A Multitask Bayesian Framework for the analysis of Motor Imagery EEG Data

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

***Abstract*— Motor Imagery EEG patterns have been found to vary across subjects, datasets or even for the same subject across sessions. This affects considerable the performance of a BCI system requiring expensive and time-consuming individual calibration sessions to adapt to new subjects. Our work tackles the aforementioned problem using the Bayesian multi-task learning (MTL) framework to share common information across subjects, sessions and datasets. The proposed framework finds a balance between universal (subject-independent) classifiers and subject-specific classifiers by using data from other subjects (even datasets) and combine them to estimate the classifier parameters for the target subject. This combination is achieved by selecting appropriate prior and hyper-prior distributions for the Bayesian MTL framework. Experimental results in three widely used Motor Imagery datasets shown that the proposed MTL framework presents superior performance compared to other state-of-the art methods.**

I. Introduction

A Brain Computer Interface (BCI) using EEG signals aims to create a communication channel between the human brain and the computer [1], [2]. 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. 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. An EEG based BCI system could use various components of the EEG signal to achieve its goal, (such as SSVEPs, ERPs etc) where special interest have attracted the systems based on motor imagery (MI) due to their endogeneneous nature [2]. However, in MI BCI systems the training effort for both the classifier and the user is considerable and represents a significant drawback for its wider adoption [3].

Event related desynchronization (ERD) and synchronization (ERS) phenomena, during the imagination of movement, are useful to MI BCI systems [2]. When a human imagines a movement of his limbs, we observe contralateral changes in the brain activities of the motor cortex [2]. But, the observed EEG signals, are high dimensional, noisy, and present high degree of correlation, making the recognition of MI tasks a very difficult procedure. In addition, during an MI experiment, the acquired EEG signals are produced by the motor cortex as well from other spatially neighboring cortical regions. Thus, it is important to isolate the desired from the undesired signals, a requirement that has motivated the use of apatial filters. Among the spatial filtering techniques reported in MI BCI literature, the one based on Common Spatial Patterns (CSP) is the most prominent due to its nice theoretical properties (such as low SNR, dimensionality reduction) and its experimental validation on various different datasets [4], [5]. However, the basic edition of CSP algorithm is sensitive to noise while overfitting effect can be observed when we have small training sets. CSP is a subject-specific algorithm, that it is unable to use information from other subjects executing the same task with the target subject. To overcome these problems regularized versions of the CSP algorithm were proposed in [5], while a conjunction of CSP algorithm with filter banks (FBCSP) were used in [6].

The design of sophisticated classification schemes with good generalization ability is an important issue for MI BCI. Linear Discriminant Analysis (LDA), or Least Squares (LS) Classifier, presents a well known classifier in BCI community due to its simplicity and efficiency in discriminating MI tasks [7]. LDA generally provides good performance under the hypothesis that the sample covariance matrices are similar between different classes. However, this might not always be the case for the classification of MI tasks due to the potential of severe noise interference. As a consequence, the overfitting problem is likely to occur, resulting in poor classification performance. To overcome this issue, classification algorithms based on regularization techniques have been employed for the classification of MI tasks. One of the most prominent representatives of this category are the Support Vector Machines (SVMs). In conjunction with CSP features, SVM provides state-of-the-art performance for MI tasks classification [7]–[9]. However, apart from the algorithms using regularization techniques, algorithms based on a Bayesian version of LDA (BLDA or SBL) have been also proposed [10], [11]. Under various circumstances these algorithms have shown better performance than LDA or SVM [10]. Finally, while BLDA variants predict the label of a test trial using a sparse linear combination of its features, the Sparse Representation Classification (SRC) scheme [12]–[14] expresses the test trial as a sparse linear combination of the training trials, and its label is determined in terms of the minimum residual norm.

