Sparse Kernel Machines for motor imagery EEG classification

Vangelis P. Oikonomou[1], Spiros Nikolopoulos[1], Panagiotis Petrantonakis [1], and Ioannis Kompatsiaris[1]

***Abstract*— Brain-computer interfaces (BCIs) make human-computer interaction more natural, especially for people with neuro-muscular disabilities. Among various data acquisition modalities the electroencephalograms (EEG) occupy the most prominent place due to their non-invasiveness. In this work, a method based on sparse kernel machines is proposed for the classification of motor imagery (MI) EEG data. More specifically, a new sparse prior is proposed for the selection of the most important information and the estimation of model parameters is performed using the bayesian framework. The experimental results obtained on a benchmarking EEG dataset for MI, have shown that the proposed method compares favorably with state of the art approaches in BCI literature.**

I. Introduction

The term Brain Computer Interface (BCI) refers to an artificial communication channel between the human brain and an external environment using machines (mostly computers) [1]. A BCI system measures and translates the brain activity into control signals that can be used to operate new assistive devices for people with motor disabilities, people who are totally paralyzed, or locked-in. Besides medical applications, BCI systems can also facilitate the communication between humans and machines/computers through more natural interfaces that extent beyond mouse and keyboards. BCI systems using EEG recordings can be divided into two major categories. In the first case, the user's activity is generated in response to an external stimulus, such as in the case of Steady State Visual Evoked Potentials (SSVEPs). In the second category, the user voluntarily changes his/her brain waves without the presence of an external stimulus, such as in the case of Motor Imagery (MI) BCI [2]. In our study we are concerned with the case of MI BCI systems.

One major challenge in MI BCI is the real-time extraction of reliable information from noisy data in the form of relevant features. The existing feature extraction approaches are dominated by methods estimating the distribution of energy in various domains, such as the time domain, the frequency domain, the time - frequency (t-f) domain, the wavelet domain and the spatial domain [3]–[9]. One of the most popular and efficient algorithms for MI BCI relies on the use of Common Spatial Patterns (CSP). More specifically, this algorithm is a feature extraction method that uses spatial filters to maximize the discriminability of two classes [6]. However, the CSP algorithm in its basic form is sensitive to noise while overfitting can rise in the case of small training sets. To overcome these problems in [7] regularized versions of the CSP algorithm were proposed, while filter banks in cooperation with the CSP algorithm (FBCSP) were used in [8].

In MI BCI, the extracted features are fed into a classifier to identify the user's mental state. In [3], various classifiers have been used for the identification of motor tasks. More specifically, a comparison between Linear Discriminant Analysis (LDA) and various extensions of Support Vector Machines (SVM) is provided. The main outcome of [3] is that the use of SVMs with a gaussian kernel is the most appropriate classifier for the examined problem. In addition, in the same work a genetic algorithm was used to fine-tune the SVM, increasing considerably the overall tuning time of the system. Lately, in the BCI community, special attention is given to bayesian versions of LDA [10]–[13]. More specifically, in [10] Bayesian LDA (BLDA) is used as the main building block for the construction of semi-supervised algorithms, while in [12] Relevance Vector Machines (RVM) are used to select the most significant FBCSP features with a linear discriminant criterion for classification.

In this work, we propose a new method for the classification of MI EEG data. In particular, a new sparse prior is proposed to select the most important information under the concept of RVM. To construct this prior we have borrowed ideas from [14] and [15] and we have employed the Variational Bayesian (VB) framework to deal with the increased computational complexity of using a prior inside a bayesian framework. The proposed method for sparse kernel machines has been tested using a benchmarking MI EEG dataset and has been found to compare favorably with SVM and Variational RVM (VRVM) [16], especially in cases where the training set is small.

