Adaptive KF-SVM Classification for Single Trial EEG in BCI

Banghua Yang[1(✉)], Chengcheng Fan[1], Jie Jia[2(✉)], Shugeng Chen[2], and Jianguo Wang[1]

[1] Key Laboratory of Power Station Automation Technology, Department of Automation, College of Mechatronics Engineering and Automation, Shanghai University, No. 149, Yanchang Road, Shanghai 200072, China
yangbanghua@shu.edu.cn

[2] Department of Rehabilitation Medicine, Huashan Hospital, Fudan University, No. 12, Wulumuqi Middle Road, Shanghai 200040, China

Abstract. Single trial electroencephalogram classification is indispensable in online brain–computer interfaces (BCIs) A classification method called adaptive Kernel Fisher Support Vector Machine (KF-SVM) is designed and applied to single trial EEG classification in BCIs. The adaptive KF-SVM algorithm combines adaptive idea, SVM and within-class scatter inspired from kernel fisher. Firstly, the within-class scatter matrix of a feature vector is calculated. And to construct a new kernel, this scatter is incorporated into the kernel function of SVM. Ultimately, the recognition result is calculated by the SVM whose kernel has been changed. The proposed algorithm simultaneously maximizes the discrimination between classes and also considers the within-class dissimilarities, which avoids some disadvantages of traditional SVM. In addition, the within-class scatter matrix of adaptive KF-SVM is updated trial by trail, which enhances the online adaptation of BCIs. Based on the EEG data recorded from seven subjects, the new approach achieved higher classification accuracies than the standard SVM, KF-SVM and adaptive linear classifier. The proposed scheme achieves the average performance improvement of 5.8%,5.2% and 3.7% respectively compared to other three schemes.

Keywords: Brain computer interface (BCI) · Support vector machine (SVM) · Adaptive classification · Kernel fisher · Within-class scatter

1 Introduction

Brain–computer interfaces (BCIs) is a direct communication pathway between brain and external device which is independent from muscle pathway [1]. The inherent nonstationarities existed in the sampled EEG data makes a principal problem in electroencephalogram based brain computer interfaces [2]. These nonstationarities are caused by many factors such as variations of the concentration and excitation level, fluctuations in the involved subjects' mental task, the impedance variations or positions movement of the electrodes, affection of feedback, fatigue, and swallowing and blinking artifacts [3–6]. In addition, the characteristics of EEG signals may vary

M. Fei et al. (Eds.): LSMS/ICSEE 2017, Part I, CCIS 761, pp. 35–45, 2017.
DOI: 10.1007/978-981-10-6370-1_4

significantly from person to person [7]. In order to track non-stationary EEG, reduce the subject training, some adaptive algorithms are investigated extensively, from its feature extraction to its classification [8–12]. This present work mainly emphasizes on the analysis of motor imagery electroencephalogram signal in BCI for the ideal of adaptive classification. As a kind of spontaneous EEG signal, Motor imagery signal is commonly used in BCIs because of the reason that it is much more natural compared with evoked EEG signal for example as P300 and VEP (Visual Evoked Potential).

Some typical classification methods include adaptive linear discriminant analysis (LDA), adaptive support vector machine (SVM) The adaptive linear discriminant analysis is investigated for left and right hand motor imagery classification. By using Kalman filter, the adaptive LDA parameters are constantly updated trial by trial, which get better results than the non-adaptive LDA classification. It is a simple and efficient method, but it cannot avoid shortcomings of linear classifiers. An adaptive SVM classification for BCIs is proposed in [10], which attains much higher classification accuracy than the non-adaptive SVM. However, the classification performance of standard SVMs is restricted by many factors such as data noise, unbalanced data points, complexity of data points and so on. When the classification examples are difficult data classification for standard SVMs, it is impossible to obtain optimal classification results [11]. To classify imagery EEG data in BCIs, this article describes a novel and adaptive method called adaptive kernel fisher SVM (KF-SVM), which combines adaptive idea, SVM, kernel fisher inspiring within-class scatter.