Typical analysis of MI EEG data treats each subject independently. In order to utilize data from all subjects, the learning of CSP spatial filters or classifiers involve three different

* Corresponding author.
[1]V. P. Oikonomou, Spiros Nikolopoulos 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, ikom}@iti.gr
This work was supported by the NeuroMkt project, co-financed by the European Regional Development Fund of the EU and Greek National Funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under RESEARCH CREATE INNOVATE (T2EDK-03661).

approaches: training models using data from many subjects (or multi-subjects covariance matrix estimation) [13], [15], multi-task learning [16] and domain adaptation [15]. Multi-task learning consists in learning simultaneously multiple tasks that share some constraints or prior information. For BCI, each such task is usually to learn a classification model for a given user, which ensures a similarity between users, and thus a better learning even for users with few training data. This has been explored successfully for multi-task linear classifier learning and multi-task CSP learning [15]. Multi-task linear classifier learning [16]–[19] is a sub-field of transfer learning where multiple classification tasks are learned jointly. In our work we propose a new multi-task linear classifier for MI classification. The novelty of our work lies on the use of Sparse Bayesian Learning framework where the shared information between tasks is revealed in the model hyperparameters (or hyperprior) (second level of model's hierarchy), while in [16]–[19] the shared information is revealed on the prior distribution (first level of hierarchy). Experimental results have shown that the proposed algorithm provides superior classification accuracy compared to state-of-the-art algorithms. Multitask learning leading to algorithms that required smaller dataset than classical, superior performance and smaller training time for the subject (calibration time) [15], [17].

The remaining sections are organized as follows. First, we present the proposed multi-task algorithm for MI classification. In this section, the Bayesian Linear Model is introduced for the multi-task learning case. Subsequently, we present the experiments using MI EEG data from 3 different datasets, resulting into 25 subjects. Also, we provide a comparison with well known algorithms. Finally, we conclude this work with a short discussion and future extensions of our research.

II. Methodology

The core problem of MI BCI is to discriminate between the imaginary movement of left vs right hand, which results into a classification problem of two classes. In addressing this problem, we typically collect EEG data from L subjects that perform the same Motor Imagery tasks and for each subject we extract features' vectors (most probably CSP-related features). Let $\boldsymbol{\phi}_1^i, \boldsymbol{\phi}_2^i, \cdots, \boldsymbol{\phi}_{N_i}^i \in \Re^M$ be the set of CSP feature vectors for i-th subject, where N_i is the number of CSP feature vectors (training samples). Also, for each $\boldsymbol{\phi}_{n_i}^i$ we have knowledge of its label $y_{n_i}^i$ (i.e. whether it corresponds to a left or a right-hand imaginary movement). Finally, by collecting all feature vectors into a matrix and all labels into a vector, we have a pair of $(\boldsymbol{\Phi}_i, \boldsymbol{y}_i)$.

We are interested in learning mappings $f_i(\cdot)$ that relate features with labels, i.e. $\boldsymbol{y}_i = f_i(\boldsymbol{\Phi}_i)$. Also, these mappings (or learning tasks) have some common ground since all subjects perform (or at least are aiming to perform) the same motor imagery tasks. Furthermore, this common ground exists, not only in subjects within the same dataset, but also in subjects from different datasets. Typically, the finding of mapping $f_i(\cdot)$ constitutes the "classifier", and the standard approach is to learn one "classifier" for each subject. However, this approach treats each subject separately and independently, failing to exploit the common elements between the various learning tasks.

In our work, we assume that each mapping is linear, and we adopt the concept of Bayesian linear regression model [20] to describe the proposed multitask framework [21]. Bayesian models are one of the most important approaches for multitask learning [21]. Such representations provide the flexibility to model both the individuality of tasks, as well as the correlations between tasks. In particular, learning of the common prior among subjects is part of the training process, and data from all subjects contribute to learning this common prior, thus making it possible to transfer information between subjects. Given the prior, individual models are learned independently. As a result, the estimation of a regressor (task) is affected by both its own training data and data from other subjects (or tasks) related through the common prior, while the interrelationships among the subjects are determined automatically through the joint learning.

In our work we assume that the relation between features and labels is linear, hence each learning task can be described by the following linear regression model:

$$\boldsymbol{y}_i = \boldsymbol{\Phi}_i \boldsymbol{w}_i + \boldsymbol{e}_i \tag{1}$$

where $\boldsymbol{y}_i$ a $N_i \times 1$ vector of labels (1 or -1) related to i-th subject, $\boldsymbol{\Phi}_i$ the $N_i \times M$ matrix containing the features of i-th subject, $\boldsymbol{w}_i$ $M \times 1$ vector of weights (or parameters), and, $\boldsymbol{e}_i$ $N_i \times 1$ vector of noise coming from a zero mean Gaussian random variable with unknown precision (inverse variance) a_0. We can observe that each of the mapping yields a corresponding regression task, and performing multiple such learning tasks has been referred to as multitask learning [21], which aims at sharing information effectively among multiple related tasks. In a more abstract view of our problem we can see that each learning task is a classification problem, and features from one task affect the classification results of another task.

