Introducing chaos behavior to kernel relevance vector machine (RVM) for four-class EEG classification

Enzeng Dong[1], Guangxu Zhu[1], Chao Chen[1], Jigang Tong[1]*, Yingjie Jiao[2], Shengzhi Du[3]

1 Tianjin Key Laboratory For Control Theory & Applications in Complicated Systems, Tianjin University of Technology, Tianjin, The People's Republic of China, **2** Xi'an Modern Control Technology Research Institute, Xi'an, Shaanxi Province, The People's Republic of China, **3** Department of Electrical Engineering, Tshwane University of Technology, Pretoria, South Africa

* tjgtjut@163.com

Abstract

This paper addresses a chaos kernel function for the relevance vector machine (RVM) in EEG signal classification, which is an important component of Brain-Computer Interface (BCI). The novel kernel function has evolved from a chaotic system, which is inspired by the fact that human brain signals depict some chaotic characteristics and behaviors. By introducing the chaotic dynamics to the kernel function, the RVM will be enabled for higher classification capacity. The proposed method is validated within the framework of one versus one common spatial pattern (OVO-CSP) classifier to classify motor imagination (MI) of four movements in a public accessible dataset. To illustrate the performance of the proposed kernel function, Gaussian and Polynomial kernel functions are considered for comparison. Experimental results show that the proposed kernel function achieved higher accuracy than Gaussian and Polynomial kernel functions, which shows that the chaotic behavior consideration is helpful in the EEG signal classification.

Introduction

Brain-Computer Interface (BCI) is an interdisciplinary cutting-edge technology that establishes communication and control channels between human brain and an external computer or other intelligent electronic equipment [1–5]. Motor imagery (MI) based BCIs focus on converting the recorded electroencephalograph (EEG) during imagining limb or body movements, the so-called 'idea', into specific codes or commands to detect EEG signal behaviour or control the intelligent equipment [6–9].

To accurately classify or decode EEG signals in BCI, pattern recognition is a vitally important step. A few EEG classification algorithms were proposed, for example, the linear discriminant analysis (LDA), the artificial neural networks (ANN), and the support vector machine (SVM), etc. The LDA [10, 11] is a two-class classification which divides the input space into two subspaces by mapping the multidimensional input vector to a hyperplane,

each subspace representing one class. It was proposed that appropriate regularization of LDA by shrinkage improves the LDA performance in single-trial ERP classification [10]. The ANN [12–15] is an artificial multi-layer "neuron" inspired by the biological neuronal structure in the human brain. In ANN, a hyperplane used for classification is obtained by computing the weighted sum between neurons. Three types of ANN structures for two-class 2-D cursor movement classification were developed in [12]. A filter based on ANN [13] was proposed to reduce EEG interference signals. The SVM finds the classification hyperplane, in which two-class samples exhibit the largest distance. When the two classes are not linearly separable, the linear SVM fails to solve the classification problems. In these cases, kernel functions can be designed to map nonlinear separable samples into a high-dimensional space first, so that the samples in the high- dimensional space become linearly separable. In recent years, successful BCI experiments were reported based on SVM or its variations [12, 16–23]. For instance, a SVM method was developed to distinguish seizure EEG epochs from normal ones [21]. A hierarchical SVM algorithm was proposed for four-class EEG signal classification [23].

With further understanding of SVM, researchers gradually found some inherent shortcomings: 1. the kernel functin for nonlinear classification must meet Mercer's condition; (i.e. for a real-valued function K(x, y), $\int \int g(x)K(x,y)g(y)dxdy \geq 0$ for all square integrable functions g (x)) 2. the experimental results are usually sensitive to the penalty factor, which can easily lead to overfitting; 3. the output is not always reliable, and so on. Based on the Bayesian framework, the relevance vector machine (RVM) was proposed, which is similar to SVM, but overcomes these shortcomings. RVM is actually a Bayesian sparse kernel method for regression problems and classification problems [24, 25], because the final result is only related to a few relevant input vectors [26–32]. Comparing with SVM, the kernel function of RVM does not need satisfying the Mercer's condition, so more options of kernel function can be considered. Comparing with ANN, RVM is based on small sample data to obtain the optimal solution, while ANN generally needs a large amount of training data. In motor imagery EEG classification, there are usually only limited amount of training samples available, therefore, RVM is more suitable when the required amount of samples is considered. Overfitting is another problem of ANN, which limits the ANN in EEG classification. Therefore, we chose RVM as the EEG signal classification algorithm in this research.

Chaos is a common phenomenon that exists in nonlinear systems. Chaos does not mean disorder, but has a delicate inner structure. Studies have shown that EEG signals possess some chaotic properties [33–39]. Furthermore, chaotic systems demonstrate rich dynamic behavior, if utilized properly in kernel based methods, which is helpful for the generalization of classifiers. Therefore, the classification capacity can be improved. Inspired by these facts and considering the flexibility of kernel function selection in RVM, a chaos kernel function for RVM is proposed in this research, and validated by 4-class MI classification.

The rest of the paper is organised as follows. In section 2, the EEG data used in the research is described and the algorithms involved in this paper are reviewed in detail. Section 3 demonstrates various experimental results. Section 4 concludes the paper with a discussion on the advantages and disadvantages of the proposed method.

Methods

CSP framework for feature extraction in four-class MI classification

Common spatial pattern (CSP) was used to extract features from the processed EEG signal [40, 41]. The CSP distinguishes two categories of samples by a spatial projection in the manner that
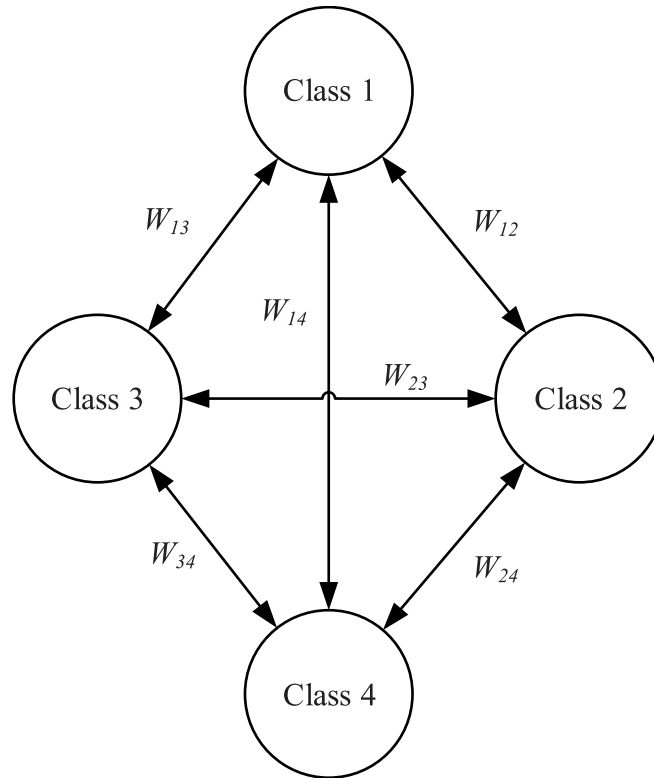
Fig 1. The illustration of OVO strategy [23].

the energy difference between the classes is maximised. For four-class classification, the OVO strategy was developed to enable CSP for the feature selection [23, 42], as illustrated in Fig 1.

For the four classes labelled as class 1, 2, 3 and 4 respectively, the OVO-CSP selects two of the four classes as the input for original CSP, which generates 6 possible selections.

Without loss of generality, for the sake of easier understanding, classification of the class 1 and class 2 is considered as an example. For the two selected classes, X_i (where $i \in 1, 2$) denotes an EEG sample of class i, X_i is a matrix of $N \times T$, where N is the number of channels, T is the product of sampling frequency and acquisition (seconds), that is, the number of sampling points in a channel for one MI epoch.

Dong et.al [43] demonstrated the method by decomposing the mixed spatial covariance matrix and then mapping the EEG signal to a feature space. The normalised covariance matrix of class i in epoch (trial) n is

$$R_{i,n} = \frac{X_i X_i^T}{trace(X_i X_i^T)} \tag{1}$$

where X_i^T is the transpose matrix of X_i, and $trace(X_i X_i^T)$ is the trace of $X_i X_i^T$, n = 1, 2,···, N_e and N_e is the total number of epochs for class i. The spatial covariance can be computed by averaging all the trials (epochs) of the class i.

The original EEG signal X is projected to the new spatial space as

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

where W is a spatial filter calculated by CSP.

