

A Survey on Machine Learning Algorithms in Little-Labeled Data for Motor Imagery-Based Brain-Computer Interfaces

Yuxi Jia¹, Feng Li^{1,2}, Fei Wang^{1,2,*} and Yan Gui^{1,2,3}

Abstract: The Brain-Computer Interfaces (BCIs) had been proposed and used in therapeutics for decades. However, the need of time-consuming calibration phase and the lack of robustness, which are caused by little-labeled data, are restricting the advance and application of BCI, especially for the BCI based on motor imagery (MI). In this paper, we reviewed the recent development in the machine learning algorithm used in the MI-based BCI, which may provide potential solutions for addressing the issue. We classified these algorithms into two categories, namely, enhancing the representation and expanding the training set. Specifically, these methods of enhancing the representation of features collected from few EEG trials are based on extracting features of multiple bands, regularization, and so on. The methods of expanding the training dataset include approaches of transfer learning (session to session transfer, subject to subject transfer) and generating artificial EEG data. The result of these techniques showed the resolution of the challenges to some extent. As a developing research area, the study of BCI algorithms in little-labeled data is increasingly requiring the advancement of human brain physiological structure research and more transfer learning algorithms research.

Keywords: Brain-Computer interface, electroencephalography (EEG), machine learning.

1 Introduction

The Brain-Computer Interface (BCI) based on electroencephalography (EEG) signals establishes a direct communication channel for healthy or disabled users from the brain to a technical device [Krauledat, Tangermann, Blankertz et al. (2008)]. Signal patterns for EEG-based BCI include: event-related potential (ERP) [Farwell and Donchin (1988)], steady-state visual evoked potential (SSVEP) [Muller-Putz, Scherer, Neuper et al. (2006)], event-related synchronization/desynchronization (ERS/ERD) [Pfurtscheller and Da Silva (1999)]. For instance, the BCI can be used to control wheelchair in the case of lower limb paralysis [Galan, Nuttin, Lew et al. (2008); Millan, Galan, Vanhooydonck et al. (2009)], restore stroke rehabilitation [Dimyan and Cohen (2011)], and support the healthy users gaming [Leeb, Lancelle, Kaiser et al. (2013); Lecuyer, Lotte, Reilly et al. (2008)].

¹ School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, 410114, China.

² Hunan Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation, Changsha University of Science and Technology, Changsha, 410114, China.

³ School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore, 639798, Singapore.

* Corresponding Author: Fei Wang. Email: cce_wf@csust.edu.cn.

The motor imagery is a typical BCI based on ERS/ERD. ERS/ERD is a spontaneous rhythm signal that does not require external stimulation, and therefore makes motor imagery more advantageous than other signal patterns for the disabled user. However, supervised learning algorithms are commonly used in the BCI require a large amount of labeled data, as the EEG signals are commonly non-stationary and high-dimensional. Furthermore, the EEG signals, especially ERS/ERD, are susceptible to subjects own physical condition. Therefore, it requires a long calibration time for subjects learning to adjust the amplitude of rhythm signal. But there is rarely numerous labeled data provided in practice [Fazli,Dahne, Samek et al. (2015)], so we need machine learning algorithms in little-labeled data to improve the performance of the BCI system.

In this paper, we reviewed the approaches which focus on addressing these problems. We explored the techniques of processing little-labeled dataset using machine learning tools in the MI-based BCI. In order to provide a broad overview, briefly, we divided these methods into two categories, enhancing the representation of the feature and expanding the training dataset. As shown in Tab. 1, the information and results of some algorithms were presented and classified into different types, which were reported by the effects of dealing the open dataset, such as BCI Competition.

