Improving classification performance using an adaptive boosting algorithm for Brain Computer Interface (BCI)

Rachel Rajan
M.Tech Student
Department of Electronics and Communication Engineering
Government Engineering College,Thrissur,Kerala-680551
Email: rachelrajan13@gmail.com

Sunny.T.D
Assistant Professor
Department of Electronics and Communication Engineering
Government Engineering College,Thrissur,Kerala-680551
Email: sunnythekkan@rediffmail.com

Abstract—**Brain-computer interfaces (BCI) are an interesting emerging technology providing an efficient communication system between human brain and external devices like computers or neuroprostheses. Among assorts of neuroimaging techniques, electroencephalogram (EEG) is among one of the non-invasive methods exploited mostly in BCI studies. Recent studies have shown that Motor Imagery (MI) based BCI can be used as a rehabilitation tool for patients with severe neuromuscular disabilities. The spatial and spectral information related to brain activities associated with BCI paradigms are usually pre-determined as default in EEG analysis without speculation, which can lead to loses effects in practical applications due to individual variability across different subjects. Hence, to solve this issue, the objective is to design a boosting based common spatial-spectral pattern for BCI based paradigms, where the channel and frequency configurations are modelled as preconditions before learning base learners and introducing a new heuristic of stochastic gradient boost for training the base learners under these preconditions. Apart from that the combination of features each specifically tailored for different physiological phenomena such as Readiness Potential (RP) and Event Related Desynchronization(ERD) might also benefit BCI making it robust against artifacts from non-Central Nervous System (CNS) activity such as eye blinks (EOG) and muscle movements (EMG). The effectiveness and robustness of this algorithm along with feature combination is evaluated on widely used benchmark dataset BCI competition III (IVa).**

Keywords : Brain Computer Interface, Motor Imagery, feature combination, spatial-spectral precondition, stochastic gradient boosting, rehabilitation training.

I. INTRODUCTION

Brain-computer interfaces (BCIs) provides a communication channel for a user to control an external device using only one's brain neural activity. They can be used as a rehabilitation tool for patients with severe neuromuscular disabilities [1], and also range of other applications including neural prosthesis, Virtual Reality (VR), internet access etc. Among different types of neuroimaging techniques, electroencephalogram (EEG) is among one of the non-invasive methods exploited mostly in BCI experiments. And, among them event related desynchronization (ERD), visually evoked potential (VEP), slow cortical potential (SCP), and P300 evoked potentials are widely used for BCI studies.

In accordance with the topographic patterns of brain rhythm modulations, feature extraction using Common Spatial Patterns (CSP) algorithm [2] provides subject-specific and discriminant spatial filters. However, CSP has some limitations, as it is sensitive to frequency bands related to neural activity, because of that the frequency band are manually selected or set to a broad band filter. Apart from that, it also results in overfitting problem when dealt with large number of channels. Hence, the problem of overfitting the classifier and spatial filter rises due to trivial channel configuration. Henceforth, a simultaneous optimization of spatial and spectral filter is highly desirable in BCI studies.

Recent years, motor imagery (MI) based BCI has proven to be an independent system with high classification accuracy. Most of the MI based BCI use brain oscillations at mu (8-12 Hz) and beta (13-26 Hz) rhythms, which displays particular areas of event related desynchronization (ERD)[3] each corresponding to respective MI states (such as right hand or right foot motion). Apart from that, Readiness-potential (RP) [4] which is a slow negative event-related potential that appears before a movement is initiated can also be used as input to BCI to predict future movements. RP is mainly divided into early RP and late RP. Early RP is slow negative potential that begins 1.5s before action, which is immediately followed by late RP that occurs 500ms before the movement. In MI based BCI, combining of features vectors [5] i.e., ERD and RP have shown a significant boost in the classification performance.

In the literature, several number of sophisticated CSP based algorithms have been witnessed especially in the BCI study. A brief review has been presented here. Taking into account of avoid overfiting and selection of optimal frequency bands for CSP algorithm, various methods were proposed.