II. Methodology

Let $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N \in \Re^D$ be a set of EEG feature vectors, where N is the number of training samples.

$$\mathbf{y} = \mathbf{X}\theta + \mathbf{e}, \tag{1}$$

The vector $\mathbf{y} \in \Re^N$ contains 1's and -1's, with the n-th element being 1 if the n-th feature vector belongs to the first class, otherwise -1. The matrix $\mathbf{X} \in \Re^{N \times D}$ contains the EEG features vectors, $\mathbf{x}_i, i = 1, \cdots, N$ and $\mathbf{e}$ denotes the noise of the model following a gaussian distribution with zero mean and precision (inverse variance) β. Finally, $\theta \in \Re^D$ is a vector containing the weights of each feature.

A kernel machine is a linear regression model where the input feature vector $\mathbf{x}_n$ has been transformed into the

[1]V. P. Oikonomou, S. Nikolopoulos, P. Petrantonakis and I. Kompatsiaris are with the Information Technologies Institute, Centre for Research and Technology Hellas, CERTH-ITI, 6th km Charilaou-Thermi Road, 57001 Thermi-Thessaloniki, Greece. {viknmu, nikolopo, ppetrant, ikom}@iti.gr

kernelized feature vector $\phi(\mathbf{x}_n)$ through the use of a kernel function $\kappa(\cdot,\cdot)$ [17],

$$\boldsymbol{\phi}(\mathbf{x_n}) = [\boldsymbol{\kappa}(\mathbf{x}_n, \boldsymbol{\mu_1}), \boldsymbol{\kappa}(\mathbf{x}_n, \boldsymbol{\mu_2}), ..., \boldsymbol{\kappa}(\mathbf{x}_n, \boldsymbol{\mu_K})] \quad (2)$$

where μ_k is a set of K prototypes. The choice of prototypes is crucial for the subsequent analysis. One approach is to find clusters in the data and then to assign one prototype per cluster center. A more simple approach is to make each example $\mathbf{x}_i$ be a prototype, so we get

$$\phi(\mathbf{x}_n) = [\kappa(\mathbf{x}_n, \mathbf{x}_1), \kappa(\mathbf{x}_n, \mathbf{x}_2), ..., \kappa(\mathbf{x}_n, \mathbf{x}_N)] \quad (3)$$

We can see that we have as many parameters as data points. However, we can use a sparse prior for regression weights to efficiently select a subset of the training examples. Then, after applying Eq. (3) the linear model of Eq. (1) can be employed. However, the design matrix $\mathbf{X}$ contains the kernelized feature vectors of EEG data. More specifically, the linear regression takes the following form:

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{w} + \mathbf{e}, \quad (4)$$

where the matrix $\boldsymbol{\Phi} \in \Re^{N\times N}$ contains the kernelized features vectors of EEG data and $\mathbf{e}$ denotes the noise of the model following a gaussian distribution with zero mean and precision (inverse variance) β. Finally, $\mathbf{w} \in \Re^N$ is a vector containing the regression weights.

A. Sparse Bayesian Learning

Sparsity is a very helpful property since the processing is faster in a sparse representation where few coefficients reveal the information we are looking for. Hence, sparse priors help us to determine the model order in an automatic way and reduce its complexity. More specifically, the weights $\mathbf{w}$ are treated as a random variable with Gaussian prior of zero mean and variance a_i^{-1} for each element in the vector $\mathbf{w}$:

$$p(\mathbf{w}|\mathbf{a}) = \mathcal{N}(\mathbf{w}|0, \boldsymbol{\Lambda}) = \prod_{i=1}^{N} \mathcal{N}(w_i|0, a_i^{-1}), \quad (5)$$