Table 1: A Summary of some approaches reported for the open dataset

Type	Approaches	Brief Description	Dataset	Mean accuracy
Enhancing the representation	Bayesian spatio-spectral filter [Suk and Lee (2013)]	Bayesian framework for discriminative feature extraction	BCI Competition IV Dataset IIa	87.03%
			BCI Competition III dataset IVa	75.46%
	Stationary subspace analysis [Samek, Vidaurre, Muller et al. (2012)]	Stationary subspace analysis for feature selection	BCI Competition IV Dataset IIa	78.67%
			BCI Competition III dataset IVa	81.94%
	Riemannian geometry [Barachant,Bonnet, Congedo et al. (2012)]	Classification using Riemannian geometry	BCI Competition IV Dataset IIa	70.20%
			BCI Competition III dataset IVa	90.00%
	Subband [Novi, Guan, Dat et al. (2007)]	Decomposing the signal into sub-bands, selecting the appropriate bands	BCI Competition III dataset IVa	89.00%
			BCI Competition III dataset IVa	89.00%
Filter bank [Ang, Chin, Zhang et al. (2008)]	Decomposing the signal into multiple bands and select the discriminative frequency bands to classify	BCI Competition III dataset IVa	89.00%	
		BCI Competition III dataset IVa	89.00%	
Filterbank optimization [Haihong,Yang, Keng et al. (2011)]	Using information theoretic approach to learn spatio-spectral filters	BCI Competition IV Dataset I	93.93%	
		BCI Competition IV Dataset I	93.93%	

	Covariance shift [Li, Kambara, Koike et al. (2010)]	Classification using importance-weighted LDA method under covariate shift	BCI Competition III dataset IVc	96.50%
session to session	Semi-supervised [Meng, Sheng, Zhang et al. (2014)]	Using a semi-supervised approach to classify the unlabeled data to expand the training set	BCI Competition IV Dataset IIa	72.20%
			BCI Competition III dataset IVa	85.70%
	Riemannian geometry [Zanini, Congedo, Jutten et al. (2017)]	Exploiting the techniques based on the Riemannian geometry of the manifold of symmetric positive definite matrices	BCI Competition IV Dataset IIa	80.00%
	Bayesian CSP [Kang and Choi (2014)]	A nonparametric Bayesian model for a multi-subject extension of CSP	BCI Competition IV Dataset IIa	70.92%
	Riemannian kernel SVM [Barachant, Bonnet, Congedo et al. (2013)]	Classification using a Riemannian-based kernel SVM	BCI Competition III dataset IVa	73.17%
			BCI Competition IV Dataset IIa	86.00%
subject to subject	Regularizing CSP [Lotte and Guan (2011)]	A unifying theoretical framework to design regularized CSP	BCI Competition IV Dataset IIa	78.78%
			BCI Competition III dataset IVa	78.83%
	Multi-task learning CSP [Devlaminck, Wyns, Grosse-Wentrup et al. (2011)]	Learning a spatial filter for a new subject from its own data and other subjects	BCI Competition III dataset IVa	87.96%
	Composite data [Kang, Nam and Choi (2009)]	Composites covariance matrices from other subjects	BCI Competition III dataset IVa	79.00%
	Riemannian geometry [Zanini, Congedo, Jutten et al. (2017)]	Exploiting the techniques based on the Riemannian geometry of the manifold of symmetric positive definite matrices	BCI Competition IV Dataset IIa	85.70%
			BCI Competition IV Dataset IIa	66.70%

2 Methods of enhancing the representation of the little-labeled dataset

A typical algorithm for feature extracting in BCI designs is the Common Spatial Pattern (CSP), which has been proven to be one of the most efficient algorithms for motor imagery tasks [Ramoser, Muller-Gerking and Pfurtscheller (2000)]. CSP aims at learning a spatial

filter by maximizing the variance of one class when minimizing the variance of the other class [Ramoser, Muller-Gerking and Pfurtscheller (2000); Ang, Chin, Zhang et al. (2008)]. However, the CSP filter is subject-specific, which is dramatically sensitive to noise, notably, in the case of the subject with much fewer training trials [Reuderink and Poel (2008)]. Currently, in order to solve the problem of little-labeled data, many machine learning algorithms were improved on the basis of CSP.