To avoid overfitting problem, Regularized CSP (RCSP)[6] was proposed, in which the regularization information was added into the CSP learning procedure . The Common Spatio-Spectral Pattern (CSSP)[7], which is an extension of CSP algorithm with time delayed sample. However, due to flexibility issues the Common Sparse Spectral-Spatial Pattern (CSSSP)[8] was proposed, where its FIR filter consists of single time delay parameter. Since, these methods were computationally expensive, a Spectrally-weighted Common Spatial Pattern (SPEC-CSP)[9] was proposed which alternatively optimizes the temporal filter in frequency domain and then the spatial filter in the iteration process. To improve the performance of SPEC-CSP, Iterative Spatio-Spectral Pattern Learning (ISSPL)[10] was proposed which does not rely on statistical assumptions and optimizes all temporal filters under a common optimization framework.

Despite of various studies and advanced algorithm, it is still a challenge to extract optimal spatial spectral filters for BCI studies, so as to be used as a rehabilitation tool especially for disabled subjects. The spatial and spectral information related to brain activities associated with BCI paradigms are usually pre-determined as default in EEG analysis without speculation, which can lead to loses effects in practical applications due to individual variability across different subjects. Hence, to solve this issue, a boosting based common spatial-spectral pattern is designed for BCI based paradigms, where the channel and frequency configurations are modelled as preconditions before learning base learners and introducing a new heuristic of stochastic gradient boost for training the base learners under these preconditions. Apart from that the combination of features each specifically tailored for different physiological phenomena such as Readiness Potential (RP) and Event Related Desynchronization(ERD), which might also benefit BCI making it more robust against artifacts from non-Central Nervous System (CNS) activity such as eye blinks (EOG) and muscle movements (EMG)[5]. The effectiveness and robustness of the designed algorithm along with feature combination is evaluated on widely used benchmark dataset BCI competition III (IVa).
The remaining part of the paper is organized as follows; a detailed design of Boosting based Common Spatial Spectral Pattern algorithm is given in Section II, the data description and experimental implementation is given in Section III, performance comparison results shown in Section IV. Finally, conclusion is given in Section V.

II. BOOSTING BASED COMMON SPATIAL SPECTRAL PATTERN

Under this section, the boosting based Common Spatial Spectral Pattern algorithm is given in detail , it includes modelling the problem, and learning algorithm for the model. The model consists of four stages, data preprocessing which includes multiple spatial and bandpass filtering, feature extraction using common spatial pattern (CSP), training the weak classifiers, and pattern recognition with the help of a combinational model. The architecture of the designed algorithm is shown in Fig 1. The EEG data is firstly spatial filtered and bandpass filtered under multiple spatial-spectral preconditions. Afterwards, the CSP algorithm is applied to extract features of the EEG training dataset, by which weak classifiers $(f_m)_{m=1}^{M}$ are trained and combined to a weighted combination model. Lastly, a new test sample $\hat{x}$ is classified using this combination model.