Due to the powerful classification ability, the SVM has become a major method to make electroencephalogram(EEG) based BCI classification, thus to overwhelm other classifiers in many other applications [13, 14]. Nevertheless, the SVM algorithm only considers the discrimination between classes, but neglects within-class scatters information. As an algorithm improvement, the proposed algorithm not only maximizes the discrimination between classes but also considers the within-class dissimilarities inspired from kernel fisher simultaneously. Firstly, the within-class scatter matrix of a feature vector is calculated. And after that this scatter is incorporated into the kernel function of SVM to reconstruct a new kernel. Finally, the recognition result is calculated by SVM whose kernel has been changed. At the same time, the adaptability of the proposed algorithm is improved by updating the within-class scatter simultaneously and continuously. The proposed method was tested on dataset collected from seven subjects. Its performance is compared to other classification algorithms including SVM, KF-SVM and adaptive LDA. The highlighted point in this paper is the employment of common spatial patterns (CSP) method for feature extraction of the three classifiers.

2 Methods

2.1 Dataset

Dataset was experimented from laboratory. A 16-channel electrode cap is used for EEG signal recording. The authors didn't use too many channels considering that fewer channels were more practical for online application of BCIs. The EEG biological amplifier was developed by Tsinghua University research group with its high quality of precision. The EEG signals were transformed by a 24-bit A/D converter and collected with the sampling frequency of 100 Hz through acquisition software. In the process of the experiment, each of seven subjects was asked to complete one session containing 60 trials. For each trial, a 4 s left or right hand motor imagination task is included. For each subject, the total accepted sessions are eight and corresponding eight datasets for each subject were acquired. These subjects had no experience of the BCI experiment. The aim of selecting them was to check if the proposed algorithm had better generalization capability for naïve BCI users. The datasets were filtered between 8 Hz and 30 Hz by band-pass filter (the usual range for motor imagery EEG data).

2.2 Research Scheme

A flowchart of the adaptive KF-SVM is illustrated in Fig. 1. In part 1, the trials from session 1 are extracted features by the CSP. For motor imagery features recognition, the PP-SVM via adding within-class scatter is used firstly and it takes five-fold cross validation secondly, so for the testing data, the average recognition accuracies are calculated across these five folds. In part 2, the parameters of adaptive KF-SVM via adding within-class scatter are initialized by the training data of all the trials in session 1. After completing initialization, the obtained trials from session 2 to session 8 are used to evaluate the performance of the classifier. At the simultaneous time, within class scatter of the proposed adaptive algorithm is updated trial by trial continuously.

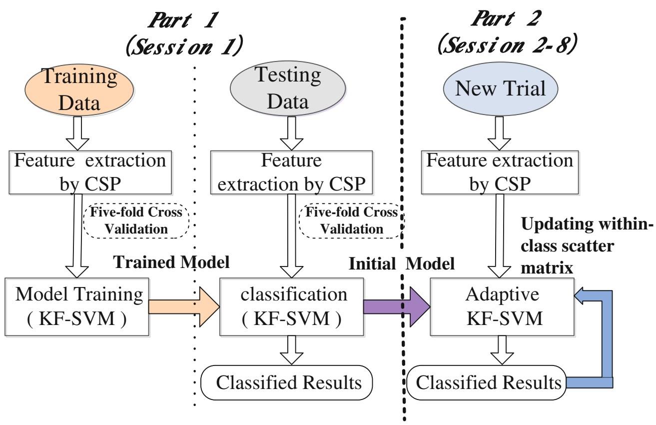

Fig. 1. Flowchart of adaptive KF-SVM.

2.3 SVM Introduction

The purpose of SVM algorithm is to search the optimal hyperplane to separate the two classes of samples [11, 12]. The SVM has good generalization ability and its optimization problem is defined as:

$$\min(1/2)w^T Iw + \gamma(1/2)\sum_{i=1}^{n+m} e_i^2 \tag{1}$$

where the parameter I represent the identity matrix, the regularisation term γ remains positive, $n + m$ reflects the training samples number, $\{e_i\}_{i=1}^{n+m}$ represents the error vector, $y_i \in \{-1, 1\}$ represents the label for sampling, $\phi(x)$ is a mapping function, the vector W is weighting vector and scalar $b \in R$ is bias vector. According to formula (1), the Lagrange multipliers are used to solve the optimisation problem:

$$\begin{bmatrix} 0 & -Y^T \\ Y & zz^T + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ I \end{bmatrix} \tag{2}$$

where $(zz^T)_{ij} = y_i y_j k(x_i, x_j)$, α reflects the dual variable vector, and this function $k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ is called a kernel function.

