A BAYESIAN FRAMEWORK TO OPTIMIZE DOUBLE BAND SPECTRA SPATIAL FILTERS FOR MOTOR IMAGERY CLASSIFICATION

Soroosh Shahtalebi, Student Member, IEEE and Arash Mohammadi, Member, IEEE

Concordia Institute for Information System Engineering (CIISE), Concordia University,
Montreal, QC, Canada H3G 1M8 Emails: {s_shahta, arashmoh}@encs.concordia.ca

ABSTRACT

The ability to discriminate and classify different tasks is a crucial requirement for any Electroencephalogram (EEG) based Brain computer Interface (BCI). However, the intra and inter subject variability in the brain signal patterns is a bottleneck for developing general BCI systems and needs to be tackled. To address this issue, recently filter banks are deployed to extract frequency specific features, which are then fused at the classification step. On the other hand, some works deploy optimization techniques to design (extract) subject-specific filters (features). While both approaches have reached compromising results, there is still a huge gap between the performance of the techniques and that of humans. In this regard, we propose a Bayesian framework to simultaneously optimize a number of filter banks and spatial filters according to the patterns of brain activity for each subject. Referred to as the Bayesian double band spectro-spatial filter optimization (B2B-SSFO), the proposed method aims at combining the advantages of the two aforementioned approaches, and consists of two bandpass filters providing frequency specific features for each subject. The proposed framework is evaluated on dataset 2b from BCI Competition IV. The proposed B2B-SSFO approach outperforms its counterparts and introduces a robust framework for motor imagery studies.

***Index Terms*— *Index Terms*: Brain-computer Interface (BCI), Common Spatial Patterns, Electroencephalogram (EEG), Motor Imagery.**

1. INTRODUCTION

Human brain is one of the most powerful signal processing units ever known to us in the sense that it can analyze and fuse multiple number of streaming signals from different modalities in an adaptive and real-time fashion. This outstanding ability has intrigued numerous researchers to develop brain-computer interfaces (BCI) [1, 2], which allow the individuals to interact with outer world using their brainwaves. The BCIs play an undeniable role in the human-in-the-loop, cyber-physical systems [3, 4], which aim at further augmenting the human's interaction with the physical world. The BCIs are also a key element in various other applications of significant importance including assistive/rehabilitative systems [5–7], and controlling a neuro-prosthesis for disabled individuals [8]. The plasticity properties of brain has enabled the BCI systems to be deployed in therapeutic applications and has improved the effectiveness of the rehabilitation [5, 9]. Rehabilitation-based BCIs nowadays are of paramount importance in practical applications such as neuro-feedback (NFB), therapy for autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), schizophrenia, and motor rehabilitation for post-stroke patients to name a few. However, the performance of the aforementioned artificial systems rarely matches that of humans, which means that the research in this area is still in its infancy and numerous deficiencies need to be tackled.

A BCI system, typically, consists of different components, which can be classified in to the following two main categories: (i) A brain imaging modality such as Electroencephalogram (EEG), Near Infra-red (NIR), Electrocorticogram (ECoG), or Magnetic Resonance Imaging (MRI), which are used to record brain activities, and; (ii) The signal processing module utilized to process and extract meaningful information form the recordings. From the first category, EEG is usually the prime choice for any practical BCI system, thanks to its unique features including affordability, portability, and high temporal resolution. Processing of EEG signals, on the other hand, can be divided into two major steps, i.e., feature generation, and feature translation. The former is mainly about the pre-processing and filtering of the EEG signals to extract informative features capable of describing the intended underlying phenomena. The latter is about deployment of the extracted features for classification and discrimination of different tasks.