The likelihood function for parameters $\boldsymbol{w}_i$ and a_0 is given by:

$$p(\boldsymbol{y}_i|\boldsymbol{w}_i, a_0) = (2\pi a_0)^{-\frac{N_i}{2}} \exp\Big\{ -\frac{a_0}{2} \|\boldsymbol{y}_i - \boldsymbol{\Phi}_i \boldsymbol{w}_i\|_2^2 \Big\} \tag{2}$$

The parameters of a regression task, $\boldsymbol{w}_i$, are assumed to be drawn from a product of zero-mean Gaussian distributions that are shared by all tasks. Letting $w_{i,j}$ be the j-th parameters for i-th task then we have:

$$p(\boldsymbol{w}_i|\boldsymbol{a}) = \prod_{j=1}^{M} \mathcal{N}(w_{i,j}|0, a_i^{-1}) \tag{3}$$

where the hyperparameters $\boldsymbol{a} = \{a_j\}_{j=1,2,\cdots,M}$ are shared among L subjects, hence, data from all subjects contribute to learning these hyperparameters. To promote sparsity over parameters, we place Gamma priors over hyperparameters $\boldsymbol{a}$ [20], [21]. Also, the same type of prior is placed over noise precision a_0.

$$p(a_0|\alpha,\beta) = Ga(a_0|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} a_0^{\alpha-1} \exp\Big\{ -\beta a_0 \Big\} \quad (4)$$

$$p(\boldsymbol{a}|c,d) = \prod_{j=1}^{M} Ga(a_j|c,d) \quad (5)$$

Also, we can observe here, that noise properties are shared among different subjects (or tasks) (i.e. the noise vectors in Eq. (1) are drawn from the same Gaussian distribution). Finally, it must be noted that we have an hierarchical model, and these types of models are natural to be "dealt" within the Bayesian framework.

Given hyperparameters $\boldsymbol{a}$ and noise precision a_0, we can apply Bayes theorem to find the posterior distribution over $\boldsymbol{w}_i$, which is a Gaussian distribution:

$$p(\boldsymbol{w}_i|\boldsymbol{y}_i,\boldsymbol{a},a_0) = \frac{p(\boldsymbol{y}_i|\boldsymbol{w}_i,a_0)p(\boldsymbol{w}_i)|\boldsymbol{a})}{\int p(\boldsymbol{y}_i|\boldsymbol{w}_i,a_0)p(\boldsymbol{w}_i)|\boldsymbol{a})d\boldsymbol{w}_i} = \mathcal{N}\Big(\boldsymbol{w}_i\Big|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\Big) \quad (6)$$

where

$$\boldsymbol{\mu}_i = a_0\boldsymbol{\Sigma}_i\boldsymbol{\Phi}_i^T\boldsymbol{y}_i \quad (7)$$

$$\boldsymbol{\Sigma}_i = \Big(a_0\boldsymbol{\Phi}_i^T\boldsymbol{\Phi}_i + \boldsymbol{A}\Big)^{-1} \quad (8)$$

where $\boldsymbol{A} = (a_1, a_2, \cdots, a_M)$.

In order to find hyperparameters $\boldsymbol{a}$ and promote sparsity in parameters, the type-II Maximum Likelihood procedure is adopted [20], [22], where the objective is to maximize the marginal likelihood (or its logarithm). Also, a similar procedure is followed for the noise precision. The marginal likelihood $\mathcal{L}(\boldsymbol{a},a_0)$ is given by:

$$\mathcal{L}(\boldsymbol{a},a_0) = \sum_{i=1}^{L}\log\int p(\boldsymbol{y}_i|\boldsymbol{w}_i,a_0)p(\boldsymbol{w}_i|\boldsymbol{a})d\boldsymbol{w}_i = -\frac{1}{2}\sum_{i=1}^{L}\Big(N_i\log(2\pi) + \log|\boldsymbol{C}_i| + \boldsymbol{y}_i^T\boldsymbol{C}_i^{-1}\boldsymbol{y}_i\Big) \quad (9)$$

where $\boldsymbol{C}_i = a_0^{-1}\boldsymbol{I} + \boldsymbol{\Phi}_i\boldsymbol{A}\boldsymbol{\Phi}_i^T$

