31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

# Regularized Common Spatial Patterns with Generic Learning for EEG Signal Classification

Haiping Lu, Konstantinos N. Plataniotis and Anastasios N. Venetsanopoulos

*Abstract*— **The common spatial patterns (CSP) algorithm is commonly used to extract discriminative spatial filters for the classification of electroencephalogram (EEG) signals in the context of brain-computer interfaces (BCIs). However, CSP is based on a sample-based covariance matrix estimation. Therefore, its performance is limited when the number of available training samples is small. In this paper, the CSP method is considered in such a small-sample setting. We propose a regularized common spatial patterns (R-CSP) algorithm by incorporating the principle of generic learning. The covariance matrix estimation in R-CSP is regularized through two regularization parameters to increase the estimation stability while reducing the estimation bias due to limited number of training samples. The proposed method is tested on data set IVa of the third BCI competition and the results show that R-CSP can outperform the classical CSP algorithm by 8.5% on average. Moreover, the regularization introduced is particularly effective in the small-sample setting.**

## I. INTRODUCTION

Noninvasive brain computer interfaces (BCIs) aim to translate brain activity into sequences of control commands so that a subject, such as a disable person, can communicate with the outside world, such as a computer, without using the peripheral nervous system [1]. Electroencephalography (EEG) [2] has been widely used to capture the electric field generated by the central nervous system and to infer the user's intention for noninvasive BCI applications because of its simplicity, inexpensiveness and high temporal resolution [3]. EEG signals are recorded from multiple electrodes placed on the scalp of a subject, resulting in multichannel time series.

Multichannel EEG signals typically have low signal-to-noise ratio (SNR), giving a rather blurred image of the brain activity [4]. Thus, they are not directly usable in BCI applications. The common spatial patterns (CSP) method is an algorithm frequently employed to extract the most discriminative information from EEG signals. It was first suggested for binary classification of EEG trials in [5]. CSP is a spatial filtering method that seeks projections with the most differing power/variance ratios in the feature space. The projections are calculated by a simultaneous diagonalization of the covariance matrices of two classes [4]. Usually, only the first a few most discriminative filters obtained are useful for classification.

Haiping Lu and K. N. Plataniotis are with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, M5S 3G4, Canada {haiping, kostas}@comm.toronto.edu

A. N. Venetsanopoulos is with Ryerson University, Toronto, ON, M5B 2K3, Canada tasvenet@ryerson.ca

However, it should be noted that the classical CSP algorithm relies on a sample-based covariance matrix estimation. Therefore, its performance is limited when there are only a small number of samples available for training. Similar small-sample problems have arisen in many other applications. In the regularized discriminant analysis (RDA) [6], regularization was first introduced to tackle the small-sample problem for linear and quadratic discriminant analysis. It was pointed out by J. H. Friedman in [6] that small number of training samples tends to result in biased estimation of eigenvalues. At the same time, the sample-based covariance estimates from such poorly-posed problem are usually highly unstable. Accordingly, two regularization parameters are introduced to take these undesirable effects into account. In more recent works, regularization has been successfully employed for small-sample problem in various applications such as face recognition [7], [8] and gait recognition [9]–[11].

In this paper, we investigate the regularization of the CSP algorithm in the small-sample setting. A regularized CSP (R-CSP) algorithm is proposed by regularizing the covariance matrix estimation in common spatial pattern extraction. Although robust covariance matrix estimation is an old topic [6], to the best of the authors' knowledge, this is the first attempt to apply the regularization techniques in [6] to the problem of EEG signal classification. As in [6], there are two regularization parameters involved. The first regularization parameter controls the shrinkage of a subject-specific covariance matrix towards a "generic" covariance matrix to improve the estimation stability, based on the principle of generic learning [12]. The second regularization parameter controls the shrinkage of the sample-based covariance matrix estimation towards a scaled identity matrix to account for the bias due to limited number of samples. Data set IVa from the BCI Competition III was used to evaluate the performance of the proposed algorithm, with significant improvement demonstrated, especially in the small-sample setting.