From formula (2), the non-zero parameter a and parameter b can be obtained. So the decision function could be shown as the following equations:

$$f(x) = \sum_{i=1}^{n+m} a_i y_i k(x_i, x) + b \tag{3}$$

According to formula (3), the classification surface of SVM depends on two classes' boundary samples and misclassified samples. In other words, the SVM depends on samples that makes a non-zero and ignores samples within the boundaries. This may cause deviation when facing strong noise interference or an uneven data distribution. The output modality of the SVM classifier is expressed as the following equation:

$$z(x) = \begin{cases} 1 & f(x) > 0 \\ -1 & f(x) < 0 \end{cases} \tag{4}$$

2.4 KF-SVM

Inspiring from Kernel fisher, the SVM could make an integration with within-class scatter.. Fisher discriminant analysis can make input data relationship conversion from non-linear into linear form. Fisher's linear discriminant is found by maximising $J(w)$ equation:

$$J(w) = w^T M w / w^T N w, \ N = \sum_{j=1,2} k_j (I - 1_{l_j}) k_j^T, \ k_j = k(x_{n+m}, x^j) \tag{5}$$

Where the parameter N represent the within-class scatter, k_j represents not only the matrix of (n + m)*n or (n + m)*m but also the kernel function matrix of class j, m represents the first class's sample number, n is the second class's sample number, the parameter I represents the unit matrix, 1_{l_j} represents the matrix with all the inner elements are n-1 or m-1, w represents the transform vector, and M denotes the distance between classes.

On practical data, the Kernel fisher has got excellent results. Its classification error rate could be as low as or even lower compared with the SVM. Considering that SVM algorithm ignores within-class scatter and meanwhile to improve the adaption of a classifier, so the current work proposes adaptive KF-SVM classification method through the way to add within-class scatter inspired from kernel fisher. Combing with the kernel fisher, the SVM can be optimized which could be described as the following formula

$$\begin{aligned} &\min(1/2)w^T(\lambda N+I)w+\gamma(1/2)\sum_{i=1}^{n+m}e_i^2 \\ &s.t.\ y_i[(w\cdot\phi(x))+b]=1-e_i\ e_i\geq 0, i=1,2,3\ldots,n+m \end{aligned} \tag{6}$$

In a method similar to the SVM, according upper Eq. (6), the optimization matter could be resolved by implementing Eq. (7), with $(zz^T)_{ij}=y_iy_jk^*(x_i,x_j)$. The kernel function has been changed after incorporating within-class scatter to the kernel function. Now, $k^*(x_i,x_j)=\varphi(x_i)*\Sigma^{-1}*\varphi(x_j)^T$, where $\Sigma=\lambda N+I$. To obtain $k^*(x_i,x_j)$, on the basis of the Mercer condition, the afterwards deduction could be achieved: suppose input space $x=x_1,x_2,\ldots,x_n$ and k(x, y) is a symmetric function, for all the involved samples, the matrix can be given by: $k=(k(x_i,x_j))(i,j=0,1,\ldots,n)$, which is apparently a symmetric matrix. Definitely, there exists an orthogonal matrix to form $P^TkP=\Lambda$, where Λ represents a diagonal matrix which consists of eigenvalue λ_i, then the eigenvector of λ_i is $v_t=(v_{t1},v_{t2},\ldots,v_{tn})^T$, where n represents the size of the sample. The input space could be mapped as the following equation:

$$\phi: x_i\rightarrow(\sqrt{\lambda_1 v_{1i}};\sqrt{\lambda_2 v_{2i}};\ldots,\sqrt{\lambda_n v_{ni}})\in R^n(i=1,2,\ldots,n)$$

Where $\phi_i(x_j)=\sqrt{\lambda_i v_{ij}}$. Then:

$$\begin{aligned} \phi(x_i)&=(\phi_1(x_i),\phi_2(x_i),\ldots,\phi_n(x_i)) \\ <\phi(x_i),\phi(x_j)> &=\sum_{t=1}^{n}\lambda_t v_{ti}v_{tj}=k(x_i,x_j) \end{aligned}$$

So $k^*(x_i,x_j)=\varphi(x_i)*\Sigma^{-1}*\varphi(x_j)^T$ is obtained and new a and b values are acquired from formula (7). On the basis of this new function with its new a and b values, f(x)function can be expressed as the following equation:

$$f(x)=\sum_{i=1}^{n+m}a_iy_ik^*(x_i,x)+b \tag{7}$$

The decision function f(x) in Eq. (7) varies the traditional SVM algorithm in formula (3). The final class label is determined by whiten the spatial coefficient matrix S and transformation matrix P jointly. So the KF-SVM is formed.

