Comparison of Feature Extraction Methods for EEG BCI Classification

Tomas Uktveris(✉) and Vacius Jusas

Software Engineering Department, Kaunas University of Technology,
Studentu St. 50, Kaunas, Lithuania
{tomas.uktveris,vacius.jusas}@ktu.lt

Abstract. This work analyzes several feature extraction methods used in today's EEG BCI (electro-encephalogram brain computer interface) classification systems. Comparison of multiple EEG energy algorithms is presented for solving a 4-class motor imagery BCI classification problem. Furthermore, multiple feature vector generation techniques are employed into analysis. The effectiveness of CSP (common spatial pattern) filtering method in preprocessing step is shown. Channel difference feature extraction method is presented. It is discussed that key aim in today's EEG signal analysis should be dedicated to finding more accurate techniques for determining better quality features. Initial tests prove that static feature extraction methods are not optimal and adaptive algorithms are required to overcome subject specific EEG signal variations. Further work and new dynamic feature extraction methods are required to solve the problem.

Keywords: Common spatial patterns · Brain-computer interface · Laplace filtering · Feature extraction · Channel difference

1 Introduction

BCI systems try to narrow the gap between human and computer interaction. Direct control of computer applications using only human mind and mental abilities can help solve many rehabilitation, multimedia and gaming challenges. One of the key parts for a BCI system is accurate and fast algorithms, that are capable of analyzing electro-encephalogram (EEG) signal potentials recorded along the human scalp. Such signals contain noise and other unwanted artifacts, which prevent from correctly determining-classifying imagined motoric actions (imagery). Though, many algorithms were developed to overcome such issues, still the problem requires extensive work.

To be able to classify motoric actions with great accuracy, correct and significant features must be extracted from the EEG signal for classifier training. In the process of motor imagery, various regions of the brain are induced differently – signal energy decreases or increases. This is called Event-Related (De-) Synchronization – ERD or ERS. Since ERD/ERS describes transient changes in the brain signal oscillatory activity, a correct pattern of such information allows classification of motor imagery tasks. However, pattern extraction is error-prone due to the nature and highly non-deterministic brain activity even for the same test subject. This work further

G. Dregvaite and R. Damasevicius (Eds.): ICIST 2015, CCIS 538, pp. 1–12, 2015.
DOI: 10.1007/978-3-319-24770-0_8

discusses the use of various signal processing algorithms for EEG feature extraction. In Sect. 2 various feature extraction techniques for EEG signal analysis are presented along with a new Channel difference feature extraction method. Section 3 gives an overview of the classification methods applied in practice, while Sect. 4 provides more detail about the experiment procedure and data selected to assess the algorithms. Finally, Sect. 5 presents results and findings, concluding remarks are given in the last Sect. 6.

2 Techniques for Feature Extraction

Over the years many multi-class BCI solutions were proposed that use different feature extraction methods. A feature extraction algorithm is one of the most critical parts in EEG classification task or any BCI system processing pipeline. Since BCI system accuracy directly depends on the quality of extracted feature vectors, care must be taken to ensure quality. The pipeline can be viewed as an EEG signal processing filters chain as shown in Fig. 1.

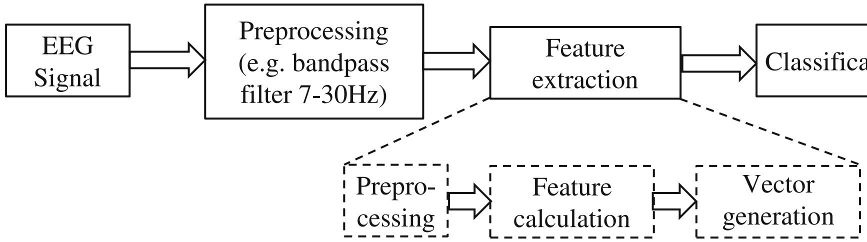

Fig. 1. EEG signal processing filter pipeline

After initial required signal preprocessing (that must be done for all signal channels) feature extraction stage comes into play. For each extraction algorithm, such stage can be decomposed into three distinct processing blocks that influence total performance of the algorithm:

1. *Preprocessing* [optional] – additional signal filtering (e.g. into different frequency bands).
2. *Feature calculation* – process EEG channel data (e.g. channel energy calculation).
3. *Vector generation* – compute elements of the final vector (e.g. mean of channel energy)

By controlling and changing the implementation in blocks many algorithm variations can be acquired for evaluation. This schema is handful in order to analyze subtle algorithm discrepancies also. Five EEG signal feature extraction methods were implemented and analyzed using this method in this work – Band Power features (BP), Time Domain Parameters (TDP), Teager-Kaiser Energy Operator, Signal power and Channel difference. Additional EEG preprocessing step was done using Common Spatial Patterns (CSP) filtering. Each of the algorithms will be briefly detailed further.