## II. REGULARIZED COMMON SPATIAL PATTERNS

This section starts with the covariance matrix estimation in the classical CSP method. Regularized covariance matrix estimation is then proposed, built upon the regularization introduced in [6] and generic learning in [12]. Lastly, the R-CSP algorithm for EEG signal classification is presented.

### A. Covariance Matrix Estimation in the CSP Method

The CSP algorithm is widely used in processing multichannel EEG signals during imagined hand movement [3],

978-1-4244-3296-7/09/$25.00 ©2009 IEEE

[5], [13], [14]. It extracts several spatial filters so that the variances of the filtered signals are the most discriminative for two classes. A given single EEG trial with $N$ channels is represented as a matrix $\mathbf{E}$ of size $N \times T$, where $T$ denotes the number of samples in each channel for a single trial. The normalized sample covariance matrix $\mathbf{S}$ of a trial $\mathbf{E}$ is obtained as [5]

$$\mathbf{S} = \frac{\mathbf{E}\mathbf{E}^T}{\mathrm{tr}(\mathbf{E}\mathbf{E}^T)}, \tag{1}$$

where the superscript '$^T$' indicates the transpose of a matrix and $\mathrm{tr}(\cdot)$ is the trace of a matrix (sum of the diagonal elements). In this paper, we consider only binary classification problems so there are two classes only, indexed by $c = \{1, 2\}$. For simplicity, we assume that there are $M$ trials in each class available for training for a subject of interest, indexed by $m$ as $\mathbf{E}_{(c,m)}$, where $m = 1, ..., M$. Thus, each trial has a corresponding covariance matrix $\mathbf{S}_{(c,m)}$.

The average spatial covariance matrix for each class is then calculated as [5]

$$\bar{\mathbf{S}}_c = \frac{1}{M}\sum_{m=1}^{M} \mathbf{S}_{(c,m)}, c = \{1, 2\}. \tag{2}$$

### *B. Regularization of the Covariance Matrix Estimation in Small-Sample Setting*

The discriminative patterns extracted by the CSP algorithm is based on the sample covariance matrix estimation in (2). However, this estimation problem could be poorly posed when there are only a small number of training trials [6]. In this case, the parameters estimated can be highly unstable, giving rise to high variance. Furthermore, the low signal-to-noise ratio for EEG signals makes the estimation variance even higher.

The method of regularization has been proved to be effective in solving the small-sample problem by biasing the estimates away from their sample-based values toward more "physically plausible" values [6]. This reduces the variance associated with sample-based estimates while tending to increase bias. The bias variance trade-off is generally regulated by one or more regularization parameters that control the strength of the biasing.

Following the regularization technique introduced in [6], the regularized average spatial covariance matrix for each class is calculated as

$$\hat{\mathbf{\Sigma}}_c(\beta, \gamma) = (1-\gamma)\hat{\mathbf{\Sigma}}_c(\beta) + \frac{\gamma}{N}\mathrm{tr}[\hat{\mathbf{\Sigma}}_c(\beta)] \cdot \mathbf{I}, \tag{3}$$

where $\beta$ ($0 \leq \beta \leq 1$) and $0 \leq \gamma \leq 1$ are two regularization parameters, $\mathbf{I}$ is an identity matrix of size $N \times N$, and $\hat{\mathbf{\Sigma}}_c(\beta)$ is defined as following:

$$\hat{\mathbf{\Sigma}}_c(\beta) = \frac{(1-\beta) \cdot \mathbf{S}_c + \beta \cdot \hat{\mathbf{S}}_c}{(1-\beta) \cdot M + \beta \cdot \hat{M}}. \tag{4}$$

In (4), $\mathbf{S}_c$ is the sum of the sample covariance matrices for all $M$ training trials in class $c$:

$$\mathbf{S}_c = \sum_{m=1}^{M} \mathbf{S}_{(c,m)}, \tag{5}$$

while $\hat{\mathbf{S}}_c$ is the sum of the sample covariance matrices for a set of $\hat{M}$ generic training trials $\{\mathbf{E}_{(c,\hat{m})}\}$ in class $c$:

$$\hat{\mathbf{S}}_c = \sum_{\hat{m}=1}^{\hat{M}} \mathbf{S}_{(c,\hat{m})}. \tag{6}$$

In the above definitions, $\mathbf{S}_{(c,m)}$ and $\mathbf{S}_{(c,\hat{m})}$ are the normalized sample covariance matrix as defined in (1).