In motor-related EEG studies, several modalities are investigated in the literature which among them, the sensorimotor activities [10] are known to be better representatives. Frontal and parietal cortices of the brain, exhibit rhythmic activities in μ and β bands, which locate in 8-12 Hz and 13-30 Hz, respectively. At the moment that a voluntary movement is about to happen, a drop in the power of these rhythms is observed, known as event related desynchronization (ERD), and once the movement is done, these rhythms emerge again and produce an event-related synchronization (ERS). Among different processing techniques on sensorimotor activities, the common spatial patterns (CSP) [11, 12] technique is known to be an effective tool for classifying the motor imagery (MI) tasks. The CSP provides spatial filters resulting in more precise detection of the ERD and the ERS waveforms. Consequently, the CSP focuses more on the channels which demonstrate higher weights of the ERD and ERS waveforms. The superior power of the CSP approach in discriminating MI tasks has motivated the researchers to further extend the method to enhance its classification performance. A number of these extensions include filter bank common spatial patterns (FBCSP) [13], regularized common spatial patterns (RCSP) [14–16], and separable common spatio-spectral patterns (SCSSP) [17]. The work in [13] showed that the performance of the CSP method drastically improves while the EEG signals are separated into different frequency bands and each frequency band is analyzed separately. This initiated several contributions in this regard, where at one hand, likewise the FBCSP [13], a number of frequency bands with deterministic limits are considered [17]. On the other hand, an optimization problem is

This work was partially supported by the Fonds de Recherche du Québec Nature et Technologies (FRQNT) Grant 2018-NC-206591.

defined to optimize one frequency band limit [18].

The methods deploying filterbanks are shown to be successful for MI classification, however, in such methods a large portion of the processing is wasted and does not provide discriminative features. On the other hand and to best of our knowledge, the methods which are based on optimizing the spectral filters inspect one frequency band and try to optimize its limits. The paper addresses this gap. In particular, we propose a framework to combine the two methods in an intuitive fashion such that the advantages of the filter bank approach align with those of the optimized techniques. In this regard, we propose a Bayesian framework to optimize the limits of two bandpass filters. This filterbank is served as the spectral filtering step of the method and is followed by CSP-based spatial filtering. Referred to as the Bayesian double band spectro-spatial filter optimization (B2B-SSFO), in the proposed framework, the uncertainty of the frequency band limits is modeled by random variables, which are optimized over several iterations by measuring the information between the extracted features and the class labels. In summary, the paper makes the following key contributions:

1. The paper integrates the idea of utilizing filter banks to extract more informative features with that of deriving one subject-specific and optimized spectral filter to analyze the MI signals.
2. In this work, we are optimizing a number of filter banks at the same time while the characteristics of the filter bank are inter-dependent. The proposed solution is superior to existing algorithms that optimize a number of filters separately.

The proposed B2B-SSFO is evaluated for accuracy and compared with different state-of-the-art techniques based of Dataset 2b from BCI Competition IV. It is shown that the proposed B2B-SSFO framework significantly outperforms its state-of-the-art counterparts. The rest of the paper is organized as follows: Section 2 formulates the problem. The proposed B2B-SSFO is developed in Section 3. Simulation results are provided in Section 4. Finally Section 5 concludes the paper.

2. PROBLEM FORMULATION

Throughout the paper, the following notation are used: non-bold letter x denotes a scalar variable; lowercase bold letter $\boldsymbol{x}$ represents a vector, and ;capital bold letter $\boldsymbol{X}$ denotes a matrix. The real domain is represented by $\mathbb{R}$. The transpose and trace of a matrix $\boldsymbol{X}$ are, respectively, denoted by $\boldsymbol{X}^T$, and $\text{Tr}(\boldsymbol{X})$.

We consider supervised learning from EEG signals based on the available set of EEG epochs (trials) denoted by

$$\boldsymbol{X}_i \in \mathbb{R}^{N_{\text{ch}} \times N_{\text{t}}}, \tag{1}$$

for $(1 \leq i \leq N_{\text{Trial}})$, where N_{Trial} is the total number of trials used for processing; N_{ch} is the number of EEG channels (electrodes), and; N_{t} is the number of time samples collected from each electrode in one trial. The training dataset is denoted by $\{(\boldsymbol{X}_i, \Omega_i)\}$, for $(1 \leq i \leq N_{\text{Trial}})$, where Ω_i represents the label corresponding to the ith trial, e.g., Ω_i could be "MI of right hand" or "MI of left hand". Before processing EEG for classifying MI tasks, typically, a pre-processing step is applied. In this stage, initially the power line interference is removed by applying a notch filter, then, desired frequency contents of the signal are extracted by means of a bandpass filter.