Ang et al. [Ang, Chin, Zhang et al. (2008)] proposed a feature selecting approach based on filter-bank which processed the feature in multiple frequency bands to enable the feature more discriminative. It comprises four stages of processing: multiple band-pass filters, CSP algorithm filters the data in spatial domain, feature selection of the CSP feature, and classification based on the selected feature. The first stage cuts the EEG data into multiple bands. The second stage uses the CSP filter to extract CSP feature of each bands. The feature selection is defined as: given a set of F with d features, find the subset $S \subset F$ with k features in the condition of class Ω that maximizes Mutual Information $I(S; \Omega)$. The Mutual Information between the two variables is

$$I(X; \Omega) = H(\Omega) - H(\Omega | X) \quad (1)$$

where the entropy of the variable X and the conditional entropy are

$$H(\Omega | X) = - \int_X \sum_{\omega=1}^{\omega} p(\omega | x) \log_2 p(\omega | x) dx \quad (2)$$

This approach addresses the problem of selecting an appropriate operational frequency band for extracting discriminating CSP features. This algorithm can achieve an average accuracy of 89% (see Tab. 1).

As the extension of filter-bank, [Suk and Lee (2013)] proposed a Bayesian framework with the spatio-spectral filter optimized by means of the probabilistic and information-theoretic approaches for extracting the discriminative feature. It defines the frequency band as a random variable. The problem of optimizing the spatio-spectral filter is formulated as the estimation of the posterior probability density function (pdf) by the Bayes rule. After the features of the proposed pdf estimation, computing the weighted label decision rule by linearly combining the output from multiple classifiers. The weight of each particle can compute as follows:

$$\pi_k = \frac{p(F_k, \Omega | b_k)}{\sum_{i=1}^K p(F_i, \Omega | b_k)} \quad (3)$$

where F_k denotes a feature vector set extracted from the spectrally and spatially filtered, b_k is a particle representing a single frequency band. The classification accuracy of this algorithm nearly improved 15% for the subject who performed ordinary. But the performance is decreased for the subjects lacking of BCI efficiency.

Some strategies are similar as filter-bank, such as Sub-band CSP(SBCSP) [Novi, Guan, Dat et al. (2007)] and Optimum Spatio-Spectral Filtering Network (OSSFN) [Haihong, Yang, Keng et al. (2011)], decomposed the EEG signals into multiple bands. Besides, stationary subspace analysis (SSA) [Von Bunau, Meinecke, Kiraly et al. (2009)] and Riemannian geometry [Barachant, Bonnet, Congedo et al. (2012)] had been used for this task. Samek et al. [Samek, Vidaurre, Muller et al. (2012)] proposed to combine CSP and SSA and achieved a better classification result. It used SSA to find the steady-state part of the EEG data before using CSP to calculate the spatial filter, and the results indicate that stationary CSP is better

than that of other methods for the subjects lacking of BCI efficiency.

3 Methods of expanding the training dataset

Supervised learning is aimed at inferring an unknown model of input-output, by which unlabeled samples can be classified. If little training dataset is available, or these data cannot completely reflect the feature distribution, the covariance matrices may be poorly estimated. This will lead to the spatial filter or classifier under-fitting. One of the most direct way to solve the problem of insufficient data is to supplement more relevant training trials from other sessions or subjects. It can be divided into the following types of approaches: session to session transfer, subject to subject transfer and generating artificial data.

3.1 Session to session transfer

Since CSP filters are subject-specific, in order to improve the performance of filters, the similarity sessions data is applied for a given subject to extend the training set. Krauledat et al. [Krauledat, Tangermann, Blankertz et al. (2008)] aimed at mitigating the estimation bias of the previous model through reusing previous sessions data of the same user in a clustering approach. It suggested that the similar filter should be found across all sessions for a given subject. To find the more densely sampled regions in the data space consisting of CSP filters, it calculates the angle between the column vectors w_i of a CSP filters matrix W :

$$m(w_1, w_2) = \arccos\left(\frac{w_1^T w_2}{|w_1| |w_2|}\right) \quad (4)$$

compute the average distance γ -index of w to its neighbors:

$$\gamma(w) = \frac{1}{k} \sum_{i=1}^k m(w, nn_i(w)) \quad (5)$$

where nn_i is the k nearest neighbors of w , the lowest γ -index denotes this filter inside a region which contain other filters, and the corresponding filter should be chosen as cluster prototype. But this approach can not applicable to a completely new subject.