The term $\hat{\mathbf{S}}_c$ is introduced in (4) to reduce the variance in the covariance matrix estimation and it tends to produce more stable results. For its construction, we adopt the idea of generic learning introduced for one-training-sample face recognition [12]. For a subject whose EEG signals are to be classified, the training process employs the corresponding EEG trials collected for other subjects in the regularization term $\hat{\mathbf{S}}_c$.

From (4), the first regularization parameter $\beta$ controls the degree of shrinkage of the training sample covariance matrix estimation to the generic covariance matrix estimation. From (3), the second regularization parameter $\gamma$ controls the degree of shrinkage toward a multiple of the identity matrix, with the average eigenvalue of $\hat{\mathbf{\Sigma}}_c(\beta)$ as the multiplier. This shrinkage effectively decreases the larger eigenvalues and increases the smaller ones since the estimates in (1) tend to bias the eigenvalues in the opposite direction, especially in the small-sample setting [6]. It should be noted that when $\beta = \gamma = 0$, the R-CSP algorithm is reduced to the conventional CSP algorithm so CSP can be considered as a special case of R-CSP.

### *C. The Regularized CSP for EEG Signal Classification*

Next, the procedures in the classical CSP method [5] is followed to get the R-CSP algorithm. The composite spatial covariance in R-CSP is formed and factorized as

$$\hat{\mathbf{\Sigma}}(\beta, \gamma) = \hat{\mathbf{\Sigma}}_1(\beta, \gamma) + \hat{\mathbf{\Sigma}}_2(\beta, \gamma) = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{U}}^T, \tag{7}$$

where $\hat{\mathbf{U}}$ is the matrix of eigenvectors and $\hat{\mathbf{\Lambda}}$ is the diagonal matrix of corresponding eigenvalues. In this paper, we adopt the convention that the eigenvalues are sorted in descending order.

Next, the whitening transformation is obtained as

$$\hat{\mathbf{P}} = \hat{\mathbf{\Lambda}}^{-1/2}\hat{\mathbf{U}}^T. \tag{8}$$

$\hat{\mathbf{\Sigma}}_1(\beta, \gamma)$ and $\hat{\mathbf{\Sigma}}_2(\beta, \gamma)$ are whitened as

$$\tilde{\mathbf{\Sigma}}_1(\beta, \gamma) = \hat{\mathbf{P}}\hat{\mathbf{\Sigma}}_1(\beta, \gamma)\hat{\mathbf{P}}^T \tag{9}$$

and

$$\tilde{\mathbf{\Sigma}}_2(\beta, \gamma) = \hat{\mathbf{P}}\hat{\mathbf{\Sigma}}_2(\beta, \gamma)\hat{\mathbf{P}}^T, \tag{10}$$

respectively. $\tilde{\mathbf{\Sigma}}_1(\beta, \gamma)$ can then be factorized as

$$\tilde{\mathbf{\Sigma}}_1(\beta, \gamma) = \hat{\mathbf{B}}\hat{\mathbf{\Lambda}}_1\hat{\mathbf{B}}^T, \tag{11}$$

and the full projection matrix is formed as

$$\hat{\mathbf{W}}_0 = \hat{\mathbf{B}}^T\hat{\mathbf{P}}. \tag{12}$$

For the most discriminative patterns, only the first and last $\alpha$ columns of $\hat{\mathbf{W}}_0$ are kept to form $\hat{\mathbf{W}}$, which is of size