2.1 Signal Power Features

One of the simplest methods for signal energy calculation. The power of a signal is the sum of squares of its time-domain samples divided by the signal length (1). The power is computed for every EEG channel and the result is used further as a feature vector.

$$P = \frac{1}{N}\sum_{k=1}^{N} x^2[k] \tag{1}$$

where x – the discrete single EEG channel signal values, N – number of EEG signal samples taken.

2.2 Band Power Features

Algorithm calculates multiple band power by band-pass filtering [1] the signal. To apply the algorithm, the signal frequency range must be divided into multiple regions. First, initial EEG signal is filtered using a band-pass filter designed for each frequency band, e.g. 4-th order Butterworth finite impulse response (FIR) filter. The resulting signal is squared (2) to obtain its power.

$$p[t] = x^2[t] \tag{2}$$

here x –the filtered single band EEG signal values, p – resulting bandpower values.

A w-sized smoothing window operation is performed (3) to smooth-out (average) the signal. Finally a logarithm of the processed signal is taken. Computed result is used for feature vector generation. Such method is already implemented in MATLAB signal processing library.

$$\bar{p}[n] = ln\left(\frac{1}{w}\sum_{k=0}^{w} p[n-k]\right) \tag{3}$$

here $\bar{p}$ – resulting smoothed bandpower values, w – the smoothing window size.

Three different frequency bands were used in the work: 8–14 Hz, 19–24 Hz and 24–30 Hz which correspond to *Mu*, *Alfa* and *Beta* brain waves. A similar approach was used in a closed loop system for navigating in a virtual environment via ERD-BCI [2]. Complex band power features were selected from three major frequency bands of cortical oscillations: μ (8–12 Hz), sensorimotor rhythm (12–15 Hz) and β (15–30 Hz).

2.3 Time Domain Parameters

Similar to the BP algorithm, time domain parameters computes time-varying power of the first k derivatives of the signal. Obtained derivative values (4) are smoothed using exponential moving average and a logarithm is taken as given by (5). The resulting signal is used in feature vector generation.

$$p_i(t) = \frac{d^i x(t)}{dt^i}, \quad i = 0, 1, \ldots, k \tag{4}$$

$$\bar{p}_i[n] = \ln(u \cdot p_i[n] - (1 - u) \cdot p_i[n - 1]) \tag{5}$$

here x – initial EEG signal, p – signal derivative values, u – moving average parameter, $u \in [0; 1]$, $\bar{p}$ – smoothed signal derivatives.

2.4 Teager-Kaiser Energy Operator (TKEO)

A non-linear algorithm for a more accurate signal energy calculation was presented by Teager and further analyzed by Kaiser [3]. Advantage of TKEO is the ability to discover high-frequency low-amplitude components and take into account frequency component and signal amplitude of the signal [4]. The algorithm for a continuous signal can be written as shown in the (6) equation, while an approximation (7) exists for discrete signals.

$$\Psi[x(t)] = \left(\frac{\partial x}{\partial t}\right)^2 - x(t) \cdot \frac{\partial^2 x}{\partial t^2} \tag{6}$$

$$\Psi[x[n]] = x^2[n] - x[n - 1]x[n + 1] \tag{7}$$

here x – the EEG signal, $\Psi[x[n]]$ – energy values computed for discrete signal.

The algorithm was applied for each of EEG channels then final feature vectors were generated.

2.5 Channel Difference Method

Channel difference algorithm is a new method presented in this paper for extracting EEG signal features. The method is an extension to the Band power feature algorithm with an extra signal filtering step. The algorithm works by computing filtered features only for EEG channels that have at least four neighboring electrodes around each of them. In this work four EEG channels, that match this criteria, i.e. electrodes C3, Cz, C4 and Pz (in international system 10–20), were selected for this as shown in Fig. 2. The locations were chosen to be symmetric and to cover both hemispheres in order to be able to capture all energy changes induced by motor-imagery for different sides of the body [5]. It was already shown that Laplace signal filtering is effective at enhancing EEG spatial resolution [6] and discerning EEG signals from the background [7] noise. Each of the selected channels was filtered by using a Laplace filter (8) in a single channel radius neighborhood with weight kernel as given in (9).

