the polynomial kernel function has a significant impact on both the data near the test point and the one away from the test point, which means the global characteristics of the polynomial kernel function is more obvious. As the d decreases, the global characteristics of the kernel function becomes more obvious. For instance, when $d = 0.4$, the global characteristics

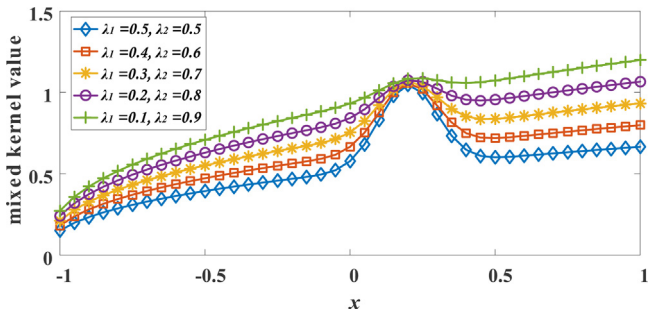

Fig. 8. Plot of hybrid kernel function.

of the polynomial kernel function are the most obvious, and the extrapolation ability is higher.

(3) Hybrid kernel function

The characteristics of the local (Gaussian) and global (polynomial) kernel function are compared in the above sections. In order to obtain both interpolation and extrapolation capacity, a hybrid kernel function of the Gaussian kernel function K_{gauss} ($\sigma = 0.1$) and the polynomial kernel function K_{poly} ($d = 0.4$) is proposed as a new kernel function K_{hybrid} as shown in Eq. (29).

$$K_{hybrid} = \lambda_1 K_{gauss} + \lambda_2 K_{poly} \tag{29}$$

Where $\lambda_1 + \lambda_2 = 1$, K_{hybrid} can be used as a kernel function for the RVM correlation vector machine classifier. Fig. 8 depicts the K_{hybrid} of different weight combinations. It can be seen from Fig. 8 that as the weight of the Gaussian kernel function increases, the weight of the polynomial kernel function decreases, the local characteristics of the hybrid kernel function become prominent, and the global characteristics decrease. Contrary to this, as the weight of the polynomial kernel function increases, the weight of the Gaussian kernel function decreases, the global characteristics of the hybrid kernel function become more obvious, while the local characteristics are reduced.

Based on the above analysis, this paper uses the relevance vector machine as the classifier, and designs a hybrid kernel function by combining the Gaussian and the polynomial kernel functions, as shown in Eq. (30).

$$y(x, w) = \sum_{i=1}^{N} w_i K_{hybrid}(x, x_i) + w_0 \tag{30}$$

Where x is the input vector with N samples and y is the output of prediction, w_0 is the bias, $K_{hybrid}(x, x_i)$ and w_i are the kernel function and its weight, respectively.

4. Experimental results and discussion

4.1. Experimental data and preprocessing

Fig. 9 is the EEG signal processing flow chart. The four-task motor imagery datasets used in this paper are from the BCI competition IV dataset 2a. First, the original EEG signal is filtered by a 0.5–30 Hz filter through a 6th-order Butterworth filter, and the noise is removed to leave our desired frequency band EEG. Then the EEGLAB toolbox was used to perform independent component analysis on the EEG signal, and obtain the independent components of 22 channels to decouple the original hybrid signal with a signal dimension of $22 \times 750 \times 72$.

The decoupled EEG is reconstructed to obtain a new set of reconstructed EEG signal matrix Z_r. The selection of embedding dimension is 3 and delay time is 6. The reconstructed EEG dimension is $66 \times 738 \times 72$. Table 1 shows the data before and after PSR. The spatial filter matrix W^2 obtained by the common spatial pattern selects the first two rows and the last two rows of the filter matrix to map the reconstructed EEG to obtain a maximum variance of one task of EEG and a minimum variance of another tasks of EEG. Then the phase space common spatial pattern (PSCSP) features as the input vectors of the RVM classifier for prediction.

The EEG signal processing algorithm flow is shown in Table 2.

Table 1
Description about dataset before and after PSR.

	m = 3; τ = 6			
	Channel	Trial	Sample	Butterworth filter
Before PSR	22	72	750	0.5–30Hz
PSR	66	72	738	0.5–30Hz

Table 2
EEG signal processing algorithm flow.