The features used for classification are obtained from (2). For each class of imagined movements, only a small amount, denoted as m, of the most distinguishing signal variances is selected for classification. Z_k ($k = 1, 2, \cdots, 2m$) is constructed by the first m and last m rows of Z, which maximize the difference of variance of two-class EEG signals.

OVO-CSP transforms the four-class classification problem into six cases of two-class classification. We pick up the first and the last vectors (corresponding to the largest and the smallest eigenvalues respectively) from the sorted feature matrix Z as the most significant two feature vectors.

Instead of using Z directly, the normalised log-variances of these components are considered to be features for classification.

The feature corresponding to Z_k is calculated as

$$f_k = \log\left(\frac{\mathrm{var}(Z_k)}{\sum_{i=1}^{2m} \mathrm{var}(Z_i)} \right) \tag{3}$$

This new feature makes the distance between two classes more significant.

RVM classification

Assume that $\{\mathbf{u}_i\}_{i=1}^N$ is the eigenvector in training data and $\{t_i\}_{i=1}^N$ ($t_i \in \{0, 1\}$) is the corresponding target value. Then the RVM classification model can be expressed as

$$y(\mathbf{u}; \mathbf{w}) = \sum_{i=1}^N w_i K(\mathbf{u}, \mathbf{u}_i) + w_0 \tag{4}$$

Where $K(\mathbf{u}, \mathbf{u}_i)$ is a kernel function, w_i is the weight of the i-th kernel function, $\mathbf{w} = [w_0, w_1, \cdots, w_N]^T$, w_0 is the bias.

For the two-class classification, we adopt the Logistic Sigmoid function to map $y(\mathbf{u}; \mathbf{w})$ to $(0, 1)$. Since the target value $\{t_i\}_{i=1}^N$ can only be 0 or 1, and each prediction is independent, the samples are assumed to be independent and identically distributed.

To avoid introducing the shortcomings similar to the SVM, such as severe over-fitting due to excessive support vectors used, the weight vector $\mathbf{w}$ is constrained with the precondition, that is, all weight vectors satisfy a zero-mean Gaussian prior distribution.

$$p(\mathbf{w}|\alpha) = \prod_{i=0}^N N(w_i|0, \alpha_i^{-1}) = \prod_{i=0}^N \frac{\alpha_i}{\sqrt{2\pi}} \exp\left(-\frac{\alpha_i w_i^2}{2}\right) \tag{5}$$

Where $\alpha = [\alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_N]^T$ is a hyper-parameter vector which determines the prior distribution of the weight vector w, and controls the degree to which the weight deviates from its zero-mean.

Given the prior probability distribution and the likelihood distribution, the Bayes' Rule is adopted to calculate the posterior probability of models $\mathbf{w}$ and α [44]

$$p(\mathbf{w}, \alpha|\mathbf{t}) = p(\mathbf{w}|\mathbf{t}, \alpha)p(\alpha|\mathbf{t}) \tag{6}$$

In Eq (6), the posterior probability $p(\mathbf{w}|\mathbf{t}, \alpha)$ and $p(\alpha|\mathbf{t})$ cannot be directly solved, the approximation procedure, as used by MacKay [45], can be adopted based on Laplace's method.

And the maximum w can be calculated as follow.

$$\mathbf{w}^{new}_{MP} = \mathbf{w}_{MP} + \Delta\mathbf{w} \tag{7}$$

$$\alpha_i^{new} = \frac{1 - \alpha_i \Sigma_{i,i}}{\mathbf{w}^2_{MPi}} \tag{8}$$

Where $\Delta\mathbf{w} = -\mathbf{H}^{-1}g$, $\mathbf{H} = -(\Phi^T B\Phi + A)$, $\Sigma = (-\mathbf{H}|_{\mathbf{w}_{WP}})^{-1}$, $\mathbf{A} = diag(\alpha_0, \alpha_1, \cdots, \alpha_N)$, $\mathbf{B} = diag$ $[y_i(1 - y_i)]$, $y_i = \sigma\{y(\mathbf{x}_i; \mathbf{w})\}$, $g = \Phi^T(\mathbf{t} - \mathbf{y}) - A\mathbf{w}$, $\mathbf{y} = [y_1, y_2, \cdots, y_N]^T$,

$$\Phi = \begin{bmatrix} 1 & K(\mathbf{u}_1, \mathbf{u}_1) & K(\mathbf{u}_1, \mathbf{u}_2) & \cdots & K(\mathbf{u}_1, \mathbf{u}_N) \\ 1 & K(\mathbf{u}_2, \mathbf{u}_1) & K(\mathbf{u}_2, \mathbf{u}_2) & \cdots & K(\mathbf{u}_2, \mathbf{u}_N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & K(\mathbf{u}_N, \mathbf{u}_1) & K(\mathbf{u}_N, \mathbf{u}_2) & \cdots & K(\mathbf{u}_N, \mathbf{u}_N) \end{bmatrix}$$

$\Sigma_{i,i}$ is the i-th diagonal element in Σ.

The RVM algorithm model training procedure is to proceed to repeat (8), concurrent with updating (7), until some appropriate convergence conditions have been met.

In fact, with the repeated updating, the majority of α_i approaches infinity, and the corresponding w_i approaches 0. The $\mathbf{u}_i$ corresponding to the non-zero weight are relevant vectors. Assume that $\{\mathbf{u}_*\}$ is the test sample vector, we make classification predictions by the weights obtained from the learning training data, as follows.

$$\mathbf{t}_* = y(\mathbf{u}_*; \mathbf{w}_{MP}) = \Phi(\mathbf{u}_*)\mathbf{w}_{MP} \tag{9}$$

Chaos kernel function for RVM

Fig 2 roughly presents the steps of classification of BCI signals by employing the RVM. The complete procedure mainly includes four parts: training data processing, the RVM training, test data processing, and the RVM test. The re-estimation in the RVM training procedure is the key step of the algorithm to achieve sparseness.

Because the kernel functions map the feature vectors to a high-dimensional space to achieve linear classification, the properties of the kernel functions play an important role in the performance of the RVM classification algorithm. In this paper, a chaos kernel (CK) is proposed, which evolves from the probability distribution of a chaotic sequence.

Consider the fact that the human brain signal is so complex that there is currently no theory or rule to fully explain its behaviours, but it is believed that there must be some rules behind the seemed "disorderly" signals. As shown in Fig 3, when our brain is in a state of motor imagination, the chaos in motor imagery might associate with some mental behaviours (known or unknown). The equation transformed from this chaos system can be considered to decode the brain activities. Furthermore, inspired by the idea of a kernel function, the low-dimensional collected brain signal is mapped to a high-dimensional space to find more intuitive features related to MI.