$$\mathrm{S} = \sum_{i=1}^{3} \sum_{j=1}^{3} \mathrm{K}_{ij} E_{ij} \tag{8}$$

$$K = \begin{bmatrix} -0.5 & -1 & -0.5 \\ -1 & 6 & -1 \\ -0.5 & -1 & -0.5 \end{bmatrix} \tag{9}$$

where E_{ij} – the neighboring EEG channel signal, K_{ij} – the corresponding weight in the kernel matrix. For example, by taking Cz as the selected channel, the K_{11} element would denote a weight of −0.5 for the 3rd EEG channel and K_{22} would denote weight 6 for the Cz channel and so on. For non-existing channels zero weight was used. Larger kernel sizes were not analyzed as they would require more EEG channels.

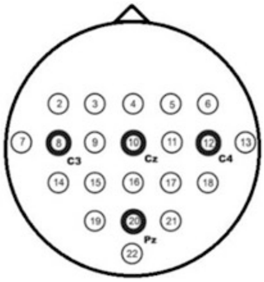

Fig. 2. EEG channels from the 10–20 system used in calculation

After filtering, the band power of frequency ranges: 8–14 Hz, 14–19 Hz, 19–24 Hz, 24–30 Hz was computed for each of the four signals. The 16 resulting energy bands were used for feature vector generation.

2.6 CSP Preprocessing

Common spatial patterns (CSP) is a preprocessing technique (filter) for separating a multivariate signal into subcomponents that have maximum differences in variance [8]. Separation allows for easier signal classification. In general, the filter can be described (10) as a spatial coefficient matrix W:

$$S = W^T E \tag{10}$$

where S – the filtered signal matrix, E – original EEG signal vector. Columns of W denote spatial filters, while inverse of, i.e. W^{-1}, are spatial patterns of EEG signal. Criterion of CSP for a two C_1, C_2 class problem is given by:

$$maximize\ tr\left(W^T \sum\nolimits_1 W\right) \tag{11}$$

$$subject\ to\ W^T \left(\sum\nolimits_1 + \sum\nolimits_2\right) W = I \tag{12}$$

here

$$\sum\nolimits_1 = exp_{E_n \in C_1}\left(\frac{E_n E_n^T}{tr\left(E_n E_n^T\right)}\right) \tag{13}$$

$$\sum\nolimits_2 = exp_{E_n \in C_2}\left(\frac{E_n E_n^T}{tr\left(E_n E_n^T\right)}\right) \tag{14}$$

The solution can be acquired by solving generalized eigenvalue problem of by decomposing the problem into multiple standard eigenvalue sub-problems. Multiclass solutions are combined of multiple spatial filters. For more information see [9].

3 Methods for Classification

From an extensive list of known EEG signal classifiers most commonly used ones were selected for analysis – Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and k-Nearest Neighbors (kNN). A brief overview of the classification algorithms will be given further.

3.1 Linear and Quadratic Discriminant Analysis

A classifier that employs the Bayes' theorem for classification. Discriminant analysis estimates the parameters of a Gaussian distribution for each class and the trained classifier finds the class with the smallest misclassification cost. The posterior probability that a point x belongs to class C is the product of the prior probability and the multivariate normal density. The density function of the multivariate normal with mean μ_c and covariance Σ_c at a point x is given by:

$$P(x|C) = \frac{1}{(2\pi|\Sigma_c|)^{1/2}} exp\left(-\frac{1}{2}(x-\mu_c)^T \Sigma_c^{-1}(x-\mu_c)\right), \tag{15}$$

where $|\Sigma_c|$ is the determinant of Σ_c and Σ_c^{-1} is the inverse matrix.

Let $P(C)$ represent the prior probability of class C. Then the posterior probability that an observation x is of class C is given by:

$$P(C|x) = \frac{P(x|C)P(C)}{P(x)}, \tag{16}$$

where $P(x)$ is a constant equal to the sum over C of $P(x|C)P(C)$.

The Linear discriminant (or Fisher discriminant) analysis model is named for its inventor, Fisher [10]. Linear method (LDA) has the same covariance matrix for each class and only the means vary. For quadratic discriminant analysis (QDA), both means and covariances of each class vary.