2.5 Comparison Between SVM and KF-SVM

To further explain the difference between SVM and KF-SVM, optimization objective, constraint function and formula are given in Table 1.

Table 1. Comparison between SVM and KF-SVM

	SVM	KF-SVM
Optimization objective, Constraint Function	$\min\ (1/2)w^T Iw + \gamma(1/2)\sum_{i=1}^{n+m} e_i^2$ $s.t.\ y_i[(w \cdot \phi(x)) + b] = 1 - e_i$ $e_i \geq 0, i = 1, 2, \ldots n+m$	$\min\ (1/2)w^T(\lambda N + I)w + \gamma(1/2)\sum_{i=1}^{n+m} e_i^2$ $s.t.\ y_i[(w \cdot \phi(x)) + b] = 1 - e_i$ $e_i \geq 0, i = 1, 2, \ldots, n+m$
Formula and steps	Step1: $k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$	Step1: $k^*(x_i, x_j) = \phi(x_i) * \Sigma^{-1} * \phi(x_j)^T$
	Step2: $f(x) = \sum_{i=1}^{n+m} a_i y_i k(x_i, x) + b$	Step2: $f(x) = \sum_{i=1}^{n+m} a_i y_i k * (x_i, x) + b$
	Step3: $z(x) = \begin{cases} 1 & f(x) > 0 \\ -1 & f(x) < 0 \end{cases}$	Step3: $z(x) = \begin{cases} 1 & f(x) > 0 \\ -1 & f(x) < 0 \end{cases}$

2.6 Adaptive KF-SVM

A whole scheme is described in section "research scheme" and we have known that the scheme is divided into two parts including part 1 and part 2. Detailed flow of adaptive KF-SVM to part 1 and part 2 is shown in Fig. 2.

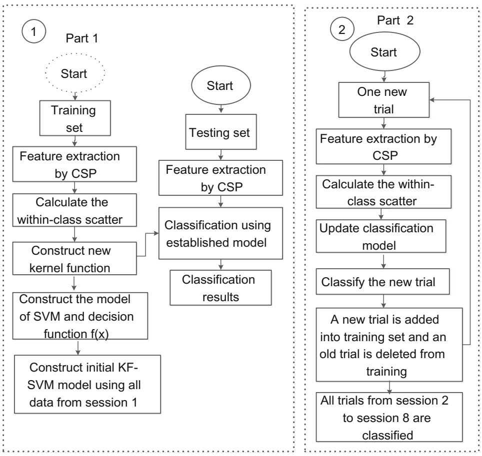

Fig. 2. Detailed flow of adaptive KF-SVMP

The realization steps involved in proposed adaptive algorithm are given below:
Initialization Process (part 1)

Step 1: The feature vector of training data are extracted by the CSP.
Step 2: The within-class scatter N of feature vectors are calculated by formula (5).
Step 3: The new kernel function is constructed by $k^*(x_i, x_j) = \phi(x_i) * \Sigma^{-1} * \phi(x_j)^T$.
Step 4: The decision function $f(x)$ and final classification result are obtained according formula (2) and CSP whiten transformation matrix and CSP spatial coefficient matrix respectively. An initial model is formed.
Step 5: The established initial model is used to classify testing data.
Step 6: The five-fold cross validation is implemented according above steps and afterwards the average recognition rate for the testing data is calculated across these five folds.
Step 7: The adaptive KF-SVM is initialized by implementing the session 1 trials as training data.

Validation Process (part 2)

Step 1: The feature vector of a new single trial i from next sessions is extracted by CSP.
Step 2: The initialized model obtained in training process is used to classify the feature vector and the corresponding values of the decision function $f(x(i))$ and class label can be obtained.
Step 3: The new trial is added into training dataset and at the same time the oldest trial is deleted from the training dataset in order to keep training number constant.
Step 4: The within-class scatter is updated by using new training dataset. New training dataset are trained by the KF-SVM to get a new model, which will replace the old model.
Step 5: When next trial comes, the above four steps are repeated again until all data from session 2 to session 8 are finished. The classification results of all data from these seven sessions are implemented to estimate the performance of the algorithm.

3 Discussions and Results

3.1 Classifier

For the purpose to check the performance of the upper proposed algorithm, the four bellowing classification methods were used in the experiments.