While chaos is a seemingly random irregular motion occurring in a deterministic system, it does hide a certain law. Therefore, in this paper, we are inspired to construct a kernel function for RVM from the chaos theory perspective. The Logistic Map in (10), a classic chaotic system
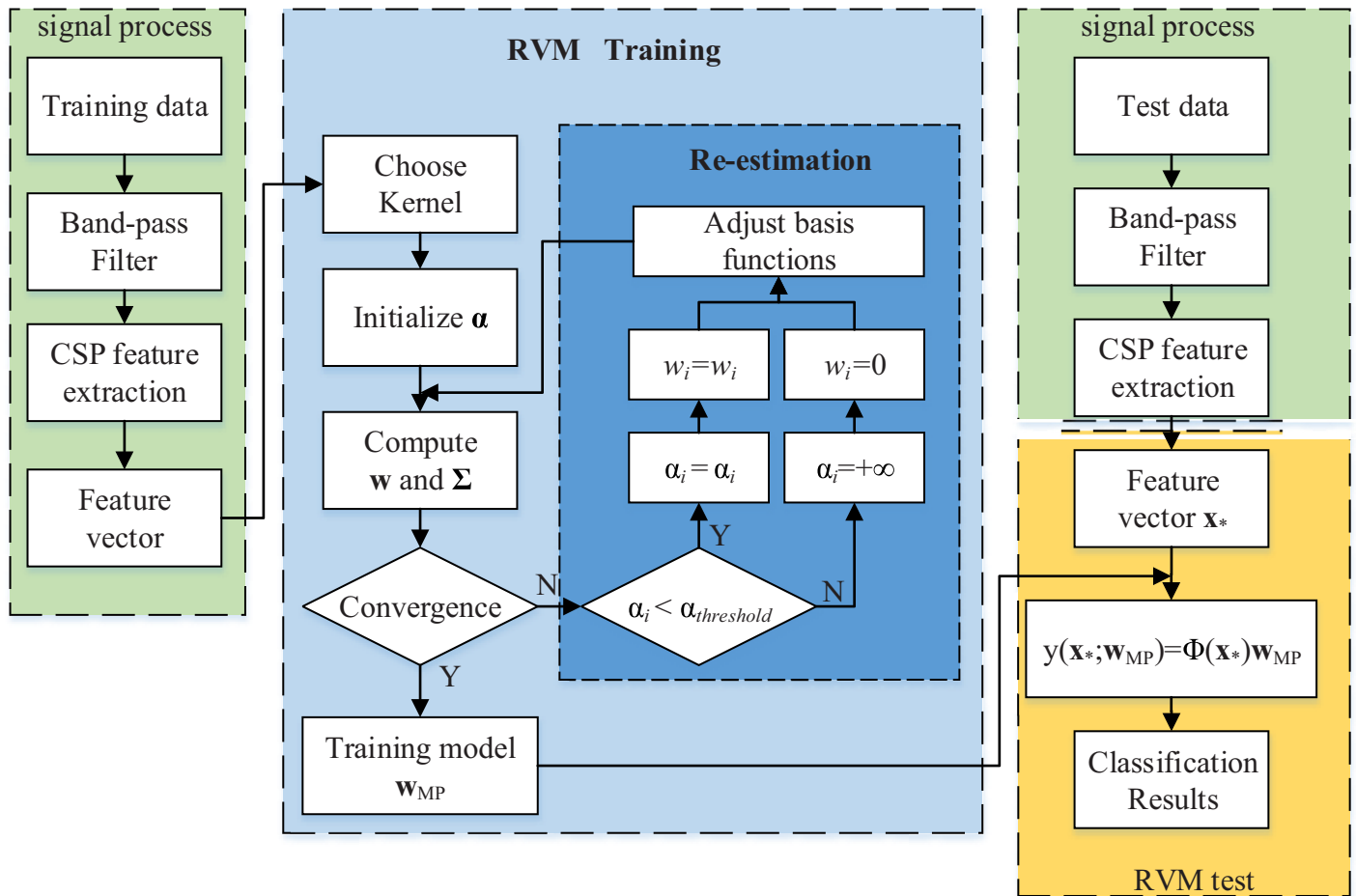
Fig 2. The flow of EEG signal processing and RVM algorithm.

model, is used in this paper.

$$Y_{n+1} = AY_n(1 - Y_n) \tag{10}$$

Fig 4 shows the bifurcation diagram of the typical Logistic map and the corresponding Lyapunov spectrum. When $A = 4$, the Lyapunov exponent of the Logistic mapping is more than 0, and the Logistic mapping is in a chaotic state. In this way, we think the following series of changes are based on the chaos-related equations.

When $A = 4$, the probability distribution of Y is

$$P(Y) = \frac{1}{\pi\sqrt{Y(1 - Y)}} \tag{11}$$

As shown in Fig 5, with enough iterations of the logistic map for $A = 4$, the orbit approaches arbitrarily close to every point in the interval $0 < Y < 1$. The probability distribution function $P(Y)$ has peaks at $Y = 0$ and $Y = 1$. But it is not very suitable for classification.

In Fig 6, we can see that the Lyapunov exponent of the transformed system at $A = 4$ is greater than 0, so the system is still a chaos system. The probability distribution of the transformed chaos system is shown in Fig 7.
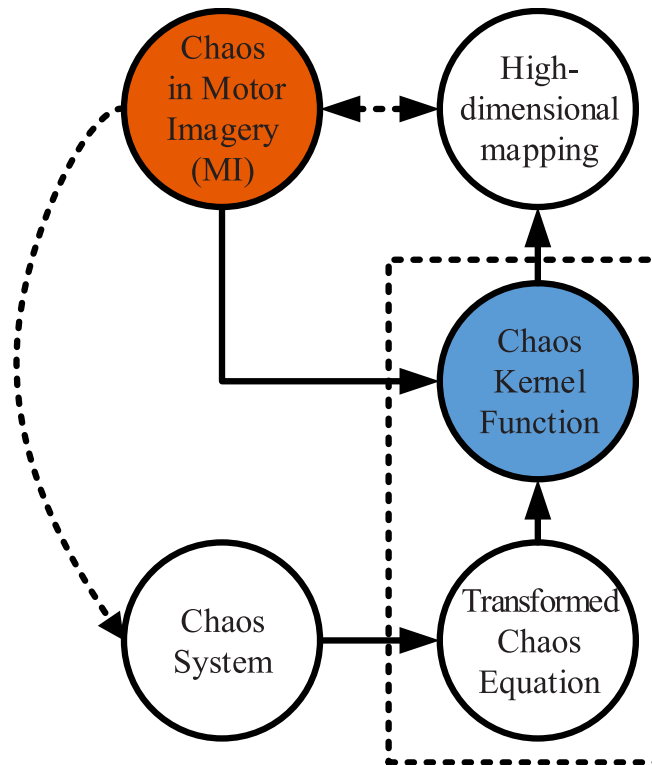
Fig 3. The relationship between chaos kernel and MI.

Applying the logit transform $f(Y) = \ln\left(\frac{Y}{1-Y}\right)$ to the iterates of the Logistic Map with $A = 4$ gives a probability distribution function

$$P(Y) = \frac{1}{\pi(e^{Y/2} + e^{-Y/2})} \tag{12}$$
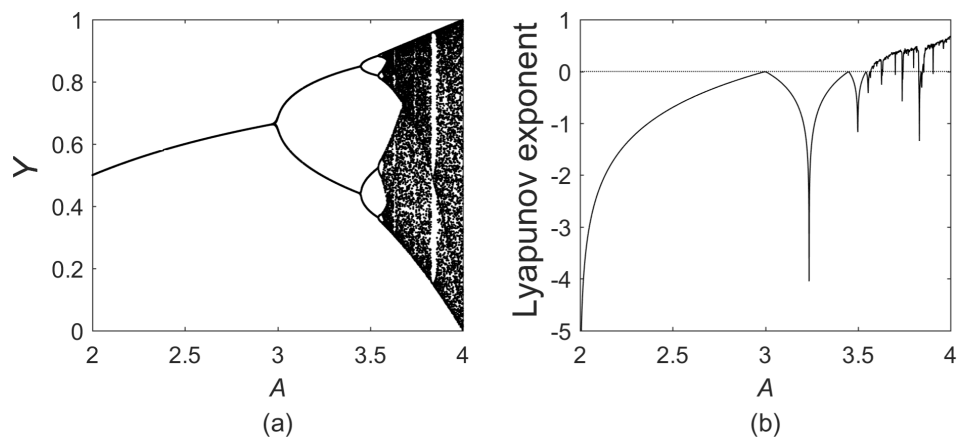

Fig 4. (a) Bifurcation diagram for the logistic map; (b) Lyapunov spectrum for the logistic map.
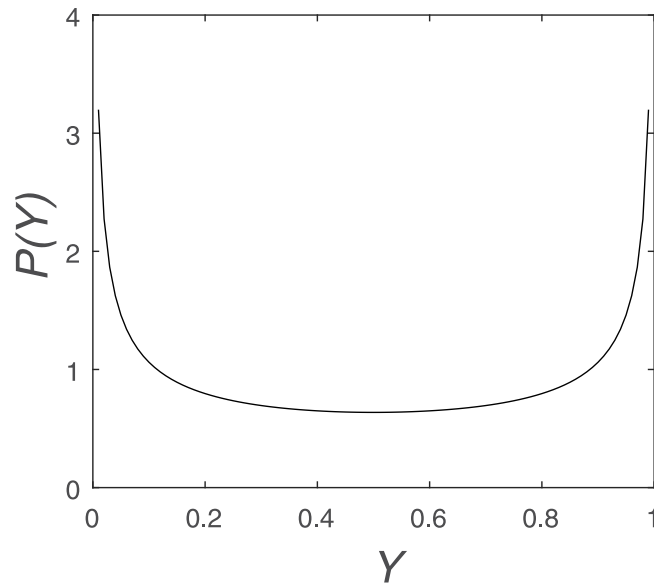
Fig 5. Probability distribution function of the logistic map with *A* = 4.

Evolve (12) into a kernel function

$$K(Y_1, Y_2) = \frac{1}{\pi(\exp(\beta \parallel Y_1 - Y_2 \parallel) + \exp(-\beta \parallel Y_1 - Y_2 \parallel))} \tag{13}$$

where β is the parameter, $\parallel \cdots \parallel$ is the 2-norm operation.