Once the aforementioned pre-processing step is complete, the proposed B2B-SSFO framework is implemented based on the training trials to obtain the optimized spectral filters for each subject. The spectral filters are optimized to achieve the maximum classification accuracy for the MI tasks. The optimization problem defines a conditional probability between the characteristics of the spectral filters and the class labels and tries to maximize the probability by adjusting the characteristics of the spectral filters. To extract informative features from the spectrally filtered signals, the CSP technique [12] is then deployed. Intuitively speaking, the CSP performs dimension reduction on the data as well as providing discriminant features from the classes of data. The reason that CSP outperforms some well regarded analytical techniques (e.g., principal component analysis (PCA) and independent component analysis (ICA)) for dimension reduction and feature extraction is that the CSP uses the labels of the data and handles the problem in a supervised fashion. This completes a brief presentation of the problem at hand. Next, we present the proposed B2B-SSFO framework in details.

3. THE B2B-SSFO FRAMEWORK

In this section, we present the proposed B2B-SSFO framework, which is designed for discriminating two classes of MI tasks by means of optimizing the spatio-spectral filters to extract the most discriminant features. We model the uncertainty in the cut-off frequencies of the spectral filters with a prior probability denoted by $p(\boldsymbol{\mathcal{B}})$ over random variable $\boldsymbol{\mathcal{B}}$. Unlike the work in [18], which defines $\boldsymbol{\mathcal{B}} = [b_s, b_e]$ as the cutoff frequencies of a bandpass filter, we define $\boldsymbol{\mathcal{B}}$ as follows

$$\boldsymbol{\mathcal{B}} \triangleq [b_s, b_m, b_e], \tag{2}$$

where $[b_s, b_m]$ defines one bandpass filter, which ideally aims at extracting the μ band contents, and $[b_m, b_e]$ defines another bandpass filter to extract information from the β band. However, the cutoff frequencies are random variables and are optimized in an iterative fashion to increase the classification accuracy. The prior density $p(\boldsymbol{\mathcal{B}})$ describes relative probabilities of different states (frequency bands) in which a single-trial EEG recording is correctly discriminated. The posterior probability distribution denoted by $p(\boldsymbol{\mathcal{B}}|\boldsymbol{X}_i, \Omega_i)$ is then computed based on a single-trial EEG recording $\boldsymbol{X}_i$, for $(1 \leq i \leq N_{\text{Trial}})$, and its corresponding label denoted by Ω_i, as follows

$$p(\boldsymbol{\mathcal{B}}|\boldsymbol{X}_i, \Omega_i) = \frac{p(\boldsymbol{X}_i, \Omega_i|\boldsymbol{\mathcal{B}})p(\boldsymbol{\mathcal{B}})}{p(\boldsymbol{X}_i, \Omega_i)}. \tag{3}$$

However, the term $p(\boldsymbol{X}_i, \Omega_i|\boldsymbol{\mathcal{B}})$ on the right hand side (RHS) of Eq. (3) is too complex in nature resulting in complex $p(\boldsymbol{\mathcal{B}}|\boldsymbol{X}_i, \Omega_i)$, which eliminates the possibility of direct evaluation in closed-form. Alternatively, particle-based approximation techniques [19, 20] are utilized to address this issue. In brief, a set of N_p particles $\{\mathbb{B}(k)\}_{k=1}^{N_p}$ generated from the prior density $p(\boldsymbol{\mathcal{B}})$ are utilized, where $\mathbb{B}(k)$ denotes a particle representing a single filter bank. To be more specific, each particle contains the weight of itself $(\pi(k))$ and the characteristics of the filter bank that it defines $(\mathbb{B}(k) = \{b_s(k), b_m(k), b_e(k), \pi(k)\})$.

We model the bandpass filtering of the EEG recordings as convolution of the input signals with system $h^{(l)}(k)$, for $(l \in \{1, 2\})$, associated with each of the two bandpass filters. Therefore, the filtered signal, denoted by $\boldsymbol{\mathcal{Z}}^{(l)}$, is deterministically obtained as follows

$$\boldsymbol{\mathcal{Z}}_i^{(l)}(k) = h^{(l)}(k) \circledast \boldsymbol{X}_i, \tag{4}$$

where $\circledast$ denotes the convolution operation. The likelihood and the evidence are , therefore, become equal to $p(\boldsymbol{\mathcal{Z}}_i^{(l)}(k), \Omega_i|\boldsymbol{\mathcal{B}}(k))$ and $p(\boldsymbol{\mathcal{Z}}_i^{(l)}(k), \Omega_i)$, respectively. Hence, we can rewrite (3) by replacing the raw EEG signal $\boldsymbol{X}_i$ with its bandpass-filtered version $\boldsymbol{\mathcal{Z}}_i^{(l)}(k)$ as