$N \times Q$, where $Q = 2\alpha$. For feature extraction, a trial $\mathbf{E}$ is first projected as

$$\hat{\mathbf{Z}} = \hat{\mathbf{W}}^T \mathbf{E}. \tag{13}$$

Then, a $Q$-dimensional feature vector $\hat{\mathbf{y}}$ is formed from the variance of the rows of $\hat{\mathbf{Z}}$ as

$$\hat{y}_q = \log\left(\frac{\mathrm{var}(\hat{\mathbf{z}}_q)}{\sum_{q=1}^{Q} \mathrm{var}(\hat{\mathbf{z}}_q)}\right), \tag{14}$$

where $\hat{y}_q$ is the $q$-th component of $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}_q$ is the $q$-th row of $\hat{\mathbf{Z}}$, and $\mathrm{var}(\hat{\mathbf{z}}_q)$ is the variance of the vector $\hat{\mathbf{z}}_q$.

For classification, the Fisher's discriminant analysis is used. A projection is solved to maximize the ratio of the between-class scatter over the within-class scatter. In classification, the class label of the nearest neighbor is assigned to a test trial.

## III. EXPERIMENTAL RESULTS

In this section, the experimental results are presented for performance evaluation.

### A. Data Description

Date set IVa from the BCI Competition III [15] is used to test the proposed R-CSP algorithm. This data set was provided by Fraunhofer FIRST Intelligent Data Analysis Group, and Campus Benjamin Franklin of the Charité, University Medicine Berlin (Neurophysics Group) [16].

In each capturing session, a subject was presented with visual cues for 3.5 seconds, indicating one of the three motor imageries that the subject should perform: left hand, right hand and right foot. The presentation of the target cues was separated by intervals of random length from 1.75 to 2.25 seconds for the subject to relax. Only the right hand and right foot motor imageries are provided for five healthy subjects ('$aa$', '$al$', '$av$', '$aw$', and '$ay$'). The EEG signals were recorded using 118 electrodes at the positions of the extended international 10/20 system.

There are 280 trials for each subject, i.e., 140 trials for each class, per subject. The number of training trials ($M$) and the number of test trials for each subject are indicated in the second and third rows of Table I, respectively. The number of testing trials is much more than $M$ for subjects '$av$', '$aw$', and '$ay$'. The EEG signals were down-sampled to 100 Hz and band-pass filtered to the 7-30 Hz frequency band [4].

### B. Performance Evaluation and Discussion

From the description above, we have $N = 118$, $T = 350$, and $M$ as indicated in Table I. In this experimental evaluation, only the two most discriminative CSP features are used for classification for preliminary evaluation, i.e., $Q = 2$ ($\alpha = 1$). To study the potential of R-CSP, we examine its performance over a testing grid of $(\beta, \gamma)$ values defined by the outer product of $\beta = [0 : 0.1 : 1]$ and $\gamma = [0 : 0.1 : 1]$. The best results obtained and the corresponding regularization parameters $(\beta^*, \gamma^*)$ are reported. In future work, schemes to determine $(\beta^*, \gamma^*)$ will be investigated,

TABLE I
CLASSIFICATION PERFORMANCE ON DATA SET IVA OF THE BCI COMPETITION III.

| Subject | $aa$ | $al$ | $av$ | $aw$ | $ay$ | Mean |
|---|---|---|---|---|---|---|
| $M$ | 84 | 112 | 42 | 28 | 14 | - |
| #Test per class | 56 | 28 | 98 | 112 | 126 | - |
| CSP CCR in % | 62.5 | 83.9 | 57.1 | 51.3 | 73.4 | 65.7 |
| R-CSP CCR in % | 69.6 | 83.9 | 64.3 | 70.5 | 82.5 | 74.2 |
| $\beta^*$ | 0 | 0/0.3 | 0.6 | 0.3 | 0 | - |
| $\gamma^*$ | 0.2 | 0 | 0.1 | 0 | 0.1 | - |

such as cross validation [6]. The generic training trials for a particular subject (e.g., '$ay$') consist of the trials from all the other subjects (e.g., '$aa$', '$al$', '$av$', and '$aw$'). Thus, $\hat{M} = 560$. As pointed out earlier, R-CSP with $\beta = \gamma = 0$ is equivalent to the classical CSP. The correct classification rate (CCR) is used to measure the classification accuracy.