Differentiating $\mathcal{L}(\boldsymbol{a},a_0)$ with respect to $\boldsymbol{a}$ and a_0 and setting the results into zero [20]–[22] (after some algebraic manipulations) we obtain:

$$a_j^{(new)} = \frac{L - a_j\sum_{i=1}^{L}\Sigma_{i,(j,j)}}{\sum_{i=1}^{L}\mu_{i,j}}, j = 1,2,\cdots,M \quad (10)$$

$$a_0^{(new)} = \frac{\sum_{i=1}L\Big(N_i - M + \sum_{j=1}^{M}a_j\Sigma_{i,(j,j)}\Big)}{\sum_{i=1}^{L}\|\boldsymbol{y}_i - \boldsymbol{\Phi}_i\boldsymbol{\mu}_i\|_2^2} \quad (11)$$

where $\mu_{i,j}$ the j-th element of $\boldsymbol{\mu}_i$ and $\Sigma_{i,(j,j)}$ the j-th diagonal element of covariance matrix $\boldsymbol{\Sigma}_i$. The above analysis suggests an iterative algorithm that iterates between Eqs. (7), (8), (10) and (11), until a convergence criterion is satisfied. Also, the same algorithm can be derived by adopting the EM framework and treating parameters $\boldsymbol{w}_i$ as hidden variables [20]. Finally, based on the above Bayesian formulation, we can derive a fast version of the above algorithm. The fast version provides an elegant treatment of feature vectors by constructing adaptively the matrix $\boldsymbol{\Phi}_i$ through three basic operators: addition, deletion and re-estimation. More information on this subject can be found in [20], [21].

Finally, when a test sample, $\boldsymbol{\phi}_{test}^i$, of features is available for subject i then we decide about its class using the following rule:

$$\boldsymbol{\mu}_i^T\boldsymbol{\phi}_{test}^i \begin{cases} > 0, \text{then class of test sample } 1 \\ \leq 0, \text{then class of test sample } -1 \end{cases} \quad (12)$$

III. Results

In our experiments we have used three MI datasets described below.

Graz Dataset B: The first dataset is the BCI competition IV dataset 2b [23]. Nine subjects participate in this dataset, and, for each participant EEG data have been collected. The EEG data of each participant were collected in five sessions. The first two sessions contain EEG data without feedback, and, the last three sessions with feedback. Each session constitutes from six runs, where each run has 20 trials, 10 trials of each class. Furthermore, the brain activity have been recorded using three electrodes (C3, Cz, and C4) with a sampling frequency of 250Hz. They were bandpass-filtered between 0.5 Hz and 100 Hz, and a notch filter at 50 Hz was enabled. Additional information for this dataset can be found in [23]. This dataset consisted of two classes, namely the motor imagery of left hand (class 1) and right hand (class 2).

MKLab MI dataset: The second dataset consists of EEG signals from 7 participants. These EEG signals were acquired with the EbNeuro cap (64 channels with a sampling frequency of 256 Hz). For each participant two sessions were recorded, In the first session EEG signals were collected without providing feedback to the participant, while, in the second session feedback was provided to the participant. Each session has four runs and in each run 20 trials (10 trials for each MI task) were collected. In order to acquire the EEG data the OpenVIBE platform [24] was adopted using the built in scenario of hand motor imagery based BCI. Finally, this dataset consists of two classes, the motor imagery of left hand (class 1) and right hand (class 2), and we have used the following EEG channels in our analysis: FC5, FC3, FC1, FC2, FC4, FC6, C5, C3, C1, C2, C4, C6, CP5, CP3, CP1, CP2, CP4, CP6.

Graz Dataset 2A: BCI Competition IV Dataset2a [23] comprised 4 classes of MI EEG data from 9 subjects, namely, left hand, right hand, feet and tongue. Two sessions, one for training and the other for evaluation, were recorded from each subject. Each session comprised 288 trials of data recorded with 22 EEG channels and 3 monopolar electrooculogram (EOG) channels (with left mastoid serving as reference). From this dataset we have used only the trials related to left and right hand, 144 trials from each session.

Finally, we can observed here that the MI EEG data from all 25 subjects are divided into two parts, the training part and the testing part. The training data are used to train the classifier, while the test data are used to see/check the performance of the classifier.