$$p(\boldsymbol{\mathcal{B}}(k)|\boldsymbol{\mathcal{Z}}_i^{(l)}(k),\Omega_i) = \frac{p(\boldsymbol{\mathcal{Z}}_i^{(l)}(k),\Omega_i|\boldsymbol{\mathcal{B}}(k))p(\boldsymbol{\mathcal{B}}(k))}{p(\boldsymbol{\mathcal{Z}}_i^{(l)}(k),\Omega_i)}. \quad (5)$$

The posterior $p(\boldsymbol{\mathcal{B}}(k)|\boldsymbol{\mathcal{Z}}_i^{(l)}(k),\Omega_i)$ provides all the required information regarding $\boldsymbol{\mathcal{B}}(k)$ which can be obtained from the bandpass-filtered signal $\boldsymbol{\mathcal{Z}}_i^{(l)}(k)$ and its corresponding class label Ω_i. The spectral filtering step is then followed by computing the common spatial patterns of each trial for each frequency band in each particle (i.e., $\boldsymbol{\mathcal{Z}}_i^{(l)}(k)$). In this regard, we first compute the spatial covariance of the trials. Please note that the spatial filter calculation is performed for the signals in each iteration (k) and in each frequency band (l). To simplify the presentation, in the following formulations we have omitted index (k) from the particles.

Since $\boldsymbol{\mathcal{Z}}_i^{(l)}$ is obtained from bandpass filtering of an EEG signal, all classes have zero mean, therefore, the normalized spatial covariance matrix is given by

$$\boldsymbol{C}_i^{(l)} = \frac{\boldsymbol{\mathcal{Z}}_i^{(l)}\boldsymbol{\mathcal{Z}}_i^{(l)^T}}{\text{Tr}(\boldsymbol{\mathcal{Z}}_i^{(l)}\boldsymbol{\mathcal{Z}}_i^{(l)^T})}. \quad (6)$$

As the goal of the CSP approach is to discriminate two classes of data (i.e., 0 and 1), we define $\bar{\boldsymbol{C}^{(l)}}_0$ and $\bar{\boldsymbol{C}^{(l)}}_1$ as the average of spatial covariance matrices of different trials. Based on the computed average covariance matrices ($\bar{\boldsymbol{C}^{(l)}}_0$ and $\bar{\boldsymbol{C}^{(l)}}_1$), the composite spatial covariance matrix denoted by $\boldsymbol{C}^{(l)}$ is computed as $\boldsymbol{C}^{(l)} = \bar{\boldsymbol{C}^{(l)}}_0 + \bar{\boldsymbol{C}^{(l)}}_1$. Next, eigenvalue decomposition is performed as $\boldsymbol{C}^{(l)} = \boldsymbol{U}^{(l)}\boldsymbol{\lambda}^{(l)}[\boldsymbol{U}^{(l)}]^T$, where $\boldsymbol{U}^{(l)}$ is the matrix of eigenvectors associated with the composite covariance, and $\boldsymbol{\lambda}^{(l)}$ is the diagonal matrix of its corresponding eigenvalues. In the next step, a whitening transform is applied on $\boldsymbol{U}^{(l)}$ as

$$\boldsymbol{P}^{(l)} = \sqrt{\left[\boldsymbol{\lambda}^{(l)}\right]^{-1}}[\boldsymbol{U}^{(l)}]^T. \quad (7)$$