The kernel function used in SVM has to satisfy Mercer's condition that the kernel matrix must be a positive semidefinite matrix. While the RVM algorithm avoids this condition. Thus, the proposed chaos kernel function does not have to satisfy Mercer's condition. Nonetheless, the kernel matrix of the chaos kernel is a positive semidefinite matrix indeed. So it can also be used in SVM.
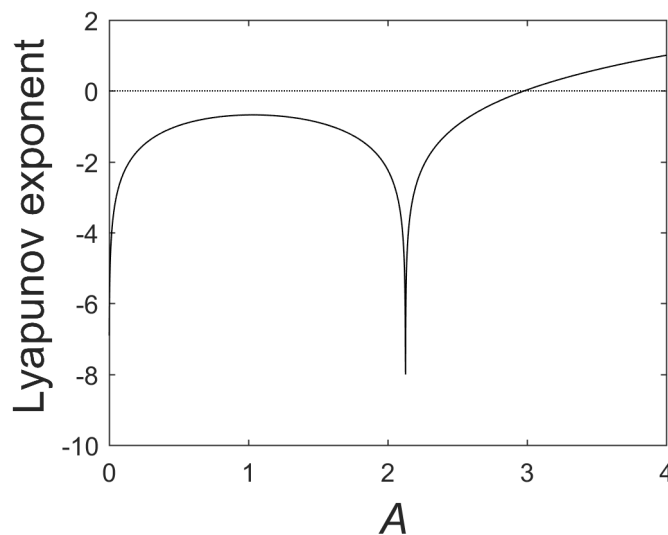

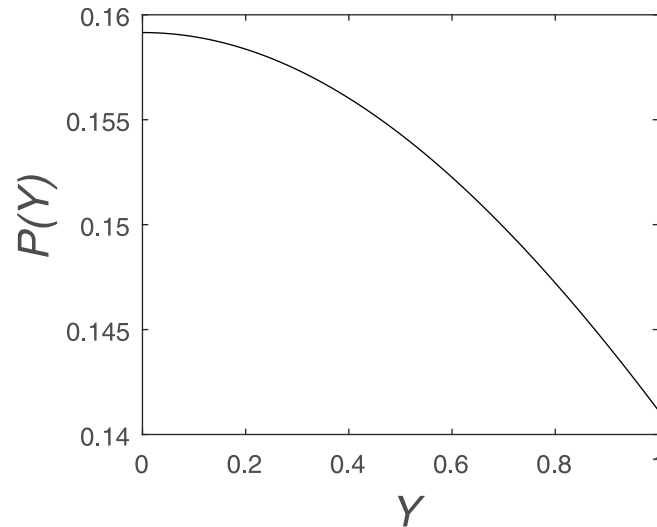**Fig 6. Lyapunov spectrum for the transformed system.**

Fig 7. Probability distribution of $P(Y) = 1/\pi(e^{Y/2} + e^{-Y/2})$.

Four-class MI classification based on the framework of OVO-CSP

The event-related frequency bands are firstly extracted from the original EEG signals containing four-class motor imagery movements. The band-pass filter (3-24Hz) is employed, and then the filtered EEG signals are randomly divided into five groups. Four groups are used for training the classifier and the rest is for the test. Six CSP projection matrices are constructed to address the four-class classification as detailed in section 2.1, denoted as W12, W13, W14, W23, W24, and W34 respectively. Then the matrices are used to extract the features of the corresponding category from the EEG data. Finally, the six sets of features are sent to the RVM as the input vector to train six models. Using these projection matrices to extract features from the test dataset, one obtains features as the input vector of the RVM test section.

The six models obtained by the RVM training are combined with the input features of the test set to predict the classification. The whole classification procedure is shown in Fig 8. The 5-fold-cross validation is used to ensure that each group has been tested once as the test set.

Experiments and results

EEG dataset illustration

The dataset for the simulation experiment in this paper was derived from the BCI competition IV-II-a [46], which provided by Graz University of Technology, Austria, in 2008. The dataset contains four-class motor imagery tasks: the imagination of movement of the left hand (class 1), the right hand (class 2), both feet (class 3), and the tongue (class 4). The data recording equipment collects EEG signals and EOG signals by utilising 22 Ag/AgCl electrode channels and three monopolar EOG channels respectively, with the sampling frequency of 250 Hz. While the EOG signals included in the dataset were not used for classification in this paper, those signals provided were bandpass filtered between 0.5Hz and 100Hz. In fact, we found that only the frequency bands [3, 24] Hz change visibly during motor imagery [23]. Thus, we re-bandpass filtered the provided EEG signals with the band [3, 24] Hz.

The BCI competition 2008—Graz data set A contains two sessions on nine subjects which were recorded on two different days, taking into account the nature of unstable state of the subjects. We named the two sessions respectively T and E. Both of them have 6 runs separated
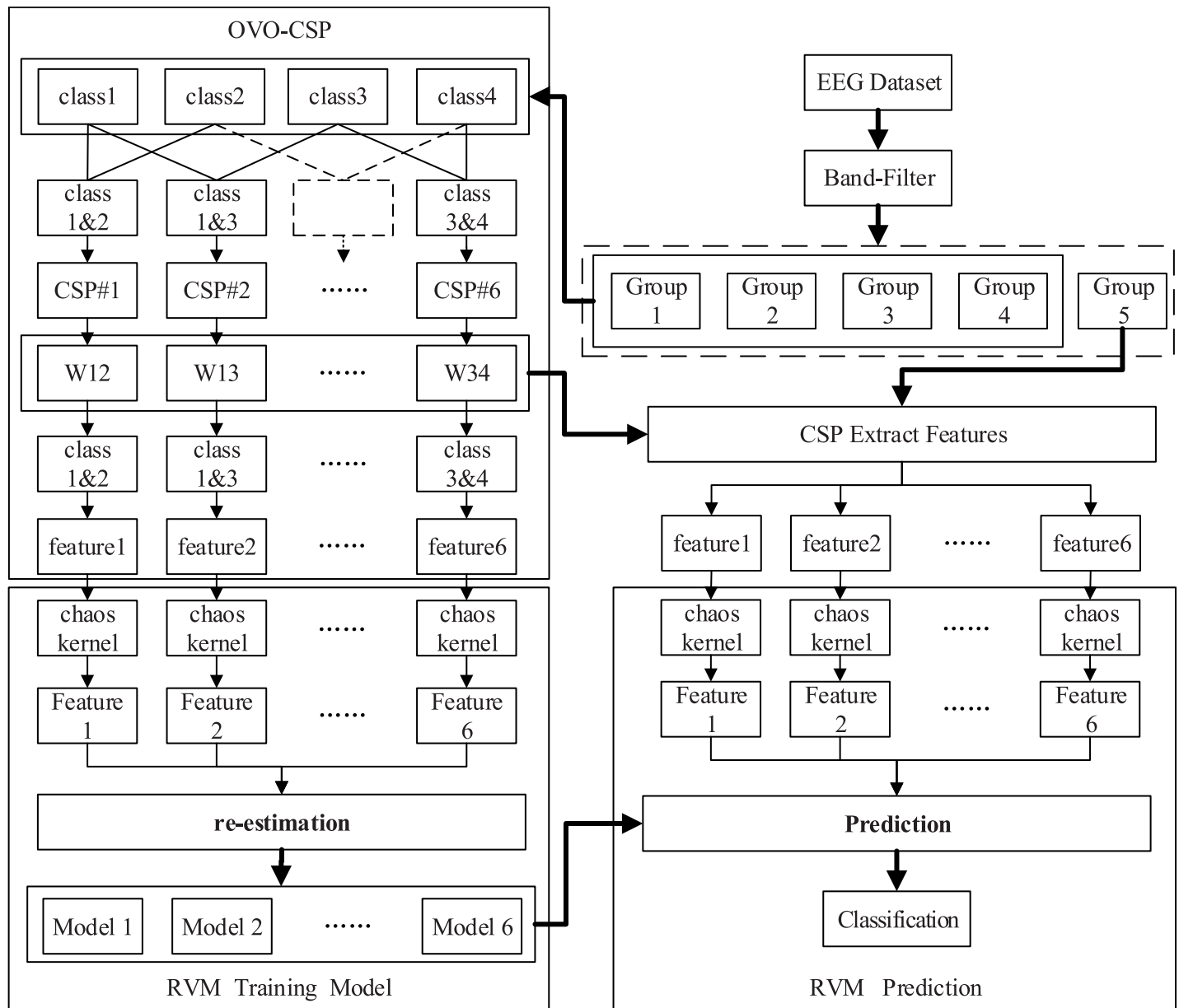
Fig 8. The framework of four-class MI classification based on OVO-CSP and chaos kernel RVM.