Four-task motor imagery EEG signal feature extraction and classification algorithm flow
Step1: Load the raw EEG data X_{raw}.
Step2: Preprocessing
(1) 6th-order Butterworth filtering for independent components of 22 channels.
(2) EEGLAB performs ICA on the data X, obtains 22 channels of independent components, and removes the EOG artifacts, the useful EEG signal of the independent source signal $Y_{22\times750\times72}$.
Step3: Calculating the embedding dimension $m = 3$ and delay time $\tau = 6$ with C-C method.
Step4: For four-task motor imagery EEG, constructing 6 spatial filters with the "OVO"strategy, obtain 6 feature vectors as the input vectors for classifier.
Step5: Designing hybrid kernel function
(1) Gaussian kernel function (local kernel function) characteristic analysis, determining parameter σ. $K_{gauss}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$
(2) Polynomial kernel function (global kernel function) characteristic analysis, determining parameter d. $K_{poly}\left(x_i, x_j\right) = (x_i^T x_j + 1)^d)$
(3) Hybrid kernel function characteristic analysis, determining weights combination (λ_1, λ_2). $(K_{hybrid} = \lambda_1 K_{gauss} + \lambda_2 K_{poly})$
Step6: Classification result
80% of the dataset is used as training data, and the remaining 20% is used as test data. The CSP feature vector as the input vector into the hybrid kernel function relevance vector machine classifier, and 10×10 cross-validation is performed to output the prediction result. The final classification result of four-task motor imagery EEG is obtained by voting method.

Table 3
Prediction result of two-class with TN, TP, FN, FP.

Real label	Predict label	
0	0(TN)	1(FP)
1	0(FN)	1(TP)

4.2. Evaluation criteria

4.2.1. Precision and classification accuracy

For a two-class (class = 0, 1) problem, the sample label (class = 1) as a positive class and (class = 0) as a negative class. There are four cases for the classification result of the two-class problem, if the real class label is 1, the prediction class label is 1, and then this is recorded as a True Positive (TP). While if the prediction class label is 0, then this is recorded as a False Negative (FN). If the real class label is 0, the prediction class label is 0, then this is recorded as True Negative (TN). While if the prediction class label is 1, then