Intuitively speaking, the whitening operator equalizes the variance in the space spanned by $\boldsymbol{U}^{(l)}$, i.e., all the eigenvalues of $\boldsymbol{P}^{(l)}\boldsymbol{C}^{(l)}[\boldsymbol{P}^{(l)}]^T$ are equal to one. Using the whitening matrix, the average covariance matrices ($\bar{\boldsymbol{C}^{(l)}}_0$ and $\bar{\boldsymbol{C}^{(l)}}_1$) are transformed as $\boldsymbol{S}_1^{(l)} = \boldsymbol{P}^{(l)}\bar{\boldsymbol{C}^{(l)}}_1[\boldsymbol{P}^{(l)}]^T$ and $\boldsymbol{S}_0^{(l)} = \boldsymbol{P}^{(l)}\bar{\boldsymbol{C}^{(l)}}_0[\boldsymbol{P}^{(l)}]^T$, therefore, $\boldsymbol{S}_0^{(l)}$ and $\boldsymbol{S}_1^{(l)}$ share common eigenvectors denoted by $\boldsymbol{B}^{(l)}$, i.e., $\boldsymbol{S}_0^{(l)} = \boldsymbol{B}^{(l)}\boldsymbol{\lambda}_0^{(l)}[\boldsymbol{B}^{(l)}]^T$ and $\boldsymbol{S}_1^{(l)} = \boldsymbol{B}^{(l)}\boldsymbol{\lambda}_1^{(l)}[\boldsymbol{B}^{(l)}]^T$, with $\boldsymbol{\lambda}_0^{(l)} + \boldsymbol{\lambda}_1^{(l)} = \boldsymbol{I}$, where $\boldsymbol{I}$ denotes an identity matrix of appropriate dimension.

The B2B-SSFO projection matrix corresponding to each bandpass filter is then given by $\boldsymbol{W}^{(l)} = [\boldsymbol{P}^{(l)}]^T\boldsymbol{B}^{(l)}$, which is used to form the decomposition (mapping) of each trial $\boldsymbol{\mathcal{Z}}_i^{(l)}$, for ($1 \leq i \leq N_\text{t}$), as follows

$$\boldsymbol{\mathcal{W}}_i^{(l)} = \left[\boldsymbol{W}^{(l)}\right]^T\boldsymbol{\mathcal{Z}}_i^{(l)}. \quad (8)$$

As the variances of only a small number (m) of signals are suitable for discrimination analysis, only the first and last m rows of $\boldsymbol{\mathcal{W}}_i^{(l)}$ are used for feature extraction. The corresponding features for the trials in each frequency band are extracted as

$$\boldsymbol{f}_i^{(l)} = \log\left(\frac{\text{var}(\boldsymbol{\mathcal{W}}_i^{(l)})}{\sum\text{var}(\boldsymbol{\mathcal{W}}_i^{(l)})}\right), \quad (9)$$

where var$(\cdot)$ denotes the variance operator. Note that, the log-transformation in Eq. (9) is included to magnify the distance between the features. After calculating the frequency specific features based on different spatial filters, we concatenate the features and form a single feature vector for each trial; hence, the feature vector of each trial is given by

$$\boldsymbol{f}_i = \left[[\boldsymbol{f}_i^{(l)}|_{l=0}]^T, [\boldsymbol{f}_i^{(l)}|_{l=1}]^T\right]^T. \quad (10)$$

A set of feature vectors ($\boldsymbol{F}(k)$) is then formed based on the features extracted from each particle as

$$\boldsymbol{F}(k) = \{\boldsymbol{f}(k)_i\}_{i=1}^{N_\text{Trial}} \in \mathbb{R}^{2m\times N_\text{Trial}}. \quad (11)$$

This matrix $\boldsymbol{F}$ incorporates all the features of the training trials into the posterior probability estimation of Eq. (3). In addition, we form vector $\boldsymbol{\Omega} = \{\Omega_i\}_{i=1}^{Ntr}$ which puts together all the trial labels for the training dataset. The goal to find the optimal spatio-spectral filters for discriminative feature extraction, therefore, can be defined as estimation of the posterior distribution given by

$$\begin{aligned} p(\boldsymbol{\mathcal{B}}(k)|\boldsymbol{\mathcal{Z}}_i^{(l)}(k),\Omega_i) &\triangleq p(\boldsymbol{\mathcal{B}}(k)|\boldsymbol{F}(k),\boldsymbol{\Omega}) \\ &= \frac{p(\boldsymbol{F}(k),\boldsymbol{\Omega}|\boldsymbol{\mathcal{B}})p(\boldsymbol{\mathcal{B}}(k))}{p(\boldsymbol{F}(k),\boldsymbol{\Omega})}. \end{aligned} \quad (12)$$

Once the posterior probabilities are estimated, the weights of the particles need to be derived. The weight of each particle is computed as

$$\pi(k) = \frac{p(\boldsymbol{F}(k),\boldsymbol{\Omega}|\boldsymbol{\mathcal{B}}_\mathbf{k})}{\sum_{k=1}^{N_p} p(\boldsymbol{F}(k),\boldsymbol{\Omega}|\boldsymbol{\mathcal{B}}(\mathbf{k}))}. \quad (13)$$