The CCRs obtained from CSP and R-CSP, as well as $(\beta^*, \gamma^*)$ are reported in Table I for each subject. The average CCRs (mean) are also included in the last column in the table, where an average improvement of $8.5\%$ is achieved, indicating the effectiveness of the regularization procedure introduced in this paper.

Fig. 1 depicts the performance gains for the five subjects as well as their average. It is observed that the regularization procedure is the least effective on subject '$al$', which has 112 trials per class for training while only 28 trials per class for test. For the other four subjects, the average improvement in CCR is $10.65\%$. In particular, there is an improvement of $19.2\%$ for subject '$aw$'. These results demonstrate the strength of R-CSP in dealing with small-sample setting in EEG signal classification.

The regularization parameter $\beta$ controls the shrinkage of covariance matrix estimation for the subject of interest to the generic covariance matrix estimation based on other subjects. From Table I, a small value of $\beta$ tends to have better performance, except for subject '$av$'. Fig. 2 illustrates

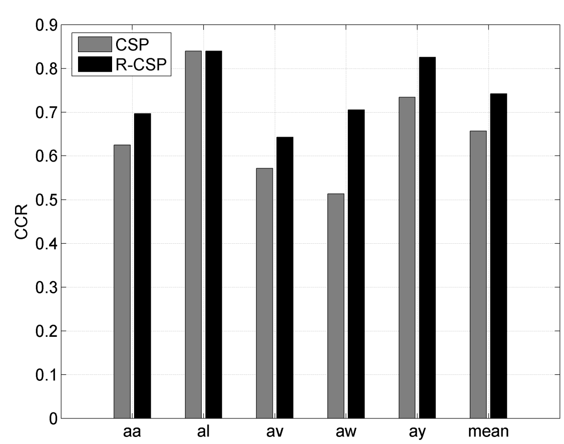
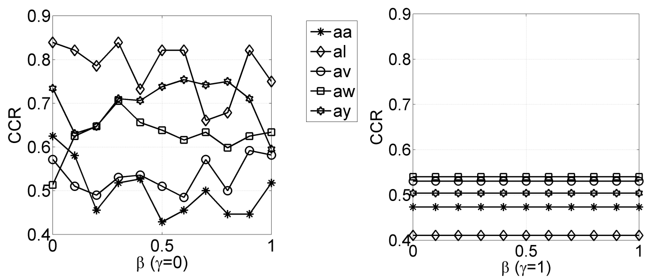



Fig. 1. Illustration of the improvement in CCRs due to regularization.




(a) (b)

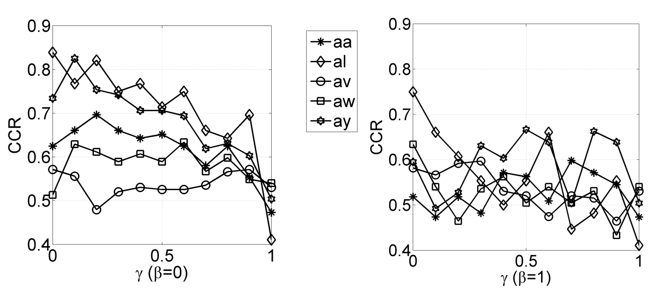
Fig. 2. Illustration of the CCRs for various values of $\beta$ with (a) $\gamma = 0$, and (b) $\gamma = 1$.

the variation of CCRs for different values of $\beta$ for each subject with $\gamma = 0$ and $\gamma = 1$. Fig. 2(a) indicates that the performance could be sensitive to the value of $\beta$. While for $\gamma = 1$, the covariance estimates become scaled identity matrices and therefore, the patterns extracted are not useful for classification, as shown in Fig. 2(b).