3.2 Support Vector Machine (SVM)

The Support Vector Machines were introduced by Boser et al. [11] in 1992. SVM is a two-class algorithm that classifies data by finding the best hyperplane (17) separating all one class points from all of the other class (with the largest margin).

$$w^T x + b = 0 \tag{17}$$

The problem is of dual quadratic programming nature and can be reduced to Lagrangian optimization problem. A scheme of One vs All or One vs One is used if more than two classes are needed (Fig. 3).

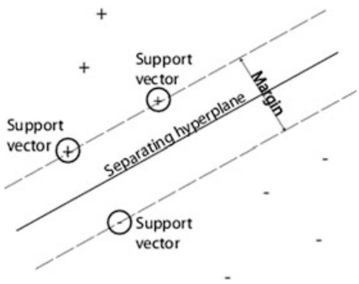

Fig. 3. Finding separating SVM hyperplane between features

3.3 k-Nearest Neighbors

kNN is one of the simplest algorithms for classification. A feature vector is classified by a majority vote of its neighbors. The object class is assigned to the most common one found among *k* nearest neighbors (e.g. to class "square" as given in Fig. 4).

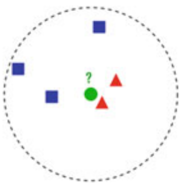

Fig. 4. Classifying object with a kNN classifier ($k = 5$)

4 Data Selection and Experiment

4.1 Evaluation Procedure

Feature extraction methods mentioned in Sect. 2 were implemented and compared in the experiment. All experiments were completed using MATLAB numerical computation environment, BioSig library for biomedical signal processing and libSVM for multiclass SVM classification tasks. The ideas of using different EEG signal energy processing methods and CSP filtering (with initial code implementation) were acquired from an earlier work of Piotr Szachewicz [12]. Classifiers were trained and validated using tenfold cross-validation. Default parameters were used for LDA and QDA classifiers as provided by MATLAB package. Grid search method was used for SVM RBF (radial basis function) gamma and cost parameter optimization [13], used values were: $C = 10$, $\gamma = 0.25$.

4.2 BCI IV 2a Dataset

A BCI signal database [14] from the BCI IV competition held in 2008 was selected for classifier training and testing. The analyzed experiment data was used from the freely available "2a" dataset for a 4-class motor imagery problem. The data consisted of 22 channels of 250 Hz sample rate recorded EEG signal for 9 test subjects and 288 motor imagery trials per subject. Using a cue-paced (synchronous) mode of operation, test subjects were asked to imagine movement of one out of four different motions (left hand, right hand, feet, tongue) for 3 s. Each of the trials (Fig. 5) in the dataset started with an audible signal (beep), followed by visual information (cue) to perform one of the mental tasks and a short break after the mental task.

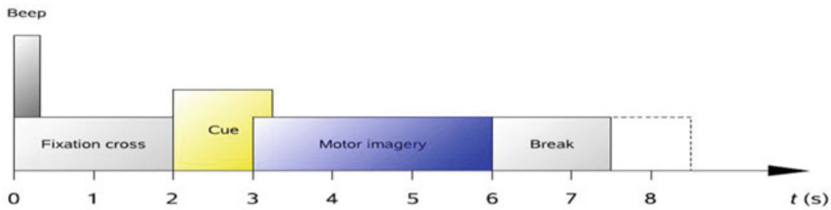

Fig. 5. Single trial timing scheme

4.3 Accuracy Calculation

The accuracy of the BCI data classification results was computed by calculating the Cohen's kappa [15] coefficient κ as given by equation:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{18}$$

here p_0 – the classification accuracy, p_e – hypothetical accuracy of a random classifier for the data ($p_e = 0.25$ for four classes).

5 Results

All experiment results are given in Tables 1 and 2. As can be seen from the tables, by using simple EEG features the mean kappa for the LDA method almost everywhere scores highest accuracy (and average). This could be an indication of good linearization

Table 1. Average classification results (kappa values) using CSP filtering

Classifier	Ch.diff	Teager	TDP	Mean power	BP	C-Avrg
kNN	N/A	0,3890	0,4327	0,4249	0,3393	0,3965
LDA	N/A	**0,4437**	**0,4950**	**0,4638**	**0,4154**	**0,4545**
QDA	N/A	0,3868	0,4344	0,4259	0,3515	0,3997
SVM	N/A	0,3645	0,3835	0,4605	0,4075	0,4040
F-avrg	N/A	0,3960	0,4364	**0,4438**	0,3784	