1. SVM: as described above, it uses the following RBF kernel function:

$$K(x, y) = \exp\left(-\frac{||x-y||^2}{\sigma^2}\right) \tag{8}$$

2. KF-SVM: the within-class scatter based on kernel fisher is added into the SVM, so the optimization object, kernel function and decision function of SVM are changed. The model of KF-SVM is established based on these changed parameters. The detailed process can be seen in above Section.
3. Adaptive LDA: the LDA is used as a classifier. The classification result from latest trial will serve as the label of the trial. Meanwhile, the latest trial is added into training set and the oldest trial in the training set will be deleted. By this manner, the classification model is trained and updated trial by trial.
4. Adaptive KF-SVM: the adaptive idea is added into the KF-SVM and the classification model is continuously updated trial by trial.

3.2 Parameters Selection in Proposed Algorithm

In the analysis of the proposed algorithm, two indispensable parameters are required to make optimal selection, that is to say, the parameter of kernel function σ and the parameter of within-class scatter λ. Here, the best σ and λ are selected from the sets $\sigma \in \{1, 2, 3, \ldots, 20\}$ and $\lambda \in \{0, 1, 2, 3. \ldots\ldots, 100\}$ respectively. Five-fold cross-validation was performed using multigroup values of σ and λ on the training data and those resulting in the minimum error are chosen.

3.3 Experiment Results

For each trial, the 2.1 s to 3.1 s time duration is selected as the signal processing period from the total 4 s for each imagination based trial. So every trial's time length was 1 s. A 4-dimensional feature vector $F = \{f_1, f_2, f_3, f_4,\}$ is calculated to every trial. Finally, a 60×4 (60 trials, four dimension of every trial) feature vector from session 1 is used for the initialization of classification model. After initialization, the trials from session 2 to session 8 are classified to make estimation of the different classifiers' performances. Table 2 lists accuracies for seven subjects (Sub1–Sub7) with different classifiers.

3.4 Discussions

The row in blue shade in Table 2 shows that the average classification accuracy of the adaptive KF-SVM is 5.8%, 5.2% and 3.7 higher than those obtained with other three methods (SVM, KF-SVM, Adaptive LDA) in part 2. Meanwhile, the row in grey shade indicates that the adaptive KF-SVM has a slight higher classification accuracy over SVM and adaptive LDA during session 1. It should be noted that the adaptive KF-SVM has the same classification accuracy and model with the KF-SVM because they have the same initialization procedure during session 1. The main difference between adaptive KF-SVM and KF-SVM lies in part 2 (from session 2 to session 8). In order to more clearly describe differences among these three classification methods, all classification results are plotted in Fig. 3. From this figure, we can see that classification accuracies varied significantly from person to person, in which subject Sub7 obtains the best performance with proposed method over other subjects. Classification performance of SVM and KF-SVM classifiers declines during part 2 in relation to part 1, implying that the nonstationarity of EEG affects the classification performance.

Table 2. Classification accuracies with different classifiers

	Subject	Classification accuracy (%)			
		SVM	KF-SVM ($\lambda = 21$, $\sigma = 5$)	Adaptive LDA	Adaptive KF-SVM ($\lambda = 21$, $\sigma = 5$)
Part 1 (Session 1, Initialization)	Sub1	75	77	74.3	77
	Sub2	71	63.33	65.6	63.33
	Sub3	87.5	89.5	85.1	89.5
	Sub4	78.5	86.25	80.8	86.25
	Sub5	75	75	73.2	75
	Sub6	77.5	79	73.8	79
	Sub7	87.5	85	78.5	85
	Mean	78.9	79.3	75.9	79.3
Part 2 (From Session 2 to Session 8)	Sub1	75	75	76.9	85
	Sub2	70	62.5	69.1	72.5
	Sub3	87.5	87.5	86.1	87.5
	Sub4	77.5	80	83.1	85
	Sub5	72.5	75	76.2	77.5
	Sub6	70	75	73.9	77.5
	Sub7	81.5	83	83.5	89.5
	Mean	76.3	76.9	78.4	82.1

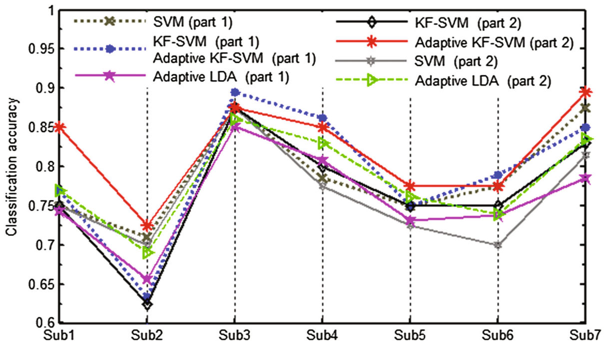

Fig. 3. Comparison of classification results among different subjects and methods

However, adaptive KF-SVM and LDA presents minimal reduction or even some increase, which implies that adaptive idea is helpful to analyze varied EEG. Meanwhile, adaptive KF-SVM achieves the best performance and has better adaptation over other three algorithms.