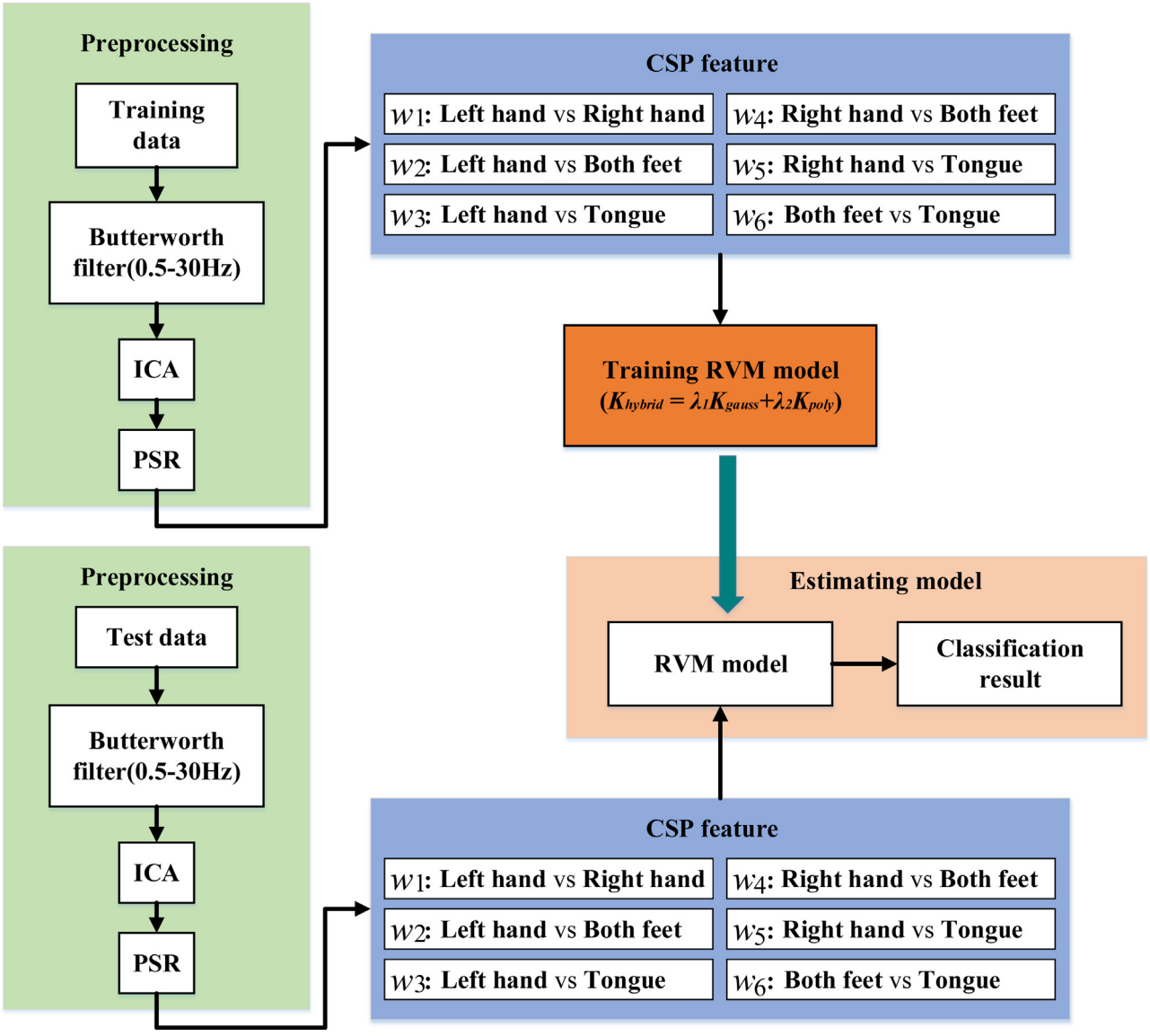

Fig. 9. EEG signal processing flow chart.

this is recorded as a False Positive (FP). As is shown in Table 3. Classification precision is:

$$Precision = \begin{cases} \dfrac{TP}{TP+FN}\,(label=1) \\ \dfrac{TN}{TN+FP}\,(label=0) \end{cases} \tag{31}$$

Classification accuracy of all sample (class = 0, 1):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{32}$$

4.2.2. Kappa coefficient

Jacob Cohen proposed the kappa coefficient to measure the confidence of the classification accuracy [39]. The calculation of the kappa coefficient requires two parameters, one is the overall agreement P_a, which is equal to the classification accuracy, and the chance agreement P_e. The estimate of kappa coefficient value k is:

$$k = \frac{P_a - P_e}{1 - P_e} \tag{33}$$

The kappa coefficient value is zero if the predicted classes show no correlation with the real classes. A kappa coefficient value of 1 indicates perfect classification. Kappa coefficient values smaller than zero indicate that the classifier suggests a different assignment between output and the true classes.

4.3. Experiments

4.3.1. Validation of the proposed kernel function

This experiment aims to demonstrate the performance of the proposed kernel function. The original motor imaginary EEG data is classified without PSR or ICA. In the BCI competition IV dataset 2a, the classification results of the hybrid kernel functions under different weight combinations are analyzed. Table 4 listed the four-task motor imagery EEG classification result, tested by 10–10 fold cross validation. In Table 4, the maximum average accuracy is highlighted by bold font. It can be seen that the average accuracy of Gaussian and polynomial kernel functions are 60.56% and 60.93%, respectively. The hybrid kernels with different weight combinations get higher accuracy. The classification accuracy of hybrid kernels is 62.86% (0.5, 0.5), 63.25% (0.4, 0.6), 62.70% (0.3, 0.7), 63.08% (0.2, 0.8) and 62.86% (0.1, 0.9), respectively. The classification accuracy of hybrid kernels with weight combinations (0.4, 0.6) is the highest.

Table 4
Classification accuracy of the hybrid kernel functions under different weight combination (acc ± SD %).

Subject	Kernel $\left(\lambda_1 K_{gauss} + \lambda_2 K_{poly}\right)$						
	K_{gauss}	K_{poly}	(0.5, 0.5)	(0.4, 0.6)	(0.3, 0.7)	(0.2, 0.8)	(0.1, 0.9)
A01T	68.55	68.43	70.00	70.54	70.89	**71.79**	70.54
A02T	57.16	56.79	59.29	59.64	58.93	**60.00**	59.46
A03T	80.43	81.32	83.04	82.32	**83.21**	82.14	82.50
A04T	50.41	51.02	53.93	54.11	53.39	**54.46**	52.86
A05T	36.00	36.52	38.04	**40.18**	37.14	37.86	39.29
A06T	42.42	41.50	44.64	44.64	43.75	**45.18**	44.29
A07T	74.05	74.98	76.25	76.25	**77.86**	75.89	76.96
A08T	81.18	81.34	81.79	82.50	81.25	82.32	**82.86**
A09T	54.84	56.45	57.14	59.11	57.86	58.04	**59.69**
Average	60.56 ± 16.38	60.93 ± 16.56	62.86 ± 16.08	**63.25 ± 15.61**	62.70 ± 16.58	63.08 ± 15.97	62.86 ± 16.16

Table 5
Comprehensive comparison of RVM based on Gaussian, polynomial and hybrid kernels (0.4Gauss+0.6Poly) of A08T.