Please refer to [19, 20], for further details on particle filtering. For each iteration of the optimization procedure, the set of particle weights are calculated according to Eq. (13). Then, the following condition is applied on the particles to evaluate the weight

$$\boldsymbol{\Pi} = \bigcup_k(\pi(k) > \tau). \quad (14)$$

It is worth mentioning that τ in Eq. (14) is a random number between 0 and 1, and in each iteration its value changes. If a particle is selected and is included in the set $\boldsymbol{\Pi}$, its associated characteristics of the filter bank remain intact for the next iteration. However, for those particles which are not included in this set, a disturbance following a normal distribution ($\sim\mathcal{N}(0,1)$) is added to the values of the band limits, which define the characteristics of the filter banks. In the next iteration, the effect of the disturbance (changing the characteristics of the filter banks) is evaluated and the same procedure goes on for a specific number of iterations until the particles converge to the optimum value of the band limits for the spectral filter and the particle weights.

In the final step, after deriving the optimized particles, N_p number of support vector machines (SVMs) classifiers are trained based on the features extracted from each particle. In the evaluation phase, the score of each classifier for an unseen trial is multiplied by its corresponding particle weight. The result of the summation over all of the weighted scores is considered as the final decision of the proposed B2B-SSFO framework. Algorithm 1 outlines the training phase of the proposed B2B-SSFO framework. This completes the development of the proposed B2B-SSFO.

4. SIMULATION RESULTS

The proposed B2B-SSFO method is evaluated on the dataset from BCI Competition IV_{2b} [21] which demonstrates the EEG recordings from 9 different subjects. The subjects are all right-handed and had

Algorithm 1 The B2B-SSFO Framework in the Training Phase

Input: EEG recordings $\{\boldsymbol{X}_i\}_{i=1}^{N_{\text{Trial}}}$ and their corresponding labels $\{\Omega_i\}_{i=1}^{N_{\text{Trial}}}$

Output: The optimized particles $\{\mathbb{B}(k)\}_{k=1}^{N_p}$ and N_p number of trained classifiers.

S1. Define N_p number of particles which $\forall k : \pi(k) = \frac{1}{N_p}$.

S2. Initialize the filter bank band limits with a random value.

S3.

for The number of iterations **do**

    for $k \in$ the number of particles **do**

        Spectrally filter the signals by Eq. (4)

        Derive the CSP filters for each frequency band ($\boldsymbol{W}^{(l)}$)

        Spatially filter the signals by Eq. (8)

        Extract features by Eqs. (9) and (10)

    end for

    Form the features matrix $\boldsymbol{F}$ by Eq. (11)

    Compute the posterior probability for Eq. (12)

    Calculate the particle weights using Eq. (13)

    if $\pi(k) < \tau$ **then**

        Add normal noise to $[b_s, b_m, b_e]$ in particle k.

    end if

end for

S4. Train N_p number of SVM classifiers based on the features $\boldsymbol{F}(k)$ from the last iteration.

normal or corrected-to-normal vision. In this dataset, the subjects were asked to perform two motor imagery tasks including “left hand MI” and “right hand MI”. From each subject 5 sessions are recorded, 3 of them for training and the rest for evaluation. Among the training sessions, 2 of them are recorded without providing feedback to the subject and one is recorded when the feedback is enabled. This dataset provides 6 channels of recording, 3 for EEG and 3 for EOG. The EEG channels are recorded form C3, Cz, and C4 points in the 10-20 EEG recording system. It is worth mentioning that we did not take the EOG recordings into account. The signals are recorded with sampling frequency of 250 Hz and bandpass filtered between 0.5-100 Hz. A notch filter to remove the 50 Hz effect of power line on the recordings is also applied.