https://doi.org/10.1371/journal.pone.0198786.g008

by short breaks. Each run includes 48 trials (12 trials per class). That is to say, both of the sessions have 288 trials to be processed. Thus, we extracted 72 valid trials corresponding with each class of the motor imagery task. The selected four-class EEG data is re-bandpass filtered to extract features using the constructed OVO-CSP. Then five-fold cross-validation is employed to eliminate the over-fitting as much as possible. Original data (72 trials) for each category of the motor imagery tasks is randomly divided into five parts, where the four-part sample (56 trials) is used to train the RVM model and the rest (14 trials) is used for the validation. The cross-validation procedure will be repeated five times, then each part of the sample can undergo validation once.
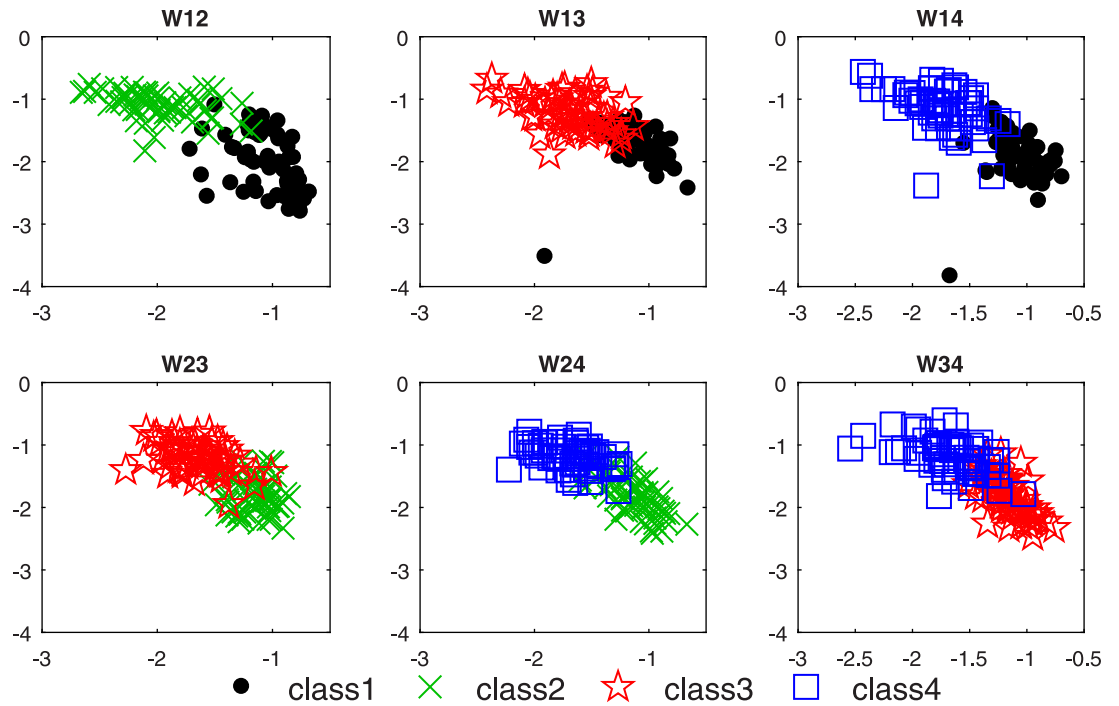
Fig 9. The distribution of most significant feature vector pairs obtained by OVO-CSP.

Results of OVO-CSP feature extraction

The four-class MI classification is transformed into six cases of two-class classification by OVO-CSP. The results of the feature extraction are depicted in Fig 9, showing the distribution of the most significant feature vector pairs obtained by OVO-CSP. Fig 9 suggests that the OVO-CSP obtains separable feature distributions used for RVM classification.

Results and comparison with existing methods

To illustrate the performance of the proposed kernel function, the Gaussian kernel and polynomial kernel, shown in (14) and (15) respectively, are considered for comparison.

$$K(Y_1, Y_2) = \exp(-\frac{\| Y_1 - Y_2 \|}{2\sigma^2}) \tag{14}$$

where σ is the width parameter.

$$K(Y_1, Y_2) = [Y_1(aY_2)^T + 1]^d \tag{15}$$

where a is a user-specified scalar parameter, and the polynomial degree d chosen in this paper is 2.

Comparison of two sessions' classification accuracy about Polynomial kernel function (PK), Gaussian kernel function (GK), and the proposed kernel function (chaos kernel function, CK) are shown in Tables 1 and 2. Each session was randomly divided into five parts (each part contains 56 epochs), four were selected for the training weight model, and the remaining one for verification. Five cross validation ensures that every part will be validated. Thirty experiments were conducted in order to gain reliable results. The average accuracy and standard deviation are calculated.

Table 1. Comparison of 3-kind kernels based on RVM on session T.

Subject	Polynomial kernel (mean ±std%) ($a = 1.26, d = 2$)	Gaussian kernel (mean ±std%) ($\sigma = 0.95$)	Chaos kernel (mean ±std%) ($\beta = 0.5$)
S1	69.4±1.7	68.9±1.5	**69.7±1.2**
S2	47.3±1.8	46.5±2.0	**48.6±1.5**
S3	**76.9±1.1**	75.9±1.3	76.4±0.8
S4	52.1±1.3	51.0±2.0	**52.9±1.2**
S5	38.7±2.2	38.7±2.1	**38.9±1.8**
S6	41.5±2.2	**41.7±1.5**	40.7±1.1
S7	72.9±1.5	73.2±2.2	**73.2±1.0**
S8	80.2±1.1	79.8±1.6	**80.7±0.7**
S9	73.3±1.2	72.9±0.8	**73.6±0.9**
Average	61.4±15.4	60.9±15.3	**61.6±15.4**

Table 1 shows that the average accuracy of classification of the three kernel functions (PK, GK and CK) is 61.4 ± 15.4%, 60.9 ± 15.3% and 61.6 ± 15.4%, respectively. The overall performance of CK is better than PK and GK. In Table 1, each subject's classification result is made up of two parts, the average accuracy and standard deviation respectively. They are two indicators in statistics. The smaller the standard deviation is, the more the statistical results are concentrated on both sides of the mean (i.e., the average accuracy). It can be seen from the Tables 1 and 2 that the standard deviation of classification results for the chaos kernel RVM are generally smaller than the others, indicating that the results are more centralized and more credible. Which suggested the proposed method is more effective for EEG signal classification.

For the individual subject case, the best of the three kernel functions are bolded. In most cases, subjects 1, 2, 4, 7, 8 and 9, the proposed chaos kernel function achieved a higher accuracy. For the remaining subjects, the proposed chaos kernel function yields a slightly lower accuracy.

Similar results are presented in Table 2 for the second session, in which the chaos kernel performance performed better on subjects 1, 3, 5, 7, 8 and 9 than with the other kernels. The polynomial kernel function performed better for subjects 2, 4 and 6. The chaos kernel function achieves better accuracy with 65.4 ± 15.3%, a little advantage over the other by 64.9 ± 15.1%, 64.7 ± 15.3%, respectively.

Table 2. Comparison of 3-kind kernels based on RVM on session E.