where $\mathcal{N}$ is the symbol for Gaussian distribution. In Sparse Bayesian Learning literature, a common approach is to assume that the covariance matrix $\boldsymbol{\Lambda}$ is a diagonal matrix with elements $a_i^{-1}, i = 1, \cdots, D$. Each parameter a_i, which controls the prior distribution of the parameters $\mathbf{w}$, follows a Gamma distribution, so the overall prior over all a_i is a product of Gamma distributions given by: $p(\mathbf{a}) = \prod_{i=1}^{D} Gamma(a_i; b_a, c_a)$. This hierarchical prior over $\mathbf{w}$ is well known for its sparse properties [14], [17] and this approach was adopted in [12]. In our study we change the above prior by introducing one more parameter. More specifically, we assume that the covariance matrix $\boldsymbol{\Lambda}$ is a diagonal matrix with elements $a_i^{-1}\lambda_i^{-1}, i = 1, \cdots, N$. In our analysis, parameters λ_i are assumed known and deterministic quantities. Now the prior distribution of weights is given by:

$$p(\mathbf{w}|\mathbf{a};\boldsymbol{\lambda}) = \mathcal{N}(\mathbf{w}|0, \boldsymbol{\Lambda}) = \prod_{i=1}^{N} \mathcal{N}(w_i|0, a_i^{-1}\lambda_i^{-1}),$$

At this point it is worth to examine the marginal prior distribution of weight w_i by eliminating the hyperparameters a_i.

$$\begin{aligned} p(w_i;\lambda_i) &= \int p(w_i|a_i;\lambda_i)p(a_i)da_i \\ &= \int \mathcal{N}(w_i|0, a_i^{-1}\lambda_i^{-1}) Gamma(a_i; b_a, c_a) da_i \\ &\propto \Big(\frac{\lambda_i}{b_a}\Big)^{1/2} \Big[1 + \frac{\lambda_i w_i^2}{b_a}\Big]^{-(c_a+1/2)} \end{aligned} \quad (6)$$

Eq. 6 can be recognized as a Student-t distribution with zero mean, shape parameter c_a and scale parameter $\frac{b_a}{\lambda_i}$. We can see that parameter λ_i controls the scale of the Student-t distribution. In addition by adopting a procedure similar to [14] we can show that the weights have the improper prior $p(w_i) \propto \frac{1}{\lambda_i^{1/2}\cdot|w_i|}$. Now, by setting $\lambda_i \to 1/|w_i|$ we obtain $p(w_i) \propto \frac{1}{\cdot|w_i|^{1/2}}$ which can be recognized as an extremely "sparse" prior.

The overall precision (inverse variance) β of the noise follows a Gamma distribution: $p(\beta) = Gamma(\beta; b, c) = \frac{1}{\Gamma(c)}\frac{\beta^{(c-1)}}{b^c}\exp\Big\{-\frac{\beta}{b}\Big\}$, where b and c is the scale and the shape of the Gamma distribution, respectively. We use the Gamma distribution for the noise components for two reasons: First, this distribution is conjugate to the Gaussian distribution, which helps us in the derivation of closed form solutions, and second it places the positivity restriction on the overall variance and the scaling parameters.

So, the overall prior over model parameters $\{\mathbf{w}, \mathbf{a}, \beta\}$ is given by: $p(\mathbf{w}, \mathbf{a}, \beta; \boldsymbol{\lambda}) = p(\mathbf{w}|\mathbf{a}; \boldsymbol{\lambda}) \prod_{i=1}^{D} p(a_i)p(\beta)$. The likelihood of the data is given by:

$$\begin{aligned} p(\mathbf{y}|\mathbf{w}, \beta; \boldsymbol{\lambda}) &= \frac{\beta^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} \cdot \\ &\exp\Big\{-\frac{\beta}{2}(\mathbf{y} - \boldsymbol{\Phi}\mathbf{w})^T(\mathbf{y} - \boldsymbol{\Phi}\mathbf{w})\Big\} \end{aligned} \quad (7)$$

To apply the VB methodology [18] we need to define an approximate posterior based on one factorization over the parameters $\{\mathbf{w}, \mathbf{a}, \beta\}$. In our study we choose the following factorization: $q(\mathbf{w}, \mathbf{a}, \beta; \boldsymbol{\lambda}) = q(\mathbf{w}|\mathbf{a}; \boldsymbol{\lambda}) \prod_{i=1}^{D} q(a_i)q(\beta)$.