Gradually, covariate shift has been introduced in BCI designs. Covariate shift is defined as the situation where the training input points and testing input points follow different distributions, under the hypothesis that the conditional distribution of these points is invariant [Sugiyama, Krauledat and MAZiller (2007)]. Li et al. [Li, Kambara, Koike et al. (2010)] had combined the labeled training data with the unlabeled test data together, assuming that the marginal distribution of the sessions changes but the decision rules are unchanged. It proposed re-weighting the data of previous sessions to fine-tune the predictive function $f_\theta(x)$ in order to correct the covariate shift. The function is described as follows:

$$\begin{aligned} & \iint Loss(x, y, f_\theta(x)) P_{te}(x) P(y | x) dx dy \\ &= \iint Loss(x, y, f_\theta(x)) w(x) P_{tr}(x) P(y | x) dx dy \end{aligned} \quad (6)$$

where x denotes training samples, $w(x) = \frac{P_{te}(x)}{P_{tr}(x)}$, $Loss(x, y, f_\theta(x))$ is a loss function.

However, it produces a large-variance estimator. Li et al. [Li, Kambara, Koike et al. (2010)] introduced a technique namely Bagging to overcome the weakness. The accuracy of the algorithm summarily improved 10% to traditional LDA algorithm.

Besides, Semi-supervised Learning is a way to exploit unlabeled data effectively [Tu, Lin, Wang et al. (2018)]. Meng et al. [Meng, Sheng, Zhang et al. (2014)] used a semi-supervised approach to solve the problem of lacking data. First, it selected less data for pre-training to get a weak spatial filter and filtered the data of test dataset to get a trial with the highest confidence. Then, it added the trial to the training dataset and iterated the processing until the number of trials met 20. But this algorithm was lacked robustness and caused the performance worse when the training set contained the outliers.

3.2 Subject to subject transfer

The early researches were to supplement the regularized data of other subjects into the training dataset [Kang, Nam and Choi (2009); Lotte and Guan (2011)]. Kang et al. [Kang, Nam and Choi (2009)] modified CSP for subject-to-subject transfer, which proposed to composite covariance matrices from other subjects. In this approach, the weights were defined according to the Kullback-Leibler (KL) divergence between subjects:

$$\tilde{C}_c^k = (1 - \lambda)C_c^k + \lambda \sum_{j \neq k} a_{jk} C_k^j \quad (7)$$

$$a_{jk} = \frac{1}{Z^k} \cdot \frac{1}{KL[p^j||p^k]} \quad (8)$$

where a_{jk} denotes the importance of subject j to subject k , $Z^k = \sum_{j \neq k} \frac{1}{KL[p^j||p^k]}$ is normalization constant, and the KL-divergence is defined as follows:

$$KL[p^j||p^k] = \frac{1}{2} \left\{ \log \left(\frac{\det C^k}{\det C^j} \right) + \text{tr} \left[(C^k)^{-1} C^j \right] - D \right\} \quad (9)$$

The modified CSP filter improved the classification performance which was caused by the lack of training trials. The accuracy of subjects aw and ay from the dataset of BCI competition III IVa had increased nearly 15%. However, the performance of this algorithm was unstable due to the potentially large variability between subjects. Thus, Barachant et al. [Barachant, Bonnet, Congedo et al. (2013)] proposed to do not use the data from all available subjects but from selected subjects only. It sequentially selected the subject from the subset, training the BCI on the data from the selected subject and tested it on the training data of the target subject. Based on the classification effect on the target subjects, it is decided that the data of the subjects should be deleted or added from the current dataset. This method guided a way to select the source data. The results showed that the best regularizing CSP algorithm can outperform CSP by almost 10% in median classification accuracy.

In a word, these methods were aimed at suppressing the influence of noise on CSP. The result shows that the best regularized CSP can outperform CSP by almost 10% in median classification accuracy. On the basis of using regularization, Samek et al. [Samek, Kawanabe and Muller (2014)] designed a divergence-based framework to compare the various divergence calculation methods and automatically selected the regularization coefficient by utilizing historical data.