For the extraction of EEG features we have used an approach similar to [6]. More specifically, EEG data from all available channels have been extracted from 0.5 sec to 2.5 sec after the presentation of the cue and then a band - pass filter between 8 to 40 Hz has been applied. We have decomposed the EEG data by applying a filter-bank with 15 bands: 8-12 Hz, 10-14 Hz, 12-16Hz,...,36-40Hz. Then, we applied the Common Spatial Filters algorithm to extract the CSP features [5]. By choosing the maximum and minimum eigenvalues for CSP algorithm, we obtained 30 features for each trial. Finally, these features are fed into the classifier. In our algorithm these features are used to construct the feature matrices $\mathbf{\Phi}_i$.

We compare the proposed approach (Multi-task Linear Classifier - mLC) with LDA, SBL, extended LDA (eLDA) and extended SBL (eSBL). The training of LDA and SBL is performed in each subject separately, while the extended versions of LDA and SBL use the training data from all subjects to train a specific classifier for one specific subject. In our first experiment, we have used the MI EEG data from all three datasets (a total of 25 subjects), where a classification task is adopted for each subject. The classification accuracy of each method is shown in Table I. We can observe that the proposed method (mLC) provides the best performance in most subjects, as well as on the best average performance. In the second experiment, we have used only the MKLab dataset, resulting into a 7 tasks problem. The results are shown in Table II. Again we can see that the mLC method provides the best performance. Finally, paired McNemar's tests have shown that the differences between the mLC method and the rest of the methods are statistically significant ($p < 0.001$ in all cases).

In both experiments we can see that the use of data from one subject to construct the classifier of another subject is a useful procedure since the mLC method consistently presents superior performance than the SBL and the LDA. The average performance for the MKLab dataset, in the case of 25 tasks, is (mLC, SBL, LDA, eSBL, eLDA) = (69.70%, 64.76%, 63.57%, 74.40%, 73.69%). Comparing the above average performance with that of Table II we can see the increase in performance for mLC, eSBL and eLDA. This observation indicates that data from one dataset can help to boost the performance of a classifier on another dataset, especially in cases where the number of training trials is small.

In Table III we provide the average classification accuracy over each dataset. We can see that the mLC method shown the best performance in terms of EEG datasets. The SBL and LDA are trained on a subject-by-subject basis so the inter-subject information is neglected. Moreover when only a few training samples are available for each subject, the performance is degraded. Our multitask learning framework is able to overcome the above problems. Also, the construction of a universal classifier (eSBL and eLDA) doesn't seem to be the most appropriate choice since it neglects to take into account the subject's specific features. In contrast, our framework shares information between subjects and datasets in order to construct the classifier for each subject. Overall, our framework seeks to find balance between universal and subject-specific classifiers.

TABLE I

CLASSIFICATION ACCURACY (%) ON THREE DATASETS (25 TASKS)

Sub No	Sub ID	mLC	SBL	LDA	eSBL	eLDA
Graz Dataset 2B						
1	B01	73.13	72.19	72.19	63.44	62.81
2	B02	61.79	58.21	58.21	55.71	55.71
3	B03	56.25	55.00	56.56	55.94	54.69
4	B04	96.25	95.94	95.94	96.88	96.88
5	B05	91.25	91.56	91.25	74.69	75.00
6	B06	82.19	81.56	80.63	76.56	75.31
7	B07	73.75	74.06	73.13	71.56	71.56
8	B08	91.56	92.19	91.56	88.75	88.13
9	B09	85.94	85.63	86.25	84.06	84.69
MKLab Dataset						
10	Mk01	77.50	82.50	81.25	82.50	81.25
11	Mk02	58.75	53.75	52.50	78.75	77.50
12	Mk03	65.00	61.25	61.25	81.25	80.00
13	Mk04	73.75	60.00	57.50	76.25	75.00
14	Mk05	57.50	55.00	53.75	61.25	61.25
15	Mk06	68.75	57.50	53.75	52.50	52.50
16	Mk07	86.67	83.33	85.00	88.33	88.33
Graz Dataset 2A						
17	A01	72.22	70.83	70.14	83.33	84.03
18	A02	49.31	52.78	50.69	44.44	42.36
19	A03	86.11	85.42	85.42	78.47	78.47
20	A04	58.33	56.94	56.25	51.39	50.69
21	A05	73.61	71.53	72.22	54.86	53.47
23	A06	56.25	51.39	52.08	52.78	52.08
24	A07	63.89	64.58	65.28	63.19	62.50
24	A08	84.72	84.03	84.03	81.25	81.94
25	A09	70.83	70.83	70.83	68.06	66.67
Average		72.61	70.72	70.31	70.65	70.11