Applying the VB methodology, and taking into account the above factorization, the following posteriors are obtained:

$$q(\mathbf{w}) = \mathcal{N}(\hat{\mathbf{w}}, \mathbf{C_w}), \quad (8)$$

$$q(\beta) = Gamma(\beta; b', c'), \quad (9)$$

$$q(\mathbf{a}) = \prod_{i=1}^{D} Gamma(a_i; b_{a_i}^{'}, c_{a_i}^{'}), \quad (10)$$

The moments of each distribution are calculated by apply-

ing iteratively the following equations until convergence:

$$\mathbf{C}_{\mathbf{w}}^{(k+1)} = (\hat{\beta}^{(k)}\mathbf{\Phi}^T\mathbf{\Phi} + \hat{\mathbf{\Lambda}}^{(k+1)})^{-1}, \quad (11)$$

$$\hat{\mathbf{w}}^{(k+1)} = (\hat{\beta}^{(k)}\mathbf{\Phi}^T\mathbf{\Phi} + \hat{\mathbf{\Lambda}}^{(k+1)})^{-1}\hat{\beta}\mathbf{\Phi}^T\mathbf{y}, \quad (12)$$

$$\frac{1}{b_{a_i}^{(k+1)'}} = \frac{\lambda_i^{(k+1)}}{2}((\hat{w}_i^{(k+1)})^2 + \mathbf{C}_{\mathbf{w}}^{(k+1)}(i,i)) + \frac{1}{b_a}, \quad (13)$$

$$c_{a_i}^{(k+1)'} = \frac{1}{2} + c_a, \quad (14)$$

$$\hat{a}_i^{(k+1)} = b_{a_i}^{(k+1)'}c_{a_i}^{(k+1)'}, \quad (15)$$

$$\frac{1}{b_\beta^{(k+1)'}} = \frac{1}{2}(\mathbf{y} - \mathbf{\Phi}\mathbf{w}^{(k+1)})^T(\mathbf{y} - \mathbf{\Phi}\mathbf{w}^{(k+1)}) + tr(\mathbf{\Phi}^T\mathbf{\Phi}\mathbf{C}_{\mathbf{w}}^{(k+1)}) + \frac{1}{b}, \quad (16)$$

$$c_\beta^{(k+1)'} = \frac{N}{2} + c, \quad (17)$$

$$\hat{\beta}^{(k+1)} = b_\beta^{(k+1)'}c_\beta^{(k+1)'}, \quad (18)$$

In the above equations the matrix $\hat{\mathbf{\Lambda}}^{(k+1)}$ is a diagonal matrix with $\hat{a}_i^{(k)} \cdot \lambda_i^{(k+1)}$ in its main diagonal. The Eqs. (11) - (18) are applied iteratively until convergence. For $\lambda_i^{(k+1)}$ we follow the considerations of [15] and we set them to $\frac{1}{|\hat{w}_i^{(k)}|}$. With respect to other similar approaches [14] we can observe the difference in Eqs. 12 and 13. More specifically, in our approach each parameter a_i is weighted by the corresponding parameter λ_i. Finally, when a new kernelized feature vector (test sample), $\mathbf{z}$, arrives, we can classify it by the following criterion:

$$\hat{y} = \begin{cases} 1 & \text{if } \hat{\mathbf{w}}^T\mathbf{z} \leq 0 \\ 2 & \text{if } \hat{\mathbf{w}}^T\mathbf{z} > 0 \end{cases}$$

III. Results

A. Motor Imagery EEG dataset

In our analysis we have used a well known motor Imagery EEG dataset, the BCI competition IV dataset 2b [19]. This dataset consists of EEG data from 9 subjects. For each subject 5 sessions are provided, whereby the first two sessions contain training data without feedback, and the last three sessions were recorded with feedback. Three bipolar recordings (C3, Cz, and C4) were recorded with a sampling frequency of 250 Hz. They were bandpass-filtered between 0.5 Hz and 100 Hz, and a notch filter at 50 Hz was enabled. The placement of the three bipolar recordings (large or small distances, more anterior or posterior) was slightly different for each subject. The electrode position Fz served as EEG ground. Further information on this dataset can be acquired in [19].