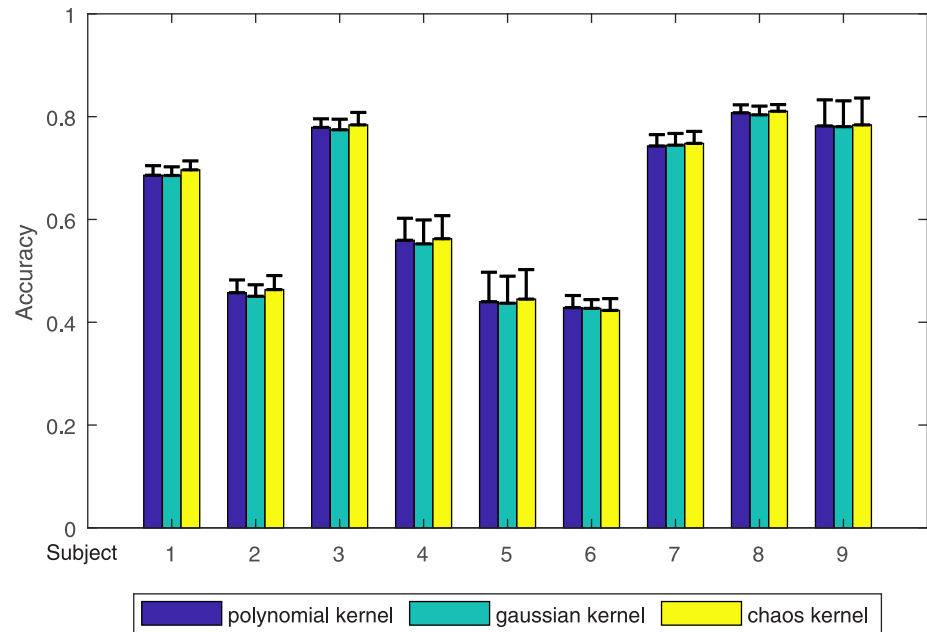| Subject	Polynomial kernel (mean ±std%) ($a = 1.26, d = 2$)	Gaussian kernel (mean ±std%) ($\sigma = 0.95$)	Chaos kernel (mean ±std%) ($\beta = 0.5$)
S1	67.8±1.5	68.2±1.4	**69.6±1.4**
S2	**44.2±1.9**	43.6±1.6	44.1±1.3
S3	78.9±1.4	79.0±1.3	**80.4±1.0**
S4	**59.8±1.9**	59.5±2.1	59.6±1.9
S5	49.3±1.9	48.7±2.0	**50.4±1.5**
S6	**44.2±1.3**	43.7±1.1	43.9±1.3
S7	75.7±1.8	75.7±1.7	**76.4±1.5**
S8	81.3±1.6	80.9±1.6	**81.4±1.2**
S9	83.1±1.2	**83.2±1.0**	**83.2±0.9**
Average	64.9±15.1	64.7±15.3	**65.4±15.3**

Fig 10. The overall classification result on two sessions.

Fig 10 exhibits the final results of 3 kernel functions in each subject. The value of the accuracy and standard deviation are computed by the datum from Tables 1 and 2. Except for subject 6, the result achieved by the proposed kernel function is better than that obtained by the other two kernel functions. The best accuracy is 81.05%, obtained by subject 8 using the chaos kernel.

Table 3 presents the comparison between the RVM algorithm, based on three kernels, and SVM. All the computations are carried on a Lenovo computer (CPU 3.3 GHz) with the software Matlab (2015b).

Table 4 presents the comparison between the proposed method and the competition methods [47]. We can see that the main difference between our method and the second method is the difference of classifiers, however, the results are very close. The result of the proposed method is obviously more effective than the third, fourth and fifth methods.

It is evident in Fig 11 that at 0.6s, the polynomial kernel RVM and the chaos kernel RVM converge, and the Gaussian kernel RVM converges at 0.65s. They yield almost the same convergence rate.

Table 3. Comprehensive comparison of RVMs based on three kernels and SVM.

	kernels	Processing time for training procedure (s)	Processing time for training procedure (s)	RVs or SVs	Accuracy (%)
RVM	Polynomial kernel ($a = 1.26; d = 2$)	0.8767	0.0074	16	63.1±15.2
	Gaussian kernel ($\sigma = 0.95$)	1.2189	0.0067	17	62.8±15.3
	Chaos kernel ($\beta = 0.5$)	1.3899	0.008	13	63.6±15.4
SVM	Polynomial kernel ($g = 3.0314$)	7.2395	0.005	183	64.5±15.7

Table 4. Comparison of the proposed method and the competition methods.

Methods	Propressing	Features	Classification	Kappa
1	bandpass filter (4-40Hz)	OVR-FBCSP	Naive Bayes Parzen Window classifier	0.57
2	bandpass filter (8-30Hz)	OVO-CSP	LDA & Bayesian classifier	0.52
3	bandpass filter (8-25Hz)	CSP	Two-hierarchical SVM classifier	0.31
4	NTSPP+CSP	CSP	LDA & SVM	0.30
5	bandpass filter (8-25Hz)	CSP	Two-hierarchical SVM classifier	0.29
Proposed method	bandpass filter (3-24Hz)	cmOVO-CSP	RVM	0.515

Fig 12 depicts the trained weights computed by three RVM kernel functions. The horizontal axis denotes the index of the RVs corresponding to the learned weight. There are no more than four learned weights in each graph, which produces the sparse classification results. The vertical axis denotes the value of the learned weight. The value of the learned weights in the different kernel functions varies so greatly, up to orders of magnitude. This is so because those weights are computed by the corresponding kernel function. While we pay attention to the difference between the positive and negative weights in each graph, which is the key indicator to distinguish the features, it is obvious that the greater the difference, the easier it is to distinguish the two-class signals.

Fig 13 shows the influence of the parameter beta, in the chaos kernel, on the classification results. It can be seen from the Fig 13 that the overall trend is that as the value of the parameter beta becomes larger, the classification accuracy is decreased, while at the point of $\beta = 0.5$, we get the best classification accuracy.
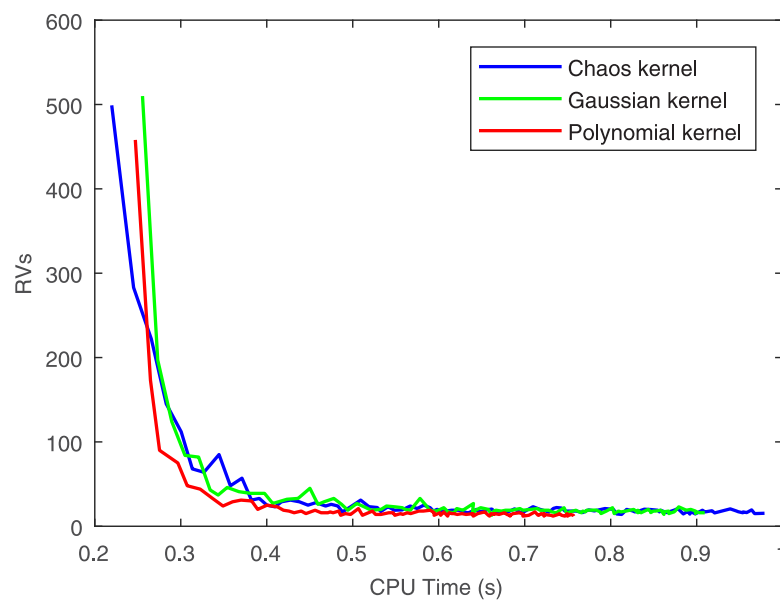

Fig 11. The relationship of RVs and CPU time.
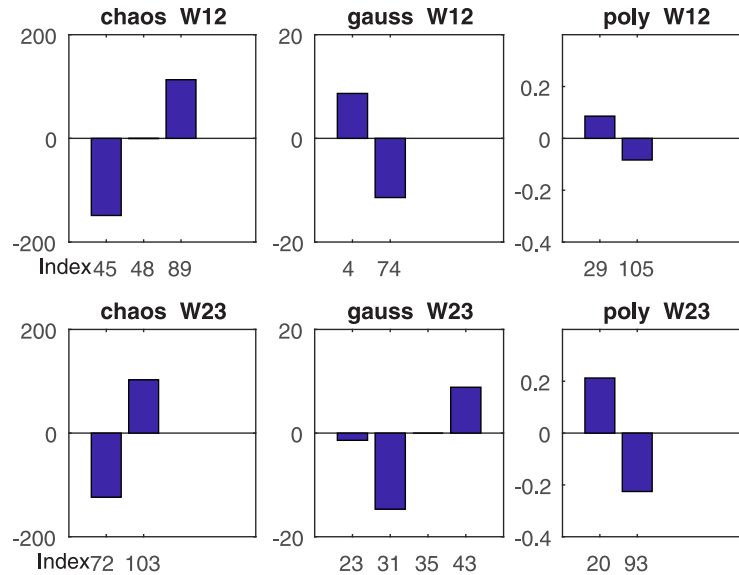
Fig 12. Comparison of two typical RVM training weights of three kernel functions.

Discussion and conclusion

In this paper, a new chaos kernel was proposed for relevance vector machine to classify four-class EEG of motor imagery. The raw EEG signals are addressed by 3-24 Hz band-pass filter to remove artifacts and uncorrelated frequency bands. And the four-class classification problem is transformed into six two-class problem under the framework of OVO-CSP method. Then the feature vectors extracted by OVO-CSP are sent to the RVM for classification.

Fig 13. The influence of parameter β on the accuracy of classification.