The regularization parameter $\gamma$ controls the shrinkage of the covariance matrices to scaled identity matrices. From Table I, for all subjects, a small value of $\gamma$ ($\leq 0.2$) gives the best classification performance. Fig. 3 demonstrates the variation of CCRs for different values of $\gamma$ for each subject with $\beta = 0$ and $\beta = 1$. Fig. 3(a) is consistent with the observation from Table I that a small value of $\gamma$ is preferred for good regularization results. When $\beta = 1$, from Fig. 3(b). the performance is more sensitive to the value of $\gamma$ and the results are poorer than the case of $\beta = 0$ in general (except for subject '$av$').

## IV. CONCLUSIONS

The classical CSP algorithm uses a sample-based covariance matrix estimation. This results in limited performance with small number of training samples. This work addresses the problem of discriminative common spatial pattern extraction in the small-sample setting. A regularized CSP algorithm has been introduced based on regularization and generic learning. Two regularization parameters are involved in reguarlizing the covariance estimates. One is to increase the estimation stability and the other is to reduce the estimation bias. Experimental results on data set IVa of BCI competition III demonstrated the effectiveness of the proposed R-CSP method, particularly in the small-sample setting. Nonetheless, further studies are needed for appropriate regularization parameter determination.




(a) (b)

Fig. 3. Illustration of the CCRs for various values of $\gamma$ with (a) $\beta = 0$, and (b) $\beta = 1$.

## V. ACKNOWLEDGMENTS


The authors would like to thank Fraunhofer FIRST, Intelligent Data Analysis Group (Klaus-Robert Müller, Benjamin Blankertz), and Campus Benjamin Franklin of the Charité - University Medicine Berlin, Department of Neurology, Neurophysics Group (Gabriel Curio) for providing the data set used in this paper.


## REFERENCES


[1] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1991–2000, Aug. 2008.

[2] S. Sanei and J. A. Chambers, *EEG signal processing*. John Wiley & Sons, 2007.

[3] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (sbcsp) for brain-computer interface," in *Proc. the 3rd Annual Int. IEEE EMBS Conf. on Neural Engineering*, May 2007, pp. 204–207.

[4] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.

[5] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined handmovement," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 4, pp. 441–446, Apr. 2000.

[6] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 165–175, Mar. 1989.

[7] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition," *Pattern Recognition Letter*, vol. 26, no. 2, pp. 181–191, 2005.

[8] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized discriminant analysis for the small sample size problem in face recognition," *Pattern Recognition Letter*, vol. 24, no. 16, pp. 3079–3087, 2003.

[9] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition," *IEEE Trans. Neural Networks*, vol. 20, no. 1, pp. 103–123, Jan. 2009.

[10] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear discriminant analysis with regularization for gait recognition," in *Proc. Biometrics Symposium 2007*, September 2007, doi:10.1109/BCC.2007.4430540.

[11] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Boosting LDA with regularization on MPCA features for gait recognition," in *Proc. Biometrics Symposium 2007*, September 2007, doi:10.1109/BCC.2007.4430542.

[12] J. Wang, K. N. Plataniotis, J. Lu, and A. N. Venetsanopoulos, "On solving the face recognition problem with one training sample per subject," *Pattern Recognition*, vol. 39, no. 9, pp. 1746–1762, 2006.

[13] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selection in motor imagery based brain-computer interface," in *Proc. the 27th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, Sept. 2005, p. 5392C5395.

[14] X. Yong, R. K. Ward, and G. E. Birch, "Robust common spatial patterns for EEG signal preprocessing," in *Proc. the 30th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, Aug. 2008, pp. 2087–2090.

[15] B. Blankertz, K.-R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. del R. Millán, M. Schröder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 14, no. 2, pp. 153–159, June 2006.

[16] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, "Boosting bit rates in non-invasive eeg single-trial classifications by feature combination and multi-class paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, June 2004.