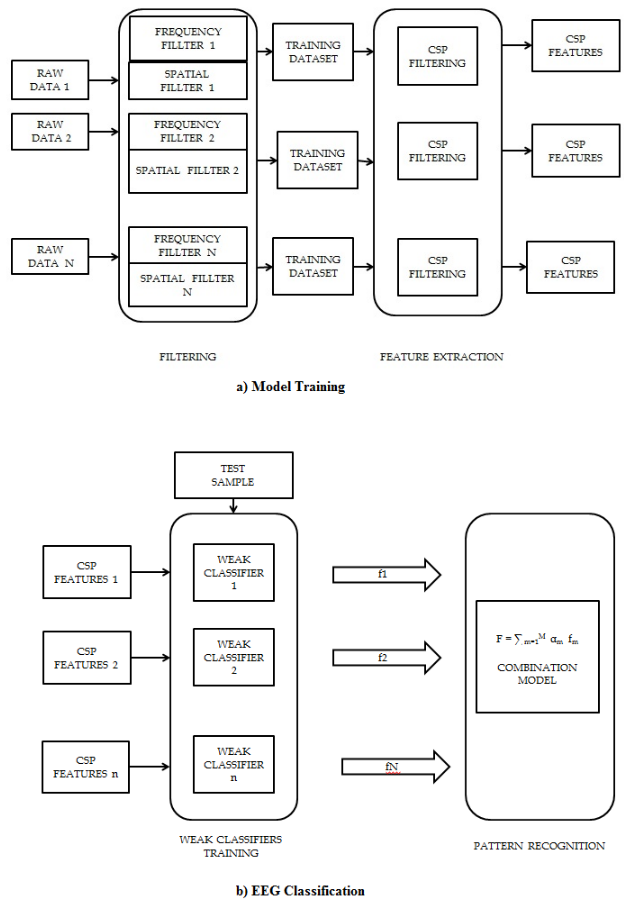

Fig. 1. Architecture of Boosting based Common Spatial Spectral Pattern

A. Problem Design

During BCI studies, the two main concerns are the channel configuration and frequency band, which are predefined as default for implementing EEG analysis. But, predefining these conditions without deliberations leads to poor performance while executing it in a real scenario due to subject variability in EEG patterns. Hence, an efficient and robust configuration is desirable in case of practical applications.

To model this problem, let us denote the training dataset as $E_{train} = (x_i, y_i)_{i=1}^{N}$, where E_i is the i^{th} sample and y_i is its corresponding label. The main aim is to find a subset $\omega \subset \nu$, by using a set of all probable preconditions ν, which generates a combination model F by incorporating all sub models trained under condition$W_M(W_M \in \omega)$ and reducing the misclassification rate on the train dataset E_{train}, given by

$$\omega = argmin_{\omega} \frac{1}{N} |E_i : F(x_i, \omega) \neq y_{i_{i=1}}^{N}| \quad (1)$$

In the following part of this section, 2 homogeneous problems are modelled in detail and then an adaptive boosting algorithm is designed to solve them.

*1) **Spatial Channel selection**:* Let's denote the set of all channels as C and U as a universal set including all possible channels subsets so that each subsetU_m in U satisfies $|U_m| \leq |C|$, here $|.|$ is used to represent the size of the corresponding set. For convenience,a $1x|C|$ binary vector is used to represent U_m, where 1 indicates the corresponding channel in C is selected while 0 not. For channel selection, the aim is to select an optimal channel set $S(S \subset U)$, which produces an optimal combination classifier F on the training data by combining base classifiers learned under different channel set preconditions. Therefore, we get

$$F(E_{train}; S) = \Sigma_{S_m \in S} \alpha_m f_m(E_train; S_m) \quad (2)$$

where F is the optimal combination model, f_m is m^{th} sub model learned with channel set precondition S_m, E_{train} is the training dataset, and α_m is combination parameter. The original EEG E_i is multiplied with the obtained spatial filter, to obtain a projection of E_i on channel set S_m, which is the alleged channel selection. In the simulation, 21 channels were selected, denoted as universal set of all channels,C =(CP6, CP4, CP2, C6, C4, C2, FC6, FC4, FC2, CPZ, CZ, FCZ, CP1, CP3, CP5, C1, C3, C5, FC1, FC3, FC5), where each one indicates an electrode channel

*2) **Frequency Band Selection**:* Generally, spectra is not a discrete variable. Here, the spectra is simplified as a closed interval, denoted as G, where the elements are all integer points, e.g., G is Hz.Here G is split into various sub-bands B and D and denotes a universal set composed of all possible sub-bands. The splitting criteria must satisfy:

1) Cover: $\cup_{B \in D} B = G$;
2) Length: $\forall B = [l, h] \in D, L_{min} \leq h - l \leq L_{max}$, where L_{min} and L_{max} and are two constants to determine the length of B;
3) Overlap: $\forall B_{min} = [l, l+1] \subset G, \exists B_1, B_2 \in D, B_{min} \subseteq B_1 \cap B_2$;
4) Equal: $\forall B_{min} = [l, l+1] \subset G, |\{B : B_{min} \subset B, B \in D\}| = C$,where C is a constant.