B. Performance Evaluation

As already mentioned, the EEG dataset consists of 5 sessions, the first 2 sessions generated without using feedback and the last 3 sessions with feedback. In our study, we have adopted the same evaluation protocol as in [8]. More specifically, the train/test split consists of the first 3 sessions for training (2 sessions without feedback and the 1 session using feedback), while the remaining 2 sessions (with feedback) are used for testing. The time segment of 0.5 - 2.5s after the onset of the visual cue was used to train the algorithms. During testing, a sliding window of 2 secs, from the visual onset of the corresponding trial until the end of it, was used. A continuous classification output for each sample in the form of class labels was provided by each algorithm. A confusion matrix was built from all trials for each time point. Using these confusion matrices, the time course of the accuracy can be obtained. From these time series the maximum value (maximum accuracy) was selected as the performance measure.

For the extraction of EEG features we have used an approach similar to [8]. More specifically, EEG data from C3, Cz and C4 have been used and a band - pass filter between 8 to 40 Hz has been applied. Following, the EEG recordings were decomposed into multiple frequency pass bands by using a filter bank with bands: 8-12 Hz, 10-14 Hz, 12-16Hz,...,36-40Hz, a total of 15 bands. Then, in each frequency band we apply the Common Spatial Filters algorithm to extract the CSP components [7]. By selecting the pair of CSP components corresponding to the maximum and minimum eigenvalues, we end-up with 30 features for each trial. Finally, these features are fed into the classifier.

To evaluate the performance of our method, we have performed comparisons with state-of-the-art methods like SVM [3], [9], VRVM [16] and the filter bank CSP (FBCSP) approach combined with mutual information-based rough set reduction (MIRSR) [8]. For SVM we have used the LIBSVM library [20] with linear kernel and we have set the regularization parameter C equal to 1. For the initialization of the proposed method and VRVM, we assumed uniformative distributions over all hyperparameters (i.e. $c_{a_k} = 10^{-6}$, $b_{a_k} = 10^6, c_{\beta_k} = 10^{-6}$, $b_{\beta_k} = 10^6$), which is typical in bayesian modeling. Also, a linear kernel was used in both cases.

The classification results are presented in Table I, highlighting the best performing algorithm(s) for each subject. The average performance of each method over the nine subjects is also reported. We can see from this table that our method outperforms the competing methods in most subjects, and overall performs 0.4% better than the second best method.

Following, we check the performance of the examined methods with respect to the size of the training set, as well as its composition on whether it incorporates training samples generated with or without feedback. More specifically, in this experiment we use as training set the trials from session 1 (generated without feedback), while the trials from sessions 4 and 5 (both of which have been generated using feedback) are used for testing. The results of our comparison with SVM and VRVM are reported in Table II, showing how the proposed method adapts better to the unfavorable conditions of a reduced and less informative training set.

IV. Conclusions

In this work we have proposed a new method for the classification of MI EEG data that simultaneously selects the most important EEG features and performs the classification

TABLE I

PERFORMANCE OF THE PROPOSED METHOD, SVM AND VRVM

sub #	Proposed	SVM	VRVM	FBCSP-MIRSR [8]
B01	**75.94**	75.62	75.31	70.00
B02	**61.79**	61.43	61.43	60.35
B03	**61.25**	59.38	59.06	60.95
B04	95.63	95.00	95.94	**97.50**
B05	**93.75**	91.87	92.81	92.80
B06	84.38	84.38	**85.31**	80.65
B07	77.81	**78.75**	78.44	77.50
B08	91.25	91.25	90.94	**92.50**
B09	87.19	**87.81**	86.88	87.20
average	**81.00**	80.61	80.68	79.94