In order to find the relationship of the dataset from different subjects, Fazli et al. [Fazli, Popescu, Danóczy et al. (2009); Morioka, Kanemura, Hirayama et al. (2015); Kang and Choi (2011)] learned an invariant sparse in multiple datasets to predict the different subjects. Sparse representation is a typical signal processing method to represent the main information of a signal using non-zero coefficients as few as possible [Wang, Shen, Li et

al. (2018)]. This kind of approach guided the algorithms for better performance, thus enabling the algorithm to generalize with little-labeled data. Simultaneously, multi-task learning made achievements in this regard. In the field of BCI, each task generally learns a classification model for a specific user to ensure the similarity between the data of users. Even for users with little-labeled data, it can be less affected to the model. Kang et al. [Kang and Choi (2014, 2011)] combined CSP with a Bayesian model for a multi-subject learning, assuming that spatial patterns across subjects shared a latent subspace.

Another frontier approach is based on the Riemannian geometry of the manifold of symmetric positive definite (SPD) matrix [Barachant, Bonnet, Congedo et al. (2012); Zanini, Congedo, Jutten et al. (2017)]. Riemannian metric learning had been proposed to process the data on the space of SPD matrix [Yger, Berar and Lotte (2017)]. One of the fundamental problems in the application of SPD manifold is to find the nearest neighbor of SPD matrix [Zheng and Song (2018)]. Formally, the Riemannian distance δ_R between two SPD matrices A and B is shown in the following:

$$\delta_R(A, B) = \|\log(A^{-1}B)\|_F = \left[\sum_{i=1}^n \log(\lambda_i)^2\right]^{\frac{1}{2}} \quad (10)$$

[Barachant et al. [Barachant, Bonnet, Congedo et al. (2013)] proposed affine transformation of the covariance matrices of each subject, in order to make the data from other subjects comparable. The results often were poor in the original data space, the affine transformation can provide a much better classification accuracy and precision. The approach is tested on the BCI competition IV dataset IIa and outperforms the conventional CSP method by 18%.

3.3 Generating artificial data

Further, for solving the situation with little-labeled data, some researches had proposed to generate artificial data. The main idea of this approach is to generate multiple trials from the few training trials in order to increase the number of training trials [Lotte (2015)]. Initially, the idea of Lotte [Lotte (2011)] was dividing the training trials into several segments in time domain for generating new artificial trials. The process is schematized in Fig. 1. This method is extremely simple, though effective, but it may conduct mismatching at the boundary and decrease the signal-to-noise ratio.

For maintaining features of both time domain and frequency domain, Lotte [Lotte (2015)] used a short-time Fourier transform to process the training trials. It also proposed a new technique based on analogy. For three selected trials A, B, and C, it computing the mapping relationship from A to B. Using the mapping method to generate a new trial D. [Dinarès-Ferran, Ortner, Guger et al. (2018)] utilized Empirical Mode Decomposition (EMD) to decompose the data to obtain the eigenmode function, and mixed their Intrinsic Mode Function (IMFs). This method also retained the information of time domain and frequency domain.

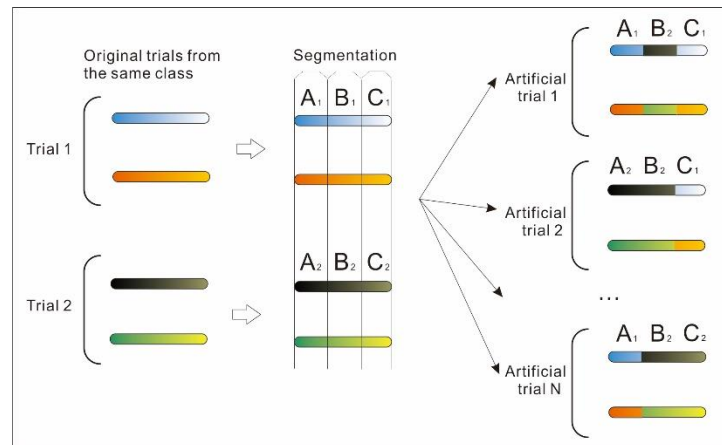


Figure 1: Principle of artificial EEG data generation in the time domain

4 Discussion and prospective

In this paper, we surveyed existing approaches to solve the problem of little-labeled data, and classify them into two categories. The method of enhancing the representation of data can dramatically improve the classification accuracy in BCI design. In the case of little available train data, expanding the training dataset was the most direct and effective way. A large amount of literature had emerged in this field, so this is also the most mature method at present. In the absence of other available data, generating artificial EEG data showed a better performance.