The above four criteria makes it possible that the set D will not under-represent the original continuous interval. Accordingly, a sliding window strategy is employed to generate all the sub-bands. The sliding window strategy comprises of four variables: 1)start offset, L ; 2) step length, S; 3) sliding window width, W; and 4) terminal offset, T. In each iteration, the four parameters (L_i, S_i, W_i, T_i) are first initialized, where $i = 1, ..., I$, then the sliding process $Slide(L, S, W, F)$ starts: a band window with width W_i shifts from the start point L_i (left edge) with a step length S_i till it reaches the end point T_i(right edge) and subbands are produced as output of sliding windows:

$$Bandset_i = Slide(L_i, S_i, W_i, T_i) \quad (3)$$

Upon varying the four parameters(L, S, W, T), one universal band set comprising of various sub-bands with different start points, widths and terminal point is produced.

$$D = \cup_{i=1}^{I} Bandset_i \quad (4)$$

While selection of optimal frequency band, the objective to obtain a optimal band set $B(B \subset D)$, so that a optimal combination classifier on the training data is produced.

$$F(E_{train}; B) = \Sigma_{B_m \in B} \alpha_m f_m(E_train; B_m) \quad (5)$$

where f_m is the m_{th} weak classifier learned by sub-band B_m. In the simulation, a bandpass filter was used to filter the raw EEG signal E_i into band B_m.

B. Model Learning Algorithm

Here, the models of channel selection and frequency selection are combined to form a two-tuple ,$\nu_m = (S_m, B_m)$, it is used to denote a spatial-spectral precondition, and ν is represented as a universal set including all these spatial-spectral preconditions. Lastly, the combination function can be computed as

$$F(E_{train}; \nu) = \Sigma_{\nu_m \in \nu} \alpha_m f_m(E_train; \nu_m) \quad (6)$$

Hence, for each spatial-spectral precondition $\nu_m \in \nu$, the training dataset E_{train} is filtered under ν_m. The, the CSP features are obtained by the filtered training dataset E_train , by which a weak classifier $f_m(E_train; \gamma(\nu_m))$is trained using the boosting algorithm, where $\gamma$is the model parameter determined by andν_m and E_{train}. Thus, the classification error defined earlier can be formulated as

$$\{\alpha, \nu\}_0^M = \min_{\{\alpha,\nu\}_0^M} \Sigma_{i=1}^{N} L(y_i, \Sigma_{m=0}^{M} \alpha_m f_m(x_i; \gamma(\nu_m))) \quad (7)$$

A Greedy approach[12] is used to solve(7), which is given in detail below,

$$F(E_{train}, \gamma, \{\alpha, \nu\}_0^M) = \Sigma_{m=0}^{M-1} \alpha_m f_m(E_{train}; \gamma(\nu_m)) + \alpha_M f_M(E_{train}; \gamma(\nu_M)) \quad (8)$$

Transforming the equation(8) into a simple recursion formula we get,

$$F_m(E_{train}) = F_{m-1}(E_{train}) + \alpha_m f_m(E_{train}; \gamma(\nu_m)) \quad (9)$$

We suppose, $F_{m-1}(E_{train})$ is known, then f_m and α_m can be determined by,

$$
\begin{aligned}
&F_m(E_{train}) = F_{m-1}(E_{train}) \\
&\quad + arg\min_f \Sigma_{i=1}^{N} L(y_i, [F_{m-1}(x_i) + \alpha_m f_m(x_i; \gamma(\nu_m))])
\end{aligned} \tag{10}
$$

The problem in (9) is solved by using a steepest gradient descent [13], and the pseudo-residuals are given by,

$$
\begin{aligned}
r_{\pi(i)m} &= -\nabla_F L(y_{\pi(i)}), F(x_{\pi(i)})) \\
&= -[\frac{\partial L(y_{\pi(i)}), F(x_{\pi(i)}))}{F(x_{\pi(i)})}]_{F(x_{\pi(i)})=F_{m-1}(x_{\pi(i)})}
\end{aligned} \tag{11}
$$