Kernel	Training time (s)	RVs	Accuracy (%)
Gauss	0.8248	14	81.18
Poly	1.9277	23	81.34
Hybrid (0.4Gauss+0.6Poly)	0.9733	18	82.50

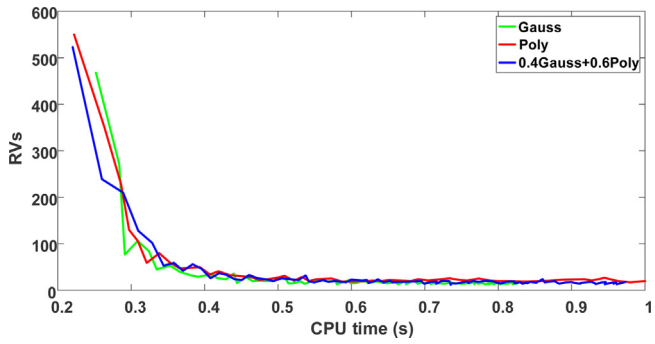

Fig. 10. The relationship of RVs and CPU time.

4.3.2. Convergence and computational complexity

As is shown in Table 5, the training time of RVM based on Gaussian kernel, polynomial kernel and hybrid kernel are 0.8248 s, 1.9277 s and 0.9733 s, respectively. When the number of the relevance vectors (RV) is concerned, the Gaussian kernel is the smallest with 14 RVs, the polynomial kernel is the largest with 23, and the hybrid kernel has 18. From Table 5, one finds that the training time increases with the number of RVs. Overall, the RVM based on hybrid kernel gets the best performance of 82.5 % in classification accuracy.

From Fig. 10, the Gaussian kernel, polynomial kernel and the hybrid kernel converge at about 0.6 s. All the computations are carried on a Lenovo computer (Inter (R) Core (TM) i5-4590 CPU 3.30 GHz) with the software MATLAB (2015b).

4.3.3. Result

This experiment is designed to show how the PSR impacts the separability of the motor imagery EEG signals. As commonly considered, 80% of the dataset for calculating the parameters of PSR and training the RVM classifier model, the rest 20% of the dataset for testing the RVM model. The one-dimensional EEG signal are reconstructed in higher dimensional phase space with the embedding dimension $m = 3$ and delay time $\tau = 6$, more useful dynamic characteristic information can be obtained, and the signal separability can be increased.

The CSP features before and after phase space reconstruction are shown in Fig. 11. Fig. 11(a) is a six-group CSP feature map before reconstruction, and Fig. 11(b) is a six-group CSP feature map after reconstruction. The distance between the CSP features (left hand vs right hand, left hand vs both feet) before reconstruction is small, and the overlap of two features cannot distinguish well. After the phase space reconstruction, the distance between the CSP features (left hand vs right hand, left hand vs both feet) become larger. Therefore, the discriminative degree of CSP features is increased after phase space reconstruction.

Table 6
Classification accuracy in different feature extraction methods on training set of BCI competition IV dataset 2a.

Subject	Feature extraction method (acc + SD %)		
	CSP	ICA + CSP	ICA + PSR + CSP
A01T	70.54	70.36	78.39
A02T	59.64	60.54	**64.29**
A03T	82.32	83.04	**84.46**
A04T	54.11	54.64	**56.61**
A05T	40.18	39.29	**46.79**
A06T	44.64	44.64	**46.25**
A07T	76.25	77.32	**88.57**
A08T	82.50	82.14	**87.14**
A09T	**59.11**	58.57	58.57
Average	63.25 ± 15.61	63.39 ± 15.89	**67.90 ± 17.04**

Based on the traditional CSP feature extraction method, the independent components of EEG can isolate the EOG components using ICA, and then reconstructed EEG with phase space reconstruction. The comparison of classification accuracy in different feature extraction methods are shown in Table 6. The classification accuracy of the traditional CSP features is 63.25%, and that of CSP features after ICA removal of EOG components is 63.39%, which improves the classification accuracy of motor imagery EEG. The classification result of CSP features extracted after ICA and PSR is 67.90%. Therefore, the classification accuracy can be improved by extracting independent components from ICA to remove EOG interference and then reconstructing EEG into high-dimensional phase space to extract CSP features.

We perform the same execution on estimating set in BCI competition IV dataset 2a. Table 7 shows the comparison result of traditional CSP features, CSP features with EOG artifacts removed using ICA, and the CSP features with EOG artifacts removed and reconstructed in phase space. The classification accuracy of these methods are 68.75%, 68.85% and 74.39% respectively. It can be seen that the CSP feature with the ICA and PSR improves the classification accuracy significantly.