Some points may be improved in the future. With the advancement of human brain physiological structure research, there are more prior knowledge could be added to the feature processing part. It is an important issue to combine more discriminative information from other brain signals. The techniques could be found to dealing with outliers in the little training dataset. Finally, we expect to introduce more transfer learning algorithms, which could make a great progressing in the BCI design.

References

- Ang, K. K.; Chin, Z. Y.; Zhang, H.; Guan, C.** (2008): Filter bank common spatial pattern (FBCSP) in brain-computer interface. *IEEE World Congress on Computational Intelligence*, pp. 2390-2397.
- Barachant, A.; Bonnet, S.; Congedo, M.; Jutten, C.** (2012): Multiclass brain-computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920-928.
- Barachant, A.; Bonnet, S.; Congedo, M.; Jutten, C.** (2013): Classification of covariance matrices using a riemannian-based kernel for BCI applications. *Neurocomputing*, vol. 112, no. 10, pp. 172-178.
- Dimyan, M. A.; Cohen, L. G.** (2011): Neuroplasticity in the context of motor rehabilitation after stroke. *Nature Reviews Neurology*, vol. 7, no. 2, pp. 76.
- Dinarès-Ferran, J.; Ortner, R.; Guger, C.; Solé-Casals, J.** (2018): A new method to

generate artificial frames using the empirical mode decomposition for an EEG-based motor imagery BCI. *Frontiers in Neuroscience*, vol. 12.

Farwell, L. A.; Donchin, E. (1988): Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510-523.

Fazli, S.; Dahne, S.; Samek, W.; Biessmann, F.; Muller, K.-R. (2015): Learning from more than one data source: data fusion techniques for sensorimotor rhythm based brain computer interfaces. *Proceedings of the IEEE*, vol. 103, no. 6, pp. 891-906.

Fazli, S.; Popescu, F.; Danóczy, M.; Blankertz, B.; Müller, K. R. et al. (2009): Subject-independent mental state classification in single trials. *Neural networks*, vol. 22, no. 9, pp. 1305-1312.

Galan, F.; Nuttin, M.; Lew, E.; Ferrez, P. W.; Vanacker, G. et al. (2008): A brain-actuated wheelchair: asynchronous and non-invasive brain-computer interfaces for continuous control of robots. *Clinical neurophysiology*, vol. 119, no. 9, pp. 2159-2169.

Haihong, Z.; Yang, C. Z.; Keng, A. K.; Cuntai, G.; Chuanchu, W. (2011): Optimum spatio-spectral filtering network for brain-computer interface. *IEEE Transactions on Neural Networks*, vol. 22, no. 1, pp. 52-63.

Kang, H.; Choi, S. (2011): Bayesian multi-task learning for common spatial patterns. *IEEE International Workshop on Pattern Recognition in Neuroimaging*, pp. 61-64.

Kang, H.; Choi, S. (2014): Bayesian common spatial patterns for multi-subject EEG classification. *Neural Networks*, vol. 57, no. 9, pp. 39-50.

Kang, H.; Nam, Y.; Choi, S. (2009): Composite common spatial pattern for subject-to-subject transfer. *IEEE Signal Processing Letters*, vol. 16, no. 8, pp. 683-686.

Krauledat, M.; Tangermann, M.; Blankertz, B.; Muller, K.-R. (2008): Towards zero training for brain-computer interfacing. *PloS one*, vol. 3, no. 8, e2967.

Lecuyer, A.; Lotte, F.; Reilly, R. B.; Leeb, R.; Hirose, M. et al. (2008): Braincomputer interfaces, virtual reality, and videogames. *Computer*, vol. 41, no. 10.

Leeb, R.; Lancelle, M.; Kaiser, V.; Fellner, D. W.; Pfurtscheller, G. (2013): Thinking penguin: multimodal brain-computer interface control of a VR game. *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 5, no. 2, pp. 117-128.