Here, the first $\hat{N}$ elements of a random permutation of $\{i\}_{i=1}^{N}$ is given by $\{\pi(i)\}_{i=1}^{\hat{N}}$. Henceforth, a new set $\{(x_{\pi(i)}, r_{\pi(i)m})\}_{i=1}^{\hat{N}}$, which signifies a stochastically partly best descent step direction, is produced and employed to learn $\gamma(\nu m)$ given by,

$$
\gamma_m = arg\min_{\gamma\rho} \Sigma_{i=1}^{\hat{N}} [r_{\pi(i)m}) - \rho f(x_{\pi(i)}; \gamma_m(\nu_m)))] \tag{12}
$$

The combination coefficient α_m is obtained with $\gamma_m(\nu_m)$ as,

$$
\alpha_m = arg\min_{\alpha} \Sigma_{i=1}^{N} L(y_i, F_{m-1}(x_i)) + \alpha f_m(x_i; \gamma(\nu_m)) \tag{13}
$$

Here, each weak classifier f_m is trained under a random subset $\pi(i)_{i=}^{N}$(without replacement) from the full training data set. This ransom subset is used instead of the full sample, to fit the base learner as shown in equation(11) and the model update is computed using equation(12) for the current iteration.

During the iteration, a self-adjusted training data pool P is maintained at background, given in detail in Algorithm 1. Then, the number of copies are computed using local classification error and this copies of incorrectly classified samples are then added to the training data pool.

C. Algorithm : Architecture of Boosting based Common Spatial Spectral Pattern Algorithm

Input: The EEG training dataset given by,$\{x_i, y_i\}_{i=1}^{N}$, L(y,x) is the squared error loss function, number of weak learners denoted by M, and ν is the set of all preconditions.

1) Initialize the training data pool $P_0 = E_{train} = (x_i, y_i)_{i=1}^{N}$.
2) for m = 1 to M.
3) Generate a random permutation $\pi(i)_{i=1}^{|P_{m-1}|}$ = randperm $(n)_{i=1}^{|P_{m-1}|}$.
4) Select the first $\hat{N}$ elements $\pi(i)_{i=1}^{\hat{N}}$ as $(x_i, y_i)_{i=1}^{\hat{N}}$ from P_0.
5) Use this $\pi(i)_{i=1}^{\hat{N}}$ elements to optimize new learner f_m and its related parameters is obtained in output as,.

   Output: F is the optimal combination classifier, weak learners obtained as $\{f_m\}_{m=1}^{M}$, where $\{\alpha_m\}_{m=1}^{M}$ is the weights of weak learners and $\{\nu_m\}_{m=1}^{M}$ is the preconditions under which these weak learners are learned.
6) Input $(x_i, y_i)_{i=1}^{N}$ and ν into a classifier using CSP and extract features to generate family of weak learners.
7) Initialize P_0, $F_0(E_{train}) = arg\min_\alpha \Sigma_{i=1}^{N} L(y_i, \alpha)$
8) Optimalize $f_m(E_{train}; \gamma(\nu_m))$ as defined in equation (12).
9) Optimalize α_m as defined in equation (13).
10) Update P_m using the following steps,
    A Use current local optimal classifier F_m to split the original training set $E_{train} = (x_i, y_i)_{i=1}^{N}$ into two parts $T_{True} = \{x_i, y_i\}_{i:y_i} = F_{m(x_i)}$ and $T_{False} = (x_i, y_i)_{i:y_i} \neq F_{m(x_i)}$.
    Re-adjust the training data pool:
    B **for** each $(x_i, y_i) \epsilon T_{False}$ do.
    C Select out all $(x_i, y_i) \epsilon P_{m-1}$ as $\{x_{n(k)}, y_{n(k)})_{k=1}^{K}\}$.
    D Copy $\{x_{n(k)}, y_{n(k)})_{k=1}^{K}\}$ with $d(d \geq 1)$ times so that we get total $(d+1)K$ duplicated samples.
    E Return these $(d+1)K$ samples into P_{m-1} and we get a new adjusted pool P_m.And $F_m(E_{train}) = F_{m-1}(E_{train}) + \alpha_m f_m(E_{train}; \gamma(\nu_m))$
    F **end for**.
11) **end for**.
12) for each $f_m(E_{train}; \gamma(\nu_m))$, use mapping $F \leftrightarrow \nu$, to obtain its corresponding precondition ν_m
13) **Return** F, $\{f_m\}_{m=1}^{M}$, $\{\alpha_m\}_{m=1}^{M}$, $\{\nu_m\}_{m=1}^{M}$.