We performed several experiments on the same subject while changing the parameters of B2B-SSFO, such as the number of particles and the number of iteration. According to our results, we noticed that 30 number of particles and 30 iterations yield almost the best result. It is worth mentioning that here, there is a compromise between selecting higher number of particles and iterations and the run time of the algorithm. In the first iteration, the particles are weighted equally ($1/N_p$); hence, the initial value for the particle weights is set to $1/30$ and the band limits of the spectral filters are initialized with random numbers in the range of 4-40 Hz. We have selected the upper band limit equal to 40 Hz, since numerous studies suggest that informative contents of the EEG signals occur in the frequencies less than $40Hz$. As it is suggested by the BCI competition, the performance for this dataset should be measured in Kappa value. The formulation for the Kappa value is given by

$$\kappa = \frac{\text{CCR} - P_{rand}}{1 - P_{rand}}, \tag{15}$$

where CCR denotes the Correct Classification Rate and P_{rand} is the probability of random classification, which in this experiment is 0.5. In addition, after spatial filtering of the signals, the first and last row ($m = 1$) of the signals are used for feature extraction.

Table 1. Performance comparison for different approaches, tested on $BCIC - IV_{2b}$ dataset. Performance measure is in Kappa (κ) value.

Subjects	**CSP**	**BSSFO**	**FBCSP**	**B2B-SSFO**
Subject 1	15	18.63±4.82	21.25	23.13±3.34
Subject 2	1.43	13.79±3.64	15.71	15±3.87
Subject 3	23.75	7.19±3.12	-4.38	5.62±3.02
Subject 4	36.88	93.63±0.92	61.25	95.63±2.11
Subject 5	23.75	15.63±2.64	55	57.5±6.43
Subject 6	-5.63	49.4±11.62	23.13	55.63±3.94
Subject 7	44.38	52.06±2.17	30.63	45.63±2.83
Subject 8	74.38	77.94±1.06	17.5	79.38±3.19
Subject 9	30.63	59.94±2.19	13.13	64.38±3.32
Average	27.17	43.15	25.91	49.1

In order to evaluate the performance of the proposed B2B-SSFO framework, we have compared the results of our algorithm with the ones from conventional CSP [12], the BSSFO [18], and the FBCSP [21]. Please note that to maintain the fidelity of the comparison, we have not performed the feature selection procedure which is discussed in the original FBCSP method, and we also have trained a linear SVM classifier, particularly for each of the methods. Moreover, the spectral filter that we deployed in this study for the B2B-SSFO, FBCSP and BSSFO is a $5^{th}-$order Butterworth IIR filter. Furthermore, for the implementation of the FBCSP and to evaluate all the methods in rather similar conditions, we used SVM classifier instead of Naive Bayes Parzen window Classifier and removed the feature selection procedure. Moreover, the spectral filter which is used is Butterworth instead of Chebychev type2.

The results of our experiment and comparison with other method are presented in Table 1. Please note that the values in this table are the kappa values multiplied by 100. As it is observed form the results, the proposed B2B-SSFO framework outperforms its counterparts in terms of accuracy. However, it is worth mentioning that the higher accuracy and optimality comes at the cost of higher computations in the training phase, where the algorithm tries to find the optimum characteristics for the spectral filters and also optimum weights for the particles. It is worth noting that as an example, the means of the optimized frequency bands for the 1^{st} subject are $7.75Hz$, $15.89Hz$ and $31.27Hz$, respectively, for b_s, b_m and b_e. Once the training phase is complete and the problem is optimized, the results can be used in the evaluation phase while keeping the execution time rather similar in comparison to other methods. It is worth mentioning that the extension of B2B-SSFO for multi-class problems, where the number of MI tasks is more than 2, could be satisfied by some techniques such as One vs. One (OVO), One vs. All (OVA) [13] or Error Correction Output Coding (ECOC) classifiers [22], which is the focus of our ongoing research.

5. CONCLUSION

In this paper, we proposed the Bayesian double band spectro-spatial filter optimization (B2B-SSFO) framework to leverage the performance of the feature extraction stage by means of an optimization step utilized to optimize the spatio-spectral filters. More specifically, we deployed a Bayesian framework to optimize the spatio-spectral filters based on a CSP feature extraction scheme which results in the most discriminative features for motor imagery tasks. The proposed framework is evaluated on dataset 2b from BCI Competition IV and the results are compared with other well known techniques in motor imagery BCIs. The results indicate that the B2B-SSFO provides a significant performance improvement in comparison to its state-of-the-are counterparts.

6. REFERENCES

[1] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Z. Zhang and K. Kreutz-Delgado, "Evolving Signal Processing for BrainComputer Interfaces," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1567-1584, 2012.