Table 2. Average classification results (kappa values) without CSP filtering

Classifier	Ch.diff	Teager	TDP	Mean power	BP	C-Avrg
kNN	0,3678	0,1538	0,2045	0,1867	0,1225	0,2071
LDA	0,4346	**0,4147**	**0,4134**	**0,4400**	**0,2487**	**0,3903**
QDA	0,3179	0,2864	0,2790	0,3185	0,1480	0,2700
SVM	**0,4372**	0,1553	0,1791	0,2950	0,1877	0,2509
F-avrg	**0,3894**	0,2525	0,2690	0,3101	0,1767	

or good linear separation of the EEG features. Still, a combination of CSP filtering and SVM classification beats the LDA for single feature type, but the difference is negligible. A maximum average kappa of 0.495 was reached in tests (i.e. accuracy of 62 %), which is still far from 90 % accuracy achieved in other [16] work that uses more advanced techniques. Since Channel difference method does not support CSP filtering, so N/A values are presented.

Average feature performance (F-avrg) indicates that Mean power and TDP features are the best feature methods when using CSP filtering. However, Channel difference method achieves best result when CSP filtering is not being done. Band power algorithm was the worst performer in tests. A graphical view of the same data is given further in Fig. 6 where CSP filtering normalization influence can be seen better.

It should be noted, that Channel difference method performance follows along Band Power performance when CSP is used due to obvious reason – the method needs to extract different signal frequency bands by using the BP algorithm. This can be seen in Fig. 7 further. Since Channel difference method cannot use CSP filtering, one result bar is not shown.

Quite important is the view showing method accuracy per subject as in Fig. 8. Some subjects are resistant to existing EEG methods, so subject-specific (adaptive) techniques are required in order to achieve higher classification accuracy. The positive normalizing effect can be seen for the CSP case - giving greater average accuracy.

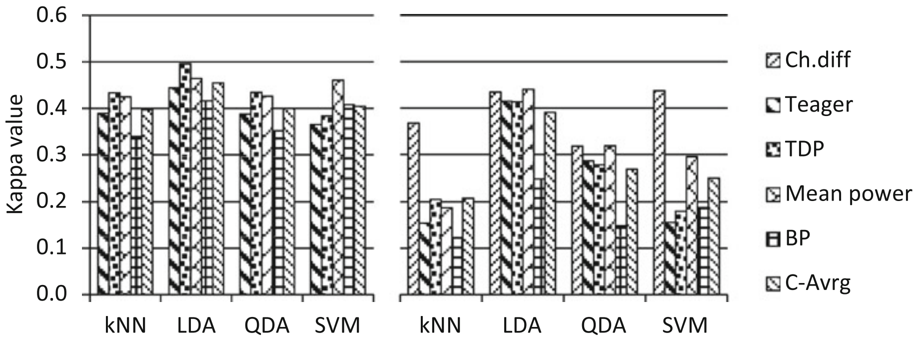

Fig. 6. Average feature results for classifiers (left – with CSP, right – no CSP)

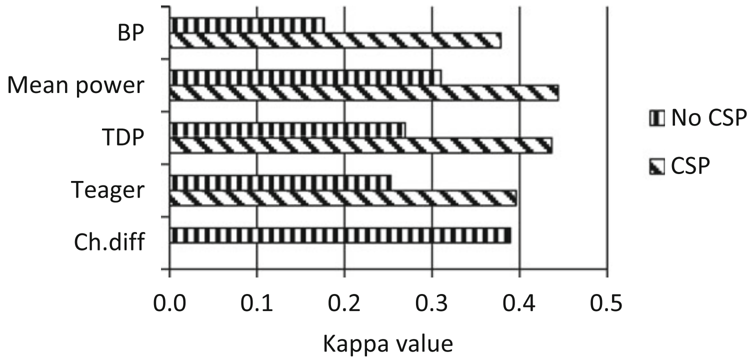

Fig. 7. Average accuracy per feature type

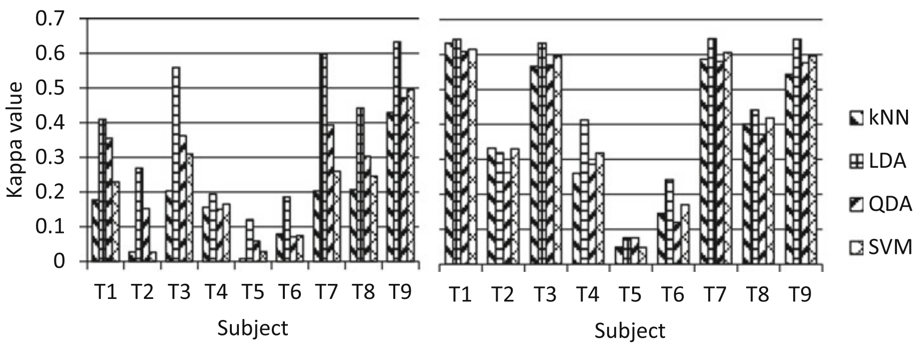

Fig. 8. Average accuracy per subject (left – no CSP, right – with CSP)