With the help of Early stopping strategy[13], the iteration time M is determined to avoid overfitting, using $\hat{N} = N$, doesn't introduce randomness, hence smaller $\frac{\hat{N}}{N}$ fraction, incorporates more overall randomness into the process. In this work, $\frac{\hat{N}}{N} \approx 0.7$ and a comparably satisfactory performance is obtained for the above approximation. While adjusting P, the copies of incorrectly classified samples, d is computed by the local classification error, $e = \frac{|T_{false}|}{N}$ given by,

$$
d = max(1, \lfloor \frac{1-e}{e+\epsilon} \rfloor) \tag{14}
$$

Here, the parameter ϵ is called as accomodation coefficient, and e is always less than 0.5, and decreases during the iterations, so that large weights on samples will be given which were incorrectly classified by strong learners.

III. Experimental Implementation

A. Dataset Description

In order to asses the robustness and performance of the designed algorithm, motor imagery-based dataset from famous BCI competitions III was downloaded, and used for simulation. The detailed description of the datasets is as follows.

Dataset was downloaded from BCI competition III (IV a), where five subjects namely (aa, al, av, aw and ay) performed the motor imagery based task. All these subjects were given instructions to imagine right hand or right foot movement when the visual cue appeared on the screen. The visual cue displayed on the screen for 3.5s and had a intermission of 1.75s to 2.25s in between the task. EEG data was recorded using 118 channels, which was further band-pass filtered between 0.05 and 200 Hz, and then down-sampled to 100 Hz. For analysis, a time frame of 500ms to 2500ms after the commencement of the visual cue was extracted. The task comprised of 280 trials, and for each subject the trials were divided into training and testing trials, and the number of training trials for each subject was 168, 224, 84, 56 and 28, respectively. All analysis were performed in MATLAB R2013a.

B. Data Pre-processing

In order to remove artifacts obtained from eye and muscle movements, FastICA was employed. As, for healthy subjects, the most contributed spectral bands in motor imagery based tasks are α rhythm (8-13 Hz) and β rhythm (14-30 Hz), hence the EEG signals obtained from dataset and are bandpass filtered between 8 and 30Hz. To quantify the lateralized Readiness Potential (LRP), signals were baseline corrected on the interval of 0-300ms and a time frame of 300ms and ending between 1500-3500ms were extracted for analysis. Afterwards the signal was low-pass filtered at 2.5Hz.

C. Feature Extraction and Classification

For comparing the performance and efficiency of the boosting based common spatial spectral pattern algorithm, Regularized CSP (RCSP)[6] was used for feature extraction. In this, model parameter λ for RCSP, were chosen on the training set using a Hold Out validation procedure.

1) Feature Combination: Different combination methods were investigated, since concatenating the features hardly improved the classification accuracy compared to result of single feature. Hence, PROB method [14] was utilized which incorporates independence between ERD and LRP features. The assumption is that, for each class the features are gaussian distributed with equal covariances and of different type are independent, and hence reducing the misclassification error.

2) Feature Selection: Feature selection was done to select relevant features, since as more features cannot improve the training accuracy. Here feature selection was done using Fisher score [15](a variant,$J = \dfrac{\|\mu_+ - \mu_-\|^2}{\sigma_+ + \sigma_-}$), it makes selection by measuring the discrimination of individual feature in the feature vector for classification. Then the features with largest fisher score are selected as most discriminative features.

Linear Discriminant Analysis (LDA)[16] which minimizes the expected risk of misclassification rate was utilized for classification.

IV. Results

The robustness of the designed algorithm was assessed on dataset obtained from BCI competition III (IV a) dataset. The results are presented in two aspects, data visualization and classification accuracy.