4 Conclusion

This paper presents a new adaptive KF-SVM classification method combining the kernel fisher, adaptive idea, and the SVM. It takes advantage of the properties of the kernel fisher and overcomes some defects inherent to the SVM. Meanwhile, the within-class scatter in the adaptive KF-SVM is continuously updated trial by trial, which could improve adaptation of the new classifier. The upper proposed method is verified by comparing it with other three algorithms. The results show that the upper proposed method could obtain satisfying recognition accuracy. It may be practical for online application in BCIs. The next-step research should take aim at the verification of the algorithm on bigger data.

References

1. Wolpaw, J.R., Birbaumer, N., Heetderks, W.J., McFarland, D.J., Peckham, P.H., Schalk, G., et al.: Brain–computer interface technology: a review of the first international meeting. IEEE Trans. Rehabil. Eng. **8**, 164–173 (2000)
2. Arvaneh, M., Guan, C.T., Ang, K.K., Quek, C.: Optimizing spatial filters by minimizing within-class dissimilarities in electroencephalogram-based brain–computer Interface. IEEE Trans. Neural Netw. Learn. Syst. **24**(4), 610–619 (2013)
3. Liu, G.Q., Zhang, D.G., Meng, J.J., Huang, G., Zhu, X.Y.: Unsupervised adaptation of electroencephalogram signal processing based on fuzzy C-means algorithm. Int. J. Adapt. Control Signal Process. **26**, 482–495 (2012)
4. Wang, Y.-T., Nakanishi, M., Wang, Y., Wei, C.-S., Cheng, C.-K., Jung, T.-P.: An online brain-computer interface based on SSVEPs measured from non-hair-bearing areas. IEEE Trans. Neural Syst. Rehabil. Eng. **25**(1), 11–18 (2017)
5. Mondini, V., Mangia, A.L., Cappello, A.: EEG-Based BCI system using adaptive features extraction and classification procedures. Comput. Intell. Neurosci. **2016**, 1–14 (2016)
6. Yan, W., Ge, Y.B.: A novel method for motor imagery EEG adaptive classification based biomimetic pattern recognition. Neurocomputing **116**, 280–290 (2013)
7. Seo, D., et al.: Wireless recording in the peripheral nervous system with ultrasonic neural dust neuron neuroresource wireless recording in the peripheral nervous system with ultrasonic neural dust. Neuron **91**, 1–11 (2016)
8. Yang, H., Sakhavi, S., Kai, K.A., Guan, C.: On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification. In: 2015 IEEE 37th Annual International Conference on Engineering in Medicine and Biology Society (EMBC), pp. 2620–2623 (2015)
9. Hsu, W.Y.: EEG-based motor imagery classification using enhanced active segment selection and adaptive classifier. Comput. Biol. Med. **41**, 633–639 (2011)
10. Wu, D.: Online and Offline Domain Adaptation for Reducing BCI Calibration Effort. IEEE Transactions on human-machine Systems, pp. 1–14 (2016)
11. Ye, Q.L., Zhao, C.X., Ye, N.: A New SVM classification approach via minimum within-class variance. J. Comput. Inf. Syst. **6**(1), 39–45 (2011)
12. Spüler, M., Rosenstiel, W., Bogdan, Martin: Adaptive SVM-based classification increases performance of a MEG-based brain-computer interface (BCI). In: Villa, A.E.P., Duch, W., Érdi, P., Masulli, F., Palm, G. (eds.) ICANN 2012. LNCS, vol. 7552, pp. 669–676. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33269-2_84

13. Kaper, M., Meinicke, P., Grossekathoefer, U., Lingner, T., Ritter, H.: BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm. IEEE Trans. Biomed. Eng. **51**, 1073–1076 (2004)
14. Hema Rajini, N., Bhavani, R.: Automatic classification of computed tomography brain images using ANN, k-NN and SVM. AI Soc. **29**, 97–102 (2014)