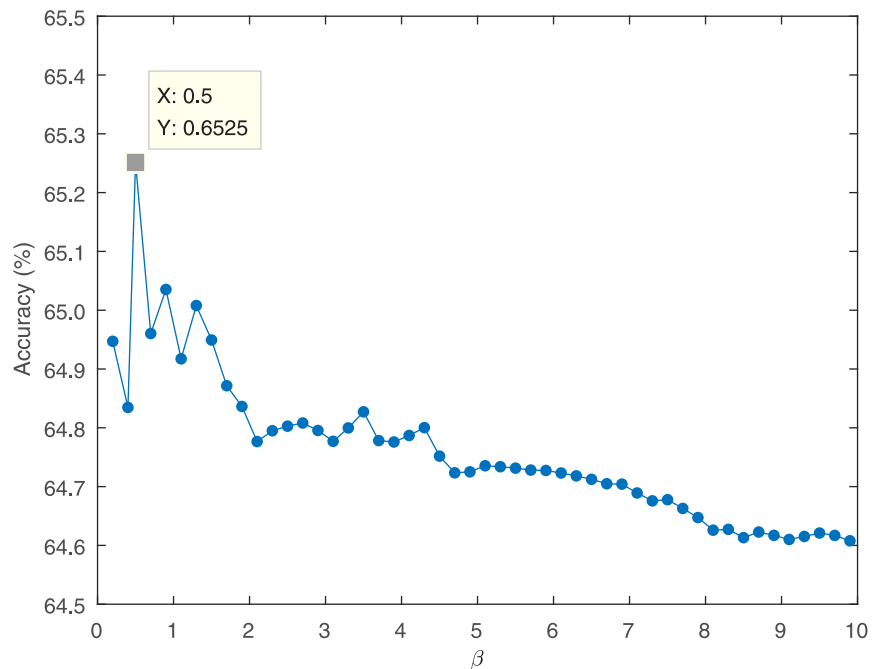Compared with SVM, RVM has a significant advantage in training time and the number of relevance vectors (or support vector for SVM), as seen in Table 3. If the number of RVs or SVs is smaller, the classification model will be simpler. Especially, as the input samples increase, the complex classification model will produce a huge amount of computation, resulting in slow classification. At the same time, the cost of smaller RVs is that the classification accuracy is slightly lower (see Table 3), thus it is also evident that the classification accuracy of RVM is not as good as that of the SVM. Although the test time of the RVM is slightly longer than that of the SVM, it requires a much shorter training time than the SVM.

The proposed kernel function evolved from the distribution function of a chaos system. For a long time, researchers have been studying the phenomenon of chaos in the brain. The EEG signal sometimes appears as a chaotic phenomenon when the neural network changes from one sequential structure to another [38]. Furthermore, the pioneers have proved that the EEG signal is controlled by several independent dynamic variables. This is very similar to the production of a chaotic system. Thus, we boldly predict that there more features of the EEG may be found by using chaos theory.

Although the proposed kernel function does not have significant advantages compared with the Gaussian and Polynomial kernel functions, it suggested another approach for EEG signal analysis, which is different from the classic SVM method. In the future, further attempts will be made to find a more suitable kernel function that stems from a chaos system.

Supporting information

S1 Table. Data for drawing Fig 5. Data for probability distribution of the logistic map with $A = 4$.
(XLSX)

S2 Table. Data for drawing Fig 6. Data for describing the Lyapunov exponent of the transformed system with A.
(XLSX)

S3 Table. Data for drawing Fig 7. Data for describing the probability distribution of the transformed chaos system.
(XLSX)

S4 Table. Data for drawing Fig 9. Data for describing the results of the feature extraction in Fig 9.
(XLSX)

S5 Table. Detailed data of the results in Table 1 (session T). This is the detailed result of 30 experiments per subject and Table 1 is the result of averaging 30 experiments.
(XLSX)

S6 Table. Detailed data of the results in Table 2 (session E). This is the detailed result of 30 experiments per subject and Table 2 is the result of averaging 30 experiments.
(XLSX)

S7 Table. Data for drawing Fig 10. Average accuracy and standard deviation of two sets (Tables 1 and 2) of data for 9 subjects.
(XLSX)

S8 Table. Data for drawing Fig 11. The data for describing convergence curves of three kernels in Fig 11.
(XLSX)

S9 Table. Data for drawing Fig 12. The data of the trained weights computed by three RVM kernel functions.
(XLSX)

S10 Table. Data for drawing Fig 13. The detailed data for explain the influence of the parameter beta, in the chaos kernel, on the classification results.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Enzeng Dong.

Data curation: Jigang Tong.

Formal analysis: Guangxu Zhu.

Funding acquisition: Enzeng Dong, Shengzhi Du.

Investigation: Yingjie Jiao.

Methodology: Enzeng Dong.

Project administration: Jigang Tong.

Resources: Yingjie Jiao.

Software: Guangxu Zhu.

Supervision: Chao Chen, Shengzhi Du.

Validation: Chao Chen.

Writing – original draft: Guangxu Zhu.

Writing – review & editing: Shengzhi Du.

References

1. Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, et al. Brain-computer interface technology:A review of the first international meeting. IEEE Trans. Rehabil. Eng. 2000; 8(2):164–173. https://doi.org/10.1109/TRE.2000.847807 PMID: 10896178

2. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. T Brain-computer interfaces for communication and control. Clin Neurophysiol. 2002 Jun; 113(6):767–791. https://doi.org/10.1016/S1388-2457(02)00057-3 PMID: 12048038

3. Chaudhary U, Birbaumer N, Ramos-Murguialday A. Corrigendum: Brain- computer interfaces for communication and rehabilitation. Nat Rev Neurol. 2017 Feb 17; 13(3):191. https://doi.org/10.1038/nrneurol.2017.23 PMID: 28211450

4. Magwire MM, Bayer F, Webster CL, Cao C, Jiggins FM. Spelling with non-invasive Brain-Computer Interfaces-Current and future trends. J Physiol Paris. 2011 Jan-Jun; 105(1-3):106–114. https://doi.org/10.1016/j.jphysparis.2011.08.003

5. Ramadan RA, Vasilakos AV. Brain computer interface: control signals review. Neurocomputing. 2017 Feb; 223:26–44. https://doi.org/10.1016/j.neucom.2016.10.024

6. Pfurtscheller G, Brunner C, Schloegl A, Lopes da Silva FH. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. Neuroimage. 2006 May 15; 31(1):153–159. https://doi.org/10.1016/j.neuroimage.2005.12.003 PMID: 16443377

7. Pfurtscheller G, Neuper C. Motor imagery and direct brain-computer communication. Proceedings of the IEEE. 2001 Jul; 89(7):1123–1134. https://doi.org/10.1109/5.939829

8. Edelman BJ, Baxter B, He B. EEG Source Imaging Enhances the Decoding of Complex Right-Hand Motor Imagery Tasks. IEEE Trans Biomed Eng.? 2016 Jan; 63(1):4–14. https://doi.org/10.1109/TBME.2015.2467312 PMID: 26276986

9. Zhang Y, Wang Y, Jin J, Wang X. Sparse Bayesian Learning for Obtaining Sparsity of EEG Frequency Bands Based Feature Vectors in Motor Imagery Classification. Int J Neural Syst. 2017 Mar; 27(2):1650032. https://doi.org/10.1142/S0129065716500325 PMID: 27377661

10. Blankertz B, Lemm S, Treder M, Haufe S, Mueller K-R. Single-trial analysis and classification of ERP components-A tutorial. Neuroimage. 2011 May 15; 56(2):814–825. https://doi.org/10.1016/j.neuroimage.2010.06.048 PMID: 20600976

11. Subasi A, Gursoy MI. EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Systems with Applications. 2010 Dec; 37(12):8659–8666. https://doi.org/10.1016/j.eswa.2010.06.065

12. Bascil MS, Tesneli AY, Temurtas F. Spectral feature extraction of EEG signals and pattern recognition during mental tasks of 2-D cursor movements for BCI using SVM and ANN. Australas Phys Eng Sci Med. 2016 Sep; 39(3):665–676. https://doi.org/10.1007/s13246-016-0462-x PMID: 27376723

13. Mateo-Sotos J, Torres AM, Sanchez-Morla EV, Santos JL. An Adaptive Radial Basis Function Neural Network Filter for Noise Reduction in Biomedical Recordings. Circuits, Systems, and Signal Processing. 2016 Dec; 35(12):4463–4485. https://doi.org/10.1007/s00034-016-0281-z

14. Triggiani AI, Bevilacqua V, Brunetti A, Lizio R, Tattoli G, Cassano F, et al. Classification of Healthy Subjects and Alzheimer's Disease Patients with Dementia from Cortical Sources of Resting State EEG Rhythms: A Study Using Artificial Neural Networks. Front Neurosci. 2017 Jan 26; 10:604. https://doi.org/10.3389/fnins.2016.00604 PMID: 28184183