TABLE II

PERFORMANCE OF THE PROPOSED METHOD, SVM AND VRVM USING A SMALL TRAINING DATASET.

sub #	Proposed	SVM	VRVM
B01	**66.87**	65.31	66.25
B02	57.86	**58.57**	57.86
B03	**56.25**	55.00	55.31
B04	**95.94**	93.13	94.06
B05	89.06	82.81	**90.63**
B06	**75.31**	72.81	74.06
B07	**73.44**	68.13	73.12
B08	80.31	76.56	**83.44**
B09	**84.69**	79.37	**84.69**
average	**75.53**	72.41	75.49

by using a linear discrimination rule. To select the most important EEG features, we proposed a new sparse prior for weighting the linear regress models, while the regression weights are estimated using the Variational Bayesian methodology. The undertaken comparisons with SVM, VRVM [16] and FBCSP-MIRSR [8] have shown how the proposed method compares favorably with state-of-the-art performance in MI BCI.

V. ACKNOWLEDGEMENTS

This work is part of project MAMEM that has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 644780.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767 – 791, 2002.

[2] B. Graimann, B. Allison, and G. Pfurtscheller, *BrainComputer Interfaces: A Gentle Introduction*, ch. 1. Springer, 2010.

[3] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, pp. 317–326, Aug 2008.

[4] Q. Xu, H. Zhou, Y. Wang, and J. Huang, "Fuzzy support vector machine for classification of eeg signals using wavelet-based features," *Medical Engineering and Physics*, vol. 31, no. 7, pp. 858 – 865, 2009.

[5] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schlogl, B. Obermaier, and M. Pregenzer, "Current trends in graz brain-computer interface (bci) research," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, pp. 216–219, Jun 2000.

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

[7] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve bci designs: Unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 355–362, Feb 2011.

[8] K. K. Ang, Z. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, no. 39, 2012.

[9] V. P. Oikonomou, K. Georgiadis, G. Liaros, S. Nikolopoulos, and I. Kompatsiaris, "A comparison study on eeg signal processing techniques using motor imagery eeg data," in *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 781–786, June 2017.

[10] J. Meng, X. Sheng, D. Zhang, and X. Zhu, "Improved semisupervised adaptation for a small training dataset in the brain - computer interface," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, pp. 1461–1472, July 2014.

[11] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, "Sparse bayesian classification of eeg for brain-computer interface," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, no. 99, pp. 1–1, 2015.

[12] Y. Zhang, Y. Wang, J. Jin, and X. Wang, "Sparse bayesian learning for obtaining sparsity of eeg frequency bands based feature vectors in motor imagery classification," *International Journal of Neural Systems*, vol. 27, no. 02, p. 1650032, 2017.

[13] V. P. Oikonomou, A. Maronidis, G. Liaros, S. Nikolopoulos, and I. Kompatsiaris, "Sparse bayesian learning for subject independent classification with application to ssvep-bci," in *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 600–604, May 2017.

[14] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Mach. Learn. Research*, vol. 1, pp. 211–244, 2001.

[15] G. Deng, "Iterative learning algorithms for linear gaussian observation models," *IEEE Transactions on Signal Processing*, vol. 52, pp. 2286–2297, Aug 2004.

[16] C. M. Bishop and M. Tipping, "Variational relevance vector machines," in *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, UAI'00, (San Francisco, CA, USA), pp. 46–53, Morgan Kaufmann Publishers Inc., 2000.

[17] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[18] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, October 2007.

[19] R. Leeb, C. Brunnera, G. R. Muller-Putz, A. Schloogl, and G. Pfurtscheller, "Bci competition 2008 - graz data set b." https://lampx.tugraz.at/ bci/database/002-2014/description.pdf, 2008.

[20] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 1–27, 2011.