4.3.4. Comparison with the state-of-the-art methods

According to the evaluation criteria kappa coefficient, proposed method is compared with some existing feature extraction methods, including the Filter Bank Common Spatial Pattern algorithm (FBCSP) [40], optimal spatial-temporal patterns CSP feature extraction algorithm (OSTP) [41], the no-homogeneous filter for

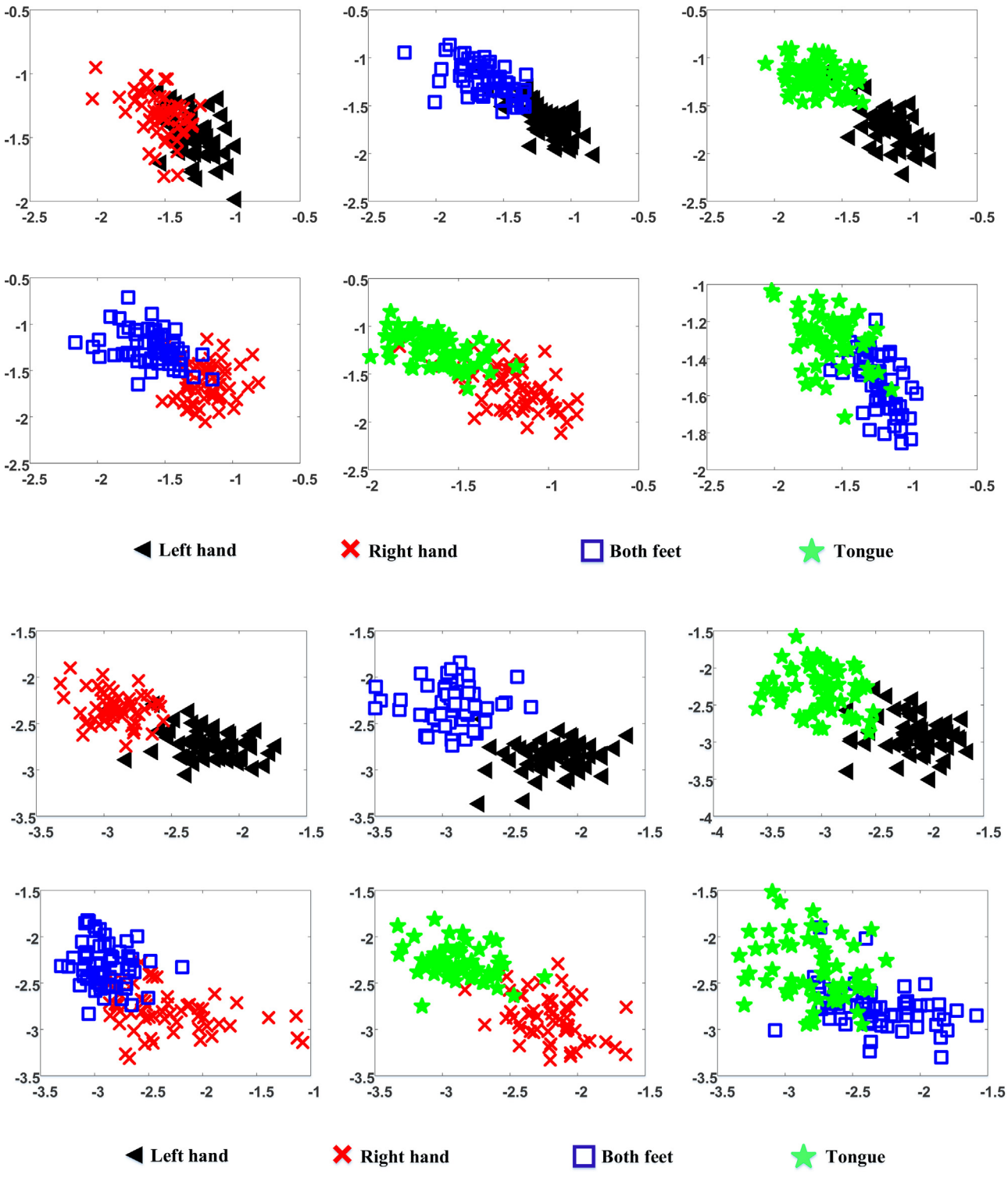

Fig. 11. CSP features before and after PSR.

distinguishing time and frequency [42], non-Homogeneous CSP (nHSCSP), separable common spatial-spectral patterns (SCSSP) [43], and discriminative filter bank common spatial pattern (DFBCSP) [44]. Table 8 shows that proposed method got the higher kappa coefficient than the FBCSP (BCI competition IV first winner) in every subject. And the average kappa coefficient of proposed method is 0.6585 which is higher than FBCSP with 0.569.