15. Tang Z, Li C, Sun S. Single-trial EEG classification of motor imagery using deep convolutional neural networks. Optik. 130. 2017 Feb; 130:11–18. https://doi.org/10.1016/j.ijleo.2016.10.117

16. Direito B, Teixeira CA, Sales F, Castelo-Branco M, Dourado A. A Realistic Seizure Prediction Study Based on Multiclass SVM. Int J Neural Syst. 2017 May; 27(3):1750006. https://doi.org/10.1142/S012906571750006X PMID: 27873554

17. Acharya UR, Sree VS, Chattopadhyay S, Yu W, Alvin APC. Application of recurrence quantification analysis for the automated identification of epileptic EEG signals. Int J Neural Syst. 2011 Jun; 21(3):199–211. https://doi.org/10.1142/S0129065711002808 PMID: 21656923

18. Huang MW, Chen CW, Lin WC, Ke SW, Tsai CF. SVM and SVM Ensembles in Breast Cancer Prediction. PLoS One. 2017 Jan 6; 12(1):e0161501. https://doi.org/10.1371/journal.pone.0161501 PMID: 28060807

19. Nicolaou N, Georgiou J. Detection of epileptic electroencephalogram based on Permutation Entropy and Support Vector Machines. Expert Systems with Applications. 2012 Jan; 39(1):202–209.? https://doi.org/10.1016/j.eswa.2011.07.008

20. Naseer N, Hong K-S. fNIRS-based brain-computer interfaces: a review. Front Hum Neurosci. 2015 Jan 28; 9:3. https://doi.org/10.3389/fnhum.2015.00003 PMID: 25674060

21. Temko A, Thomas E, Marnane W, Lightbody G, Boylan G. EEG-based neonatal seizure detection with Support Vector Machines. Clin Neurophysiol. 2011 Mar; 122(3):464–73. https://doi.org/10.1016/j.clinph.2010.06.034 PMID: 20713314

22. Hsu WY. Assembling A Multi-Feature EEG Classifier for Left-Right Motor Imagery Data Using Wavelet-Based Fuzzy Approximate Entropy for Improved Accuracy. Int J Neural Syst. 2015 Dec; 25(8):1550037. https://doi.org/10.1142/S0129065715500379 PMID: 26584583

23. Dong E, Li C, Li L, Du S, Belkacem AN, Chen C. Classification of multi-class motor imagery with a novel hierarchical SVM algorithm for brain-computer interfaces. Med Biol Eng Comput. 2017 Oct; 55(10):1809–1818. https://doi.org/10.1007/s11517-017-1611-4 PMID: 28238175

24. Zhang Y, Zhou G, Jin J, Zhao Q, Wang X, Cichocki A. Sparse bayesian classification of EEG for brain-computer interface. IEEE Trans Neural Netw Learn Syst. 2016 Nov; 27(11):2256–2267. https://doi.org/10.1109/TNNLS.2015.2476656 PMID: 26415189

25. Zhang Y, Zhou G, Jin J, Zhang Y, Wang X, Cichocki A. Sparse Bayesian multiway canonical correlation analysis for EEG pattern recognition. Neurocomputing. 2017 Feb 15; 225:103–110. https://doi.org/10.1016/j.neucom.2016.11.008

26. Tipping ME, Smola A Sparse Bayesian learning and the relevance vector machine. Journal of Machine Learning Research. 2001 Jan; 1(3):211–244.

27. Zhang Y, Zhou W, Yuan S. Multifractal Analysis and Relevance Vector Machine-Based Automatic Seizure Detection in Intracranial EEG. Int J Neural Syst. 2015 Sep; 25(6):1550020. https://doi.org/10.1142/S0129065715500203 PMID: 25986754

28. Belardinelli P, Jalava A, Gross J, Kujala J, Salmelin R. Optimal spatial filtering for brain oscillatory activity using the Relevance Vector Machine. Cogn Process. 2013 Nov; 14(4):357–69. https://doi.org/10.1007/s10339-013-0568-y PMID: 23729235

29. De Martino F, de Borst AW, Valente G, Goebel R, Formisano E. Predicting EEG single trial responses with simultaneous fMRI and Relevance Vector Machine regression. Neuroimage. 2011 May 15; 56(2):826–36. https://doi.org/10.1016/j.neuroimage.2010.07.068 PMID: 20691274

30. Gupta R, Laghari KuR, Falk TH. Relevance vector classifier decision fusion and EEG graph-theoretic features for automatic affective state characterization. Neurocomputing. 2016 Jan; 174:875–884. https://doi.org/10.1016/j.neucom.2015.09.085

31. Han M, Sun LL, Hong XJ, Han J. Automatic classification of epileptic EEG signals based on AR model and relevance vector machine. Chinese Journal of Biomedical Engineering. 2011 Dec; 30:864–870.

32. Matsumoto M, Hori J. Classification of silent speech using support vector machine and relevance vector machine. Applied Soft Computing. 2014 Jul; 20:95–102. https://doi.org/10.1016/j.asoc.2013.10.023

33. Pijn JP, Van Neerven J, Noest A, Lopes da Silva FH. Chaos or noise in EEG signals; dependence on state and brain site. Electroencephalogr Clin Neurophysiol. 1991 Nov; 79(5):371–81. https://doi.org/10.1016/0013-4694(91)90202-F PMID: 1718710

34. Stam CJ. Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. Clin Neurophysiol. 2005 Oct; 116(10):2266–301. https://doi.org/10.1016/j.clinph.2005.06.011 PMID: 16115797

35. Wang X, Meng J, Tan G, Zou L. Research on the relation of EEG signal chaos characteristics with high-level intelligence activity of human brain. Nonlinear Biomed Phys. 2010 Apr 27; 4(1):2. https://doi.org/10.1186/1753-4631-4-2 PMID: 20420714

36. Sarbadhikari S, Chakrabarty K. Chaos in the brain: a short review alluding to epilepsy, depression, exercise and lateralization. Med Eng Phys. 2001 Sep; 23(7):445–55. https://doi.org/10.1016/S1350-4533(01)00075-3 PMID: 11574252

37. Tsai ZR, Chang YZ, Zhang HW, Lin CT. Relax the chaos-model- based human behavior by electrical stimulation therapy design. Computers in Human Behavior. 2016 Nov; 67:151–160. https://doi.org/10.1016/j.chb.2016.10.006

38. Xiao D, Yang H, Zhou S. Extraction of 40 Hz EEG bursts for chaos analysis of brain function. IEEE Eng Med Biol Mag. 1997 Jul-Aug; 16(4):27–32. https://doi.org/10.1109/51.603645 PMID: 9241517

39. Vikram TSH, Sreenithi P, Harikumar R. Analysis of chaos in EEG signals for estimation of drowsiness and classification of epilepsy risk levels. International Conference on Networking. 2010 Feb; pp. 147-156.

40. Ge S, Wang R, Yu D. Classification of Four-Class Motor Imagery Employing Single-Channel Electroencephalography. PLoS One. 2014 Jun 20; 9(6):e98019. https://doi.org/10.1371/journal.pone.0098019 PMID: 24950192

41. Lotte F, Guan C. Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms. IEEE Trans Biomed Eng. 2011 Feb; 58(2):355–62. https://doi.org/10.1109/TBME.2010.2082539 PMID: 20889426

42. Sun H, Bi L, Chen B. Partitioned Common Spatial Pattern Method for single trial EEG Signal classification in Brain-Computer Interface System. Automatika. 2016 Jul; 57(1):66–75. https://doi.org/10.7305/automatika.2016.07.1078

43. Dong E, Zhu G, Chen C. Classification of four categories of EEG signals based on relevance vector machine. IEEE International Conference on Mechatronics and Automation. 2017 Aug; pp. 1024-1029.

44. Chagas S L, Eisencraft M, Lima CAM. Relevance Vector Machine Applied to EEG Signals. Conference: Anais do XXVI Simpósio Brasileiro de Telecomunicações. 2008 Jan;pp. 1-6.

45. MacKay D. The evidence framework applied to classification networks. Neural Computation. 1992 Sep; 4(5):720–736. https://doi.org/10.1162/neco.1992.4.5.720

46. Brunner C, Leeb R, Müller-Putz G, Schlögl A, Pfurtscheller G. BCI Competition 2008-Graz data set A. Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology: 136-142.

47. http://www.bbci.de/competition/iv/results/index.html♯dataset2a.