[2] K.R. Muller and B. Blankertz, "Toward noninvasive brain-computer interfaces," *IEEE Signal Processing Magazine*, vol. 23, no. 5, pp. 128-126, 2006.

[3] A.S. Aghaei, B. Donmez, C.C. Liu, D.He, G. Liu, K.N. Plataniotis, H.W. Chen, and Z. Sojoudi, "Smart Driver Monitoring: When Signal Processing Meets Human Factors: In the Driver's Seat," *IEEE Signal Processing Magazine*, vol. 33, no. 6, pp. 35-48, 2016.

[4] G. Schirner, D. Erdogmus, K. Chowdhury, and T. Padir, "The Future of Human-in-the-Loop Cyber-Physical Systems," *Computer*, vol. 46, no. 1, pp. 36–45, 2013.

[5] J.J. Daly and J.R. Wolpaw, "Brain–Computer Interfaces in Neurological Rehabilitation," *The Lancet Neurology*, vol. 7, no. 11, pp. 1032–1043, 2008.

[6] R. Xu, N. Jiang, N. Mrachacz-Kersting, C. Lin, G.A. Prieto, J.C. Moreno, J.L. Pons, K. Dremstrup, and D. Farina, "A Closed-Loop Brain–Computer Interface Triggering an Active Ankle–Foot Orthosis for Inducing Cortical Neural Plasticity," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 7, pp. 2092–2101, 2014.

[7] M. Grosse-Wentrup, D. Mattia, and K. Oweiss, "Using Brain–Computer Interfaces to Induce Neural Plasticity and Restore Function," *Journal of Neural Engineering*, vol. 8, no. 2, pp. 025004, 2011.

[8] R. Leeb, L. Tonin, M. Rohm, L. Desideri, T. Carlson, and J.R. Millán, "Towards Independence: A BCI Telepresence Robot for People with Severe Motor Disabilities," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 969–982, 2015.

[9] L.V. Dokkum, T. Ward, and I. Laffont, "Brain Computer Interfaces for Neurorehabilitation–Its Current Status as a Rehabilitation Strategy Post-Stroke," *Annals of Physical and Rehabilitation Medicine*, vol. 58, no. 1, pp. 3–8, 2015.

[10] B.J. Edelman, B. Baxter, and B. He, "EEG Source Imaging Enhances the Decoding of Complex Right-hand Motor Imagery Tasks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 1, pp. 4–14, 2016.

[11] Z.J. Koles, "The Quantitative Extraction and Topographic Mapping of the Abnormal Components in the Clinical EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 6, pp. 440–447, 1991.

[12] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal Spatial Filtering of Single Trial EEG During Imagined Hand Movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[13] K.K. Ang, Z.Y. 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, pp. 39, 2012.

[14] H. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Regularized Common Spatial Patterns with Generic Learning for EEG Signal Classification," in *IEEE Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, 2009, pp. 6599–6602.

[15] H. Lu, H.L. Eng, C. Guan, K.N. Plataniotis, and A.N. Venetsanopoulos, "Regularized Common Spatial Pattern with Aggregation for EEG Classification in Small-sample Setting," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2936–2946, 2010.

[16] 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, no. 2, pp. 355–362, 2011.

[17] A.S. Aghaei, M.S. Mahanta, and K.N. Plataniotis, "Separable Common Spatio-spectral Patterns for Motor Imagery BCI Systems," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 1, pp. 15–29, 2016.

[18] H.I. Suk and S.W. Lee, "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, 2013.

[19] A. Mohammadi and A. Asif, "Distributed Particle Filter Implementation with Intermittent/Irregular Consensus Convergence," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2572–2587, 2013.

[20] A. Mohammadi and A. Asif, "Distributed Consensus + Innovation Particle Filtering for Bearing/Range Tracking with Communication Constraints," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 620–635, 2015.

[21] M. Tangermann, K.R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K.J. Miller, G. Mueller-Putz, et al., "Review of the BCI Competition IV," *Frontiers in Neuroscience*, vol. 6, pp. 55, 2012.

[22] S. Shahtalebi and A. Mohammadi, "Ternary ECOC Classifiers Coupled with Optimized Spatio-spectral Patterns for Multiclass Motor Imagery Classification," *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 2231–2236, 2017.