6 Conclusion

This work analyzed multiple signal energy feature extraction methods and their usage for 4-class motor imagery BCI classification problem. Tested algorithms showed that simple EEG signal energy feature extraction methods such as Mean power or TDP are one of the best when doing EEG signal classification with CSP filtering. A positive influence on accuracy and test results were visible when the CSP filter was applied.

Presented results also confirmed that the Linear Discriminant analysis (LDA) algorithm can be successfully used for EEG signal classification. Best classification performance was demonstrated by LDA among tested classifiers. Such results also provide insight, that EEG features can be linearly separable.

Proposed Channel difference algorithm for signal feature extraction was able to achieve best average classification result among other tested feature extraction algorithms. Ability to reach classification results close to CSP by using simpler filtering is one of the advantages of the method. However, many optimization possibilities exist for future work including development of better feature extraction algorithms and adaptation to subject specific EEG information.

All experiment source code, test data and detailed results can be acquired from repository: https://github.com/tomazas/icist2015.

References

1. Brodu, N., et al.: Comparative study of band-power extraction techniques for motor imagery classification. In: IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), pp. 1–6 (2011)
2. Pfurtscheller, G., et al.: Graz-Brain-Computer Interface: State of Research, pp. 65–84. MIT Press, Cambridge (2007)
3. Kaiser, JF.: On a simple algorithm to calculate the energy of a signal. In: IEEE International Conference on Acoustic Speech Signal Process, Albuquerque, NM (1990)
4. Martišius, I., et al.: Using higher order nonlinear operators for SVM classification of EEG data. Elektronika ir Elektrotechnika **119**(3), 99–102 (2012)
5. Dolezal, J., Cerny, V., Stastny, J.: Online motor-imagery based BCI. In: International Conference on Applied Electronics (AE), pp. 65–68, 5–7 (2012)
6. Tandonnet, C., Burle, B., Hasbroucq, T., Vidal, F.: Spatial enhancement of EEG traces by surface Laplacian estimation: comparison between local and global methods. Clin. Neurophysiol. **116**, 18–24 (2005)
7. Qin, L., He, B.: A wavelet-based time-frequency analysis approach for classification of motor imagery for brain-computer interface applications. J. Neural Eng. **2**, 65–72 (2005)
8. Müller-Gerking, J., et al.: Designing optimal spatial filters for single-trial EEG classification in a movement task. Clin. Neurophysiol. **110**(5), 787–798 (1999)
9. Thang, L.Q., Temiyasathit, C.: Increase performance of four-class classification for motor-imagery based brain-computer interface. In: 2014 International Conference on Computer, Information and Telecommunication Systems (CITS), pp. 1–5, 7–9 (2014)
10. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Ann. Eugenics **7**, 179–188 (1936)

11. Ben-Hur, A., Weston, J.: A user's guide to support vector machines. In: Carugo, O., Eisenhaber, F. (eds.) Data Mining Techniques for the Life Sciences. Methods in Molecular Biology, vol. 609, pp. 223–239. Humana Press, New York (2010)
12. Szachewicz, P.: Classification of Motor Imagery for Brain-Computer Interfaces. Master's thesis, Poznan University of Technology, Poznan (2013)
13. Hsu, C.-W., et al.: A Practical Guide to Support Vector Classification. National Taiwan University, Taiwan (2010)
14. Brunner, C., et al.: BCI Competition 2008 – Graz data set A (2008). https://www.bbci.de/competition/iv/desc_2a.pdf
15. Schlogl, A., et al.: Evaluation criteria in BCI research. In: Dornhege, G., del Millan, J.R., Hinterberger, T., McFarland, D.J., Muller, K.-R. (eds.) Toward Brain-Computer Interfacing, pp. 327–342. MIT Press, Cambridge (2007)
16. Ang, K.K., et al.: Filter bank common spatial pattern (FBCSP) in brain-computer interface. In: Neural Networks, IJCNN 2008. IEEE World Congress on Computational Intelligence, pp. 2390–2397, 1–8 June 2008