The paired t-test revealed a significant difference ($p = 0.003$) between the proposed method and FBCSP. The average kappa coefficient of OSTP is 0.595 (lower than the proposed method). The paired t-test revealed a significant difference ($p = 0.034$) between the proposed method and OSTP. The average kappa coefficient of nHSCSP is 0.6 (lower than the proposed method). The paired t-test obtained a significant level $p = 0.013$ indicating the difference between the proposed method and nHSCSP. For SCSSP, The average kappa coefficient is 0.61, and the paired t-test revealed no significant difference between proposed method and SCSSP ($p = 0.159$). When the DFBCSP is concerned, its average kappa coefficient of

Table 7
Classification accuracy in different feature extraction methods on Estimating set of BCI competition IV dataset 2a.

Subject	Feature extraction method (acc±SD %)		
	CSP	ICA + CSP	ICA + PSR + CSP
A01E	69.82	70.36	**80.00**
A02E	54.64	56.43	**65.36**
A03E	83.21	83.57	**87.14**
A04E	64.11	63.39	**67.50**
A05E	50.00	49.46	**55.54**
A06E	44.29	44.11	**50.18**
A07E	83.93	83.21	**91.79**
A08E	82.32	82.14	**84.11**
A09E	86.43	86.96	**87.86**
Average	68.75 ± 16.23	68.85 ± 16.21	**74.39 ± 15.18**

Table 8
Classification performance on the kappa value of estimating data on BCI competition IV Dataset 2a performed using the proposed PSCSP, FBCSP, OSTP, non-homogeneous-CSP (nHSCSP), SCSSP and DFBCSP.

Subject	PSCSP	FBCSP	OSTP	nHSCSP	SCSSP	DFBCSP
A01E	0.7333	0.676	0.731	0.74	**0.75**	0.68
A02E	**0.5381**	0.417	0.398	0.35	0.31	0.36
A03E	**0.8286**	0.745	0.787	0.76	0.82	0.69
A04E	0.5667	0.481	0.574	0.53	0.56	**0.62**
A05E	0.4071	0.398	0.412	0.38	**0.47**	0.6
A06E	0.3357	0.273	0.255	0.31	0.38	**0.45**
A07E	**0.8905**	0.773	0.829	0.84	0.75	0.71
A08E	**0.7881**	0.755	0.75	0.74	0.74	0.72
A09E	**0.8381**	0.606	0.62	0.74	0.67	0.66
Average	**0.6585**	0.569	0.595	0.6	0.61	0.61
p-value	–	0.003	0.034	0.013	0.159	0.326

For each subject, the highest kappa coefficient is marked in bold. Significance of the result difference is investigated using the paired *t*-test between proposed method and each of other methods.

is 0.61, and the paired *t*-test indicates no significant difference between proposed method and DFBCSP ($p = 0.326$).

It can be seen that, compared with other methods, the kappa coefficients of 9 subjects obtained by the proposed method are the highest in 5 subjects and the highest average kappa coefficient, which means the proposed method has the most meaningful classification performance for the motor imagery EEG data.

5. Conclusion

This paper proposed a novel hybrid kernel function for relevance vector machine, which combines the local and global kernel functions. The hybrid kernel function sustains the interpolation of the local kernel function and the extrapolation of the global kernel function.

The ICA algorithm was employed to isolate the independent components from multi-channel motor imagery EEG signals, which were used to remove the EOG components. Due to the nonlinear characteristics of EEG signals, the phase space reconstruction method is used to reconstruct the one-dimensional EEG signal into the high-dimensional phase space. In the phase space, the hidden nonlinear dynamic characteristics in the EEG signal are revealed. The one-dimensional EEG signals were reconstructed into high dimensional phase space. With the "one versus one" CSP feature extraction strategy, constructing six sets of spatial filters were constructed. Then, the extracted six sets of CSP features in phase space (PSCSP) were classified by the hybrid kernel function RVM. The RVM classifier has a large advantage in computing time because of its sparsity. Compared with the classification results of other methods, the proposed method has obvious advantages in the average classification accuracy as well.

However, for nonlinear time series, when calculating the phase space reconstruction parameters - embedding dimension and delay time, due to the complexity of the algorithm, it may need some computation load. Therefore, depending on the computer performance and the amount of calculated data, there may be some differences in the reconstruction of the signal. At the same time, this paper selects the hybrid kernel function RVM with different weights in the motor imagery EEG signal processing, and the corresponding weight can be adjusted according to the signals to be concerned.

In general, the EEG signal is firstly reconstructed in phase space. The reconstructed signal contains more separable features in the phase space. Then, by extracting the PSCSP feature vector, the classification results showed that the proposed method achieved better performance in four-task motor imagery EEG comparing to the state-of-the-art methods.