TABLE II

CLASSIFICATION ACCURACY (%) ON MKLAB DATASET (7 TASKS)

Sub No	Sub ID	mLC	SBL	LDA	eSBL	eLDA
1	Mk01	83.75	82.50	81.25	76.25	76.25
2	Mk02	55.00	53.75	52.50	53.75	50.00
3	Mk03	61.25	61.25	61.25	72.50	72.50
4	Mk04	58.75	60.00	57.50	61.25	62.50
5	Mk05	53.75	55.00	53.75	47.50	48.75
6	Mk06	58.75	57.50	53.75	51.25	51.25
7	Mk07	83.33	83.33	85.00	78.33	76.67
Average		64.94	64.76	63.57	62.98	62.56

IV. CONCLUSIONS

In this work we have proposed a Bayesian multi-task learning framework to share common information across subjects, sessions and datasets. The novelty of our work lies in the use of Sparse Bayesian Learning framework where the shared information between tasks (or subjects) is revealed in the model hyperparameters. Experimental results have shown that our framework succeeds in finding a balance between universal classifiers and subject-specific classifiers, and it

TABLE III
AVERAGE CLASSIFICATION ACCURACY (%) WITH RESPECT TO THE DATASET (25 TASKS)

Dataset	mLC	SBL	LDA	eSBL	eLDA
Graz 2B	79.12	78.48	78.41	74.18	73.86
MKLab	69.70	64.76	63.57	74.40	73.69
Graz 2A	68.36	67.59	67.44	64.20	63.58

presents better overall performance than other competing methods. In the future, we intend to examine various modifications of our approach with respect to the learning of CSP filters and to the prior distributions. Also, we intend to adjust our framework for allowing its application in BCI systems that are based on Steady State Visual Evoked Potentials.

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, *Brain–Computer Interfaces: A Gentle Introduction*, ch. 1. Springer, 2010.

[3] 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.

[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 Magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] N. Brodu, F. Lotte, and A. Lécuyer, "Comparative study of band-power extraction techniques for motor imagery classification," in *2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, pp. 1–6, April 2011.

[10] 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.

[11] V. P. Oikonomou, S. Nikolopoulos, P. Petrantonakis, and I. Kompatsiaris, "Sparse kernel machines for motor imagery eeg classification," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 207–210, July 2018.

[12] Y. Shin, S. Lee, J. Lee, and H.-N. Lee, "Sparse representation-based classification scheme for motor imagery-based brain–computer interface systems," *Journal of Neural Engineering*, vol. 9, p. 056002, Aug 2012.

[13] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, and A. Cichocki, "Sparse group representation model for motor imagery eeg classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, pp. 631–641, March 2019.

[14] V. P. Oikonomou, S. Nikolopoulos, and I. Kompatsiaris, "Robust motor imagery classification using sparse representations and grouping structures," *IEEE Access*, vol. 8, pp. 98572–98583, 2020.

[15] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces," *Proceedings of the IEEE*, vol. 103, pp. 871–890, June 2015.

[16] M. Alamgir, M. Grosse-Wentrup, and Y. Altun, "Multitask learning for brain-computer interfaces," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Y. W. Teh and M. Titterington, eds.), vol. 9 of *Proceedings of Machine Learning Research*, (Chia Laguna Resort, Sardinia, Italy), pp. 17–24, PMLR, 13–15 May 2010.

[17] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Computational Intelligence Magazine*, vol. 11, no. 1, pp. 20–31, 2016.

[18] M. Arvaneh, L. Mihaylova, and A. Azab, "Weighted multi-task learning in classification domain for improving brain-computer interface," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1093–1098, 2018.

[19] K.-H.Fiebig, V. Jayaram, J. Peters, and M. Grosse-Wentrup, "Multi-task logistic regression in brain-computer interfaces," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 002307–002312, 2016.

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

[21] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, pp. 92–106, Jan 2009.

[22] D. J. MacKay, "Bayesian interpolation," *Neural Computation*, vol. 4, pp. 415–447, 1992.

[23] M. Tangermann, K.-L. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. Miller, G. Mueller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the bci competition iv," *Frontiers in Neuroscience*, vol. 6, p. 55, 2012.

[24] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, O. Bertrand, and A. Lécuyer, "Openvibe: An open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments," *Presence: Teleoper. Virtual Environ.*, vol. 19, pp. 35–53, Feb. 2010.