Li, Y.; Kambara, H.; Koike, Y.; Sugiyama, M. (2010): Application of covariate shift adaptation techniques in brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 6, pp. 1318-1324.

Lotte, F. (2011): Generating artificial EEG signals to reduce BCI calibration time. *5th International Brain-Computer Interface Workshop*, pp. 176-179.

Lotte, F. (2015): Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces. *Proceedings of the IEEE*, vol. 103, no. 6, pp. 871-890.

Lotte, F.; Guan, C. (2011): Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms. *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355-362.

Meng, J.; Sheng, X.; Zhang, D.; Zhu, X. (2014): Improved semisupervised adaptation for a small training dataset in the brain-computer interface. *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1461-1472.

Millan, J. D. R.; Galan, F.; Vanhooydonck, D.; Lew, E.; Philips, J. et al. (2009): Asynchronous non-invasive brain-actuated control of an intelligent wheelchair. *Engineering in Medicine and Biology Society*, pp. 3361-3364.

Morioka, H.; Kanemura, A.; Hirayama, J. I.; Shikauchi, M.; Ogawa, T. et al. (2015): Learning a common dictionary for subject-transfer decoding with resting calibration. *NeuroImage*, vol. 111, pp. 167-178.

Muller-Putz, G. R.; Scherer, R.; Neuper, C.; Pfurtscheller, G. (2006): Steady-state somatosensory evoked potentials: suitable brain signals for brain-computer interfaces? *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 1, pp. 30-37.

Novi, Q.; Guan, C.; Dat, T. H.; Xue, P. (2007): Sub-band common spatial pattern (SBCSP) for brain-computer interface. *3rd International IEEE/EMBS Conference on Neural Engineering*, 204207.

Pfurtscheller, G.; Da Silva, F. L. (1999): Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology*, vol. 110, no. 11, pp. 1842-1857.

Ramoser, H.; Muller-Gerking, J.; Pfurtscheller, G. (2000): Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE transactions on rehabilitation engineering*, vol. 8, no. 4, pp. 441-446.

Reuderink, B.; Poel, M. (2008): *Robustness of the Common Spatial Patterns Algorithm in the BCI-Pipeline*. Centre for Telematics and Information Technology University of Twente.

Samek, W.; Kawanabe, M.; Muller, K. R. (2014): Divergence-based framework for common spatial patterns algorithms. *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 50-72.

Samek, W.; Vidaurre, C.; Muller, K. R.; Kawanabe, M. (2012): Stationary common spatial patterns for brain-computer interfacing. *Journal of neural engineering*, vol. 9, no. 2, 026013.

Sugiyama, M.; Krauledat, M.; MAZller, K. R. (2007): Covariate shift adaptation by importance weighted cross validation. *Journal of Machine Learning Research*, vol. 8, no. May, pp. 985-1005.

Suk, H. I.; Lee, S. W. (2013): A novel bayesian framework for discriminative feature extraction in brain-computer interfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 286-299.

Tu, Y.; Lin, Y.; Wang, J.; Kim, J. U. (2018): Semi-supervised learning with generative adversarial networks on digital signal modulation classification. *Computers, Materials & Continua*, vol. 55, no. 2, pp. 243-254.

Von Bunau, P.; Meinecke, F. C.; Kiraly, F. C.; Muller, K. R. (2009): Finding stationary subspaces in multivariate time series. *Physical review letters*, vol. 103, no. 21, 214101.

Wang, R.; Shen, M.; Li, Y.; Gomes, S. (2018): Multi-task joint sparse representation

classification based on fisher discrimination dictionary learning. *Computers, Materials & Continua*, vol. 57, pp. 25-48.

Yger, F.; Berar, M.; Lotte, F. (2017): Riemannian approaches in brain-computer interfaces: a review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1753-1762.

Zanini, P.; Congedo, M.; Jutten, C.; Said, S.; Berthoumieu, Y. (2017): Transfer learning: a riemannian geometry framework with applications to brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, no. 99, pp. 1.

Zheng, L.; Song, C. (2018): Fast near-duplicate image detection in riemannian space by a novel hashing scheme. *Computers, Materials & Continua*, vol. 56, no. 3, pp. 529-539.