CRediT authorship contribution statement

Enzeng Dong: Conceptualization, Funding acquisition, Methodology, Supervision, Writing - review & editing. **Kairui Zhou:** Formal analysis, Investigation, Resources, Software, Writing - original draft. **Jigang Tong:** Data curation, Project administration. **Shengzhi Du:** Funding acquisition, Supervision, Validation, Writing - review & editing.

Acknowledgments

This work was partially supported by the Natural Science Foundation of China (No. 61603274), the Natural Science Foundation of Tianjin (No. 18JCYBJC87700).

Declaration of interest statement

The authors declare that no known competing financial interests or personal relationship that could have appeared to influence the work reported in this paper.

References

[1] N.E.M. Isa, A. Amir, M.Z. Ilyas, et al., Motor imagery classification in brain computer interface (BCI) based on EEG signal by using machine learning technique, Bull. Electr. Eng. Inform. 8 (2019) 269–275.
[2] T. Kuremoto, Y. Baba, M. Obayashi, et al., Enhancing EEG signals recognition using ROC curve, J. Robot. Netw. Artif. Life 4 (2018) 283–286.
[3] V. Bajaj, S. Taran, A. Sengur, Emotion classification using flexible analytic wavelet transform for electroencephalogram signals, Health Inf. Sci. Syst. 6 (2018) 12.
[4] E. Alickovic, J. Kevric, A. Subasi, Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction, Biomed. Signal Process. Control 39 (2018) 94–102.
[5] Y. Zhang, S. Zhang, X. Ji, EEG-based classification of emotions using empirical mode decomposition and autoregressive model, Multimed. Tools Appl. 77 (2018) 26697–26710.
[6] S. Dutta, M. Singh, A. Kumar, Automated classification of non-motor mental task in electroencephalogram based brain-computer interface using multivariate autoregressive model in the intrinsic mode function domain, Biomed. Signal Process. Control 43 (2018) 174–182.
[7] S.X. Wang, H.G. Gong, N. Li, Detection analysis of epileptic EEG using a novel random forest model combined with grid search optimization, Front. Hum. Neurosci. 13 (2019) 52.
[8] Z. Tao, C. Wanzhong, L. Mingyang, Classification of inter-ictal and ictal EEGs using multi-basis MODWPT, dimensionality reduction algorithms and LS-SVM: a comparative study, Biomed. Signal Process. Control 47 (2019) 240–251.
[9] C. Zhang, J. Xu, S. Pan, et al., Identification and classification of electroencephalogram signals based on independent component analysis, NeuroQuantology 16 (2018) 832–838.
[10] A. Fiorenzo, D. Arnaud, M. Scott, A visual working memory dataset collection with bootstrap independent component analysis for comparison of electroencephalographic preprocessing pipelines, Data Brief 22 (2019) 787–793.