A. Data Visualization

At first, the frequency component of the signal was assessed. For this, the fast fourier transform (FFT) was computed by extracting trials related to each motor imagery class (Hand and foot) and then calculating the FFT and averaging the results. The frequency component of the signal after applying a 8-30 Hz band pass filter as shown in Fig 2 , so as to remove eye and muscle movements. Here, it can be seen that the energy obtained the lower frequency band is higher than the energy in the higher frequency band.

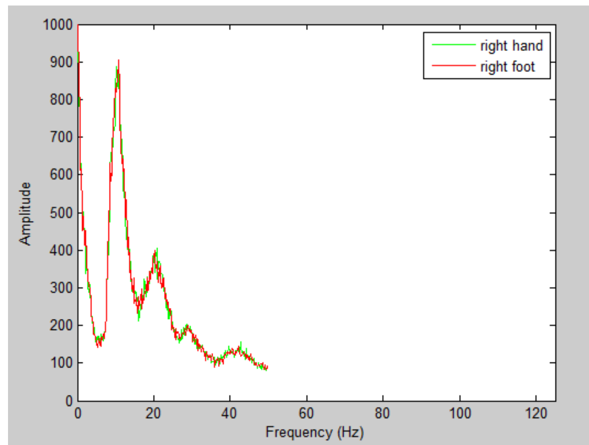

Fig. 2. FFT plots calculated for each motor imagery class (subject aa, electrode C3).

The spatial pattern of hand and foot imagery is given by SP_H and SP_F, most optimal channels can be determined by [17] , $CH_H = find(|SP_H| == max(|SP_H|))$ and $CH_F = find(|SP_F| == max(|SP_F|))$, Here, the most optimal channel are FC1 and CP6. The ERD/ERS plot is shown in Fig 3, the time courses of mu and beta bands from most optimal channels FC1 and CP6 while subject aa was performing hand and foot imagery task. In the figure, the beginning of event related desynchronization (ERD), can be seen during the time frame of 0.5s to 2s in both frequency bands. For ERD quantization, all power values were computed in relation to a baseline power, the percent value shows how much stronger (or weaker) the band power was compared to the baseline value.

To compute the spatial weight for each channel, the quantitative vector, $L = \Sigma_{S_i \in S} \alpha_i S_i$ [2] was used where S_i is the channel sets and α_i are their weights. The spectral weights were computed as given in [9] and then projected onto the frequency bands. In addition, the temporal information

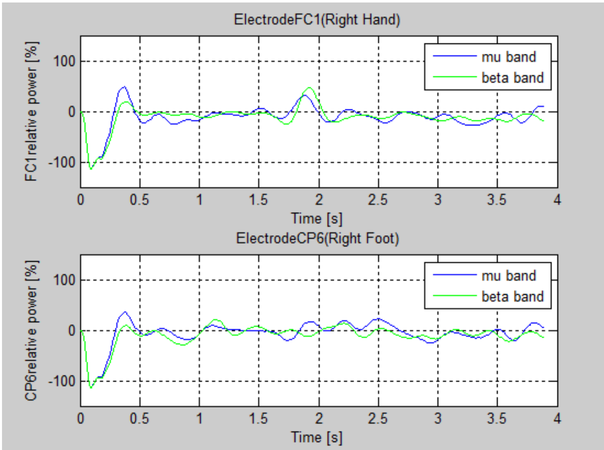

Fig. 3. The ERD plot for Subject aa, right hand and foot motor imagery, channels FC1 and CP6. The mu band and beta band ranges were 8-12 and 14-18 Hz respectively.

were also obtained and visualized.The training dataset are preprocessed under the spatial-spectral precondition $\nu_m \epsilon \nu$, which results in a new dataset on which spatial filtering is done using CSP to obtain the spatial patterns. Then the first two components obtained by CSP are projected onto the space yielding the CSP filtered signal E_m.The peak amplitude P_{mC_i} for E_m and each channel $C_i \epsilon C$. Then the P_{mC_i} was averaged over all set of preconditions $\nu_m \epsilon \nu$, computed as $P_{C_i} = (1)/(|\nu|)\Sigma_{\nu_m \epsilon \nu} \alpha_m P_{mC_i}$ where α_m is the corresponding weight for the m^{th} condition, which is then visualized using a 2-D topoplot map. The 2-D topoplot maps of peak amplitudes of boosting based CSSP filtered EEG in each electrode for all the 5 subjects is shoen in Fig. 4.

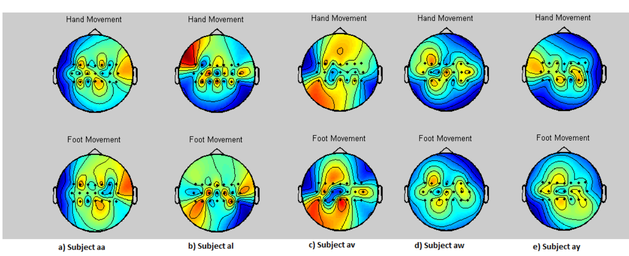

Fig. 4. 2-D topoplot maps of peak amplitude of Boosting based CSSP filtered EEG in each channel for all the 5 subjects in BCI competition III(IVa) dataset.

From the 2-D topoplot maps, it can be observed that for almost all the subjects, the activation related to motor imagery task obtained by designed method were more at the motor cortex areas. Moreover, the channels contributed to right hand imagination were concentrated on the left cortical area, whereas for the right foot imagination activation was more on the central cortical area.

B. Classification Results

The classification results of the test datasets between CSSBP and the other competing methods Regularized CSP (RCSP). In all the subjects the maximum number of iterations, M of the boosting algorithm was set to 180, which was computed using early stopping strategy so as to avoid overfitting, and ϵ was set to 0.05. The percentage of correctly classified samples for different number of iterations(M) in case of Subject aais shown in Fig 5. The accuracy obtained during the cross validation loop with data obtained from all the five subjects is shown in Fig 6. The classification accuracies obtained after combing feature vectors i.e., ERD and RP is shown in Fig.7. From the accuracy plots, it can be seen that the designed algorithm outperformed RCSP, also when the feature vectors were combined accuracies were reduced in all subjects (except al and av) in case of RCSP, whereas it showed a significant improvement in the classification accuracy for the designed method.

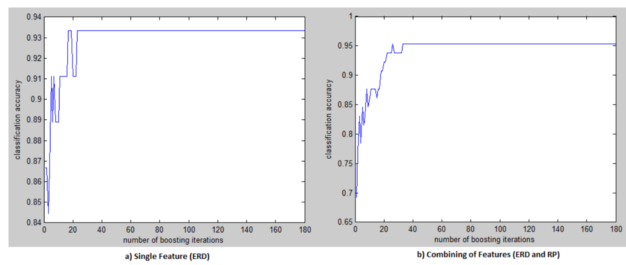

Fig. 5. Percentage of correctly classified samples for different number of iterations(M) in case of Subject aa, where a) ERD features ,b) Combined feature vectors ERD and RP

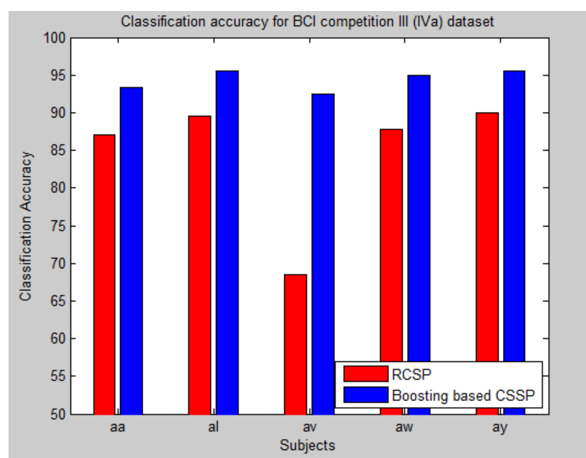

Fig. 6. Classification performances obtained for Regularized CSP (RCSP) and Boosting based CSSP.

V. Conclusion

In this work, a boosting based common spatial-spectral pattern algorithm has been designed for EEG classification.