[11] J. Wang, Z. Feng, N. Lu, et al., An information fusion scheme based common spatial pattern method for classification of motor imagery tasks, Biomed. Signal Process. Control 46 (2018) 10–17.
[12] K. Lafleur, K. Cassady, A. Doud, et al., Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface, J. Neural Eng. 10 (2013), 046003.
[13] Y. Li, J. Long, T. Yu, et al., An EEG-based BCI system for 2-D cursor control by combining Mu/beta rhythm and P300 potential, IEEE Trans. Biomed. Eng. 57 (2010) 2495–2505.
[14] A.J. Doud, J.P. Lucas, M.T. Pisansky, et al., Continuous three-dimensional control of a virtual helicopter using a motor imagery based brain-computer interface, PLoS One 6 (2011), e26322.
[15] Y. Zhang, G. Zhou, J. Jin, et al., Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface, J. Neurosci. Methods 255 (2015) 85–91.
[16] Y. Zhang, C. Nam, G. Zhou, et al., Temporally constrained sparse group spatial patterns for motor imagery BCI, IEEE Trans. Cybern. 49 (2019) 3322–3332.
[17] S. Dodia, D.R. Edla, A. Bablani, et al., An efficient EEG based deceit identification test using wavelet packet transform and linear discriminant analysis, J. Neurosci. Methods 314 (2019) 31–40.
[18] E. Dong, C. Li, L. Li, et al., Classification of multi-class motor imagery with a novel hierarchical SVM algorithm for brain–computer interfaces, Med. Biol. Eng. Comput. 55 (2017) 1809–1818.
[19] E. Dong, G. Zhu, C. Chen, et al., Introducing chaos behavior to kernel relevance vector machine (RVM) for four-class EEG classification, PLoS One 13 (2018), e0198786.
[20] Z. Jin, G. Zhou, D. Gao, et al., EEG classification using sparse Bayesian extreme learning machine for brain–computer interface, Neural Comput. Appl. (2018) 1–9.
[21] Y. Zhang, G. Zhou, J. Jin, et al., Sparse bayesian classification of EEG for brain-computer interface, IEEE Trans. Neural Netw. Learn. Syst. 27 (2016) 2256–2267.
[22] S. Dutta, M. Singh, A. Kumar, Classification of non-motor cognitive task in EEG based brain-computer interface using phase space features in multivariate empirical mode decomposition domain, Biomed. Signal Process. Control 39 (2018) 378–389.
[23] M. Chen, Y. Fang, X. Zheng, Phase space reconstruction for improving the classification of single trial EEG, Biomed. Signal Process. Control 11 (2014) 10–16.
[24] Y. Ma, W. Shi, C.K. Peng, et al., Nonlinear dynamical analysis of sleep electroencephalography using fractal and entropy approaches, Sleep Med. Rev. 37 (2018) 85–93.
[25] K. Sayed, M. Kamel, M. Alhaddad, et al., Characterization of phase space trajectories for brain-computer interface, Biomed. Signal Process. Control 38 (2017) 55–66.
[26] H. Niknazar, A.M. Nasrabadi, M.B. Shamsollahi, A new similarity index for nonlinear signal analysis based on local extrema patterns, Phys. Lett. A 382 (2018) 288–299.
[27] B. Sharif, A.H. Jafari, Prediction of epileptic seizures from EEG using analysis of ictal rules on Poincaré plane, Comput. Methods Programs Biomed. 145 (2017) 11–22.
[28] K. Natarajan, A.U. Rajendra, F. Alias, et al., Nonlinear analysis of EEG signals at different mental states, Biomed. Eng. Online 3 (2004), 7-7.
[29] A. Banerjee, S. Sanyal, A. Patranabis, et al., Study on brain dynamics by non linear analysis of music induced EEG signals, Phys. A Stat. Mech. Appl. 444 (2016) 110–120.
[30] Y. Fang, M. Chen, X. Zheng, Extracting features from phase space of EEG signals in brain–computer interfaces, Neurocomputing 151 (2015) 1477–1485.
[31] C. Brunner, M. Naeem, R. Leeb, et al., Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis, Pattern Recognit. Lett. 28 (2007) 957–964.
[32] Ricardo Vigário, Jaakko Särelä, Veikko Jousmäki, et al., Independent component approach to the analysis of EEG and MEG recordings, IEEE Trans. Biomed. Eng. 47 (2000) 589–593.
[33] N.H. Packard, J.P. Crutchfield, J.D. Farmer, et al., Geometry from a time series, Phys. Rev. Lett. 45 (1980) 712–716.
[34] F. Takens, Detecting Strange Attractors in Turbulence, Dynamical Systems and Turbulence, Warwick 1980, Springer, Berlin, Heidelberg, 1981, pp. 366–381.
[35] H.S. Kim, R. Eykholt, J.D. Salas, Nonlinear dynamics, delay times, and embedding windows, Physica D 127 (1999) 48–60.
[36] M.E. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Mach. Learn. Res. 1 (2001) 211–244.
[37] Robert A. Lordo, Learning from data: concepts, theory, and methods, Technometrics 43 (2001) 105–106.
[38] G.F. Smits, E.M. Jordaan, Improved SVM regression using mixtures of kernels, international joint conference on neural networks, IEEE Xplore (2002) 2785–2790.
[39] A. Cohen, Coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1960) 37–46.
[40] A.K. Keng, C.Z. Yang, W. Chuanchu, et al., Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b, Front. Neurosci. 6 (2012), http://dx.doi.org/10.3389/fnins.2012.00039.
[41] K.K. Ang, Z.Y. Chin, H. Zhang, et al., Mutual information-based selection of optimal spatial–temporal patterns for single-trial EEG-based BCIs, Pattern Recognit. 45 (2012) 2137–2144.
[42] T.E. Kam, H.I. Suk, S.W. Lee, Non-homogeneous spatial filter optimization for ElectroEncephaloGram (EEG)-based motor imagery classification, Neurocomputing 108 (2013) 58–68.
[43] A.S. Aghaei, M.S. Mahanta, K.N. Plataniotis, Separable common spatio-spectral patterns for motor imagery BCI systems, IEEE Trans. Biomed. Eng. 63 (2015) 15–29.
[44] B.E. Olivas-Padilla, M.I. Chacon-Murguia, Classification of multiple motor imagery using deep convolutional neural networks and spatial filters, Appl. Soft Comput. 75 (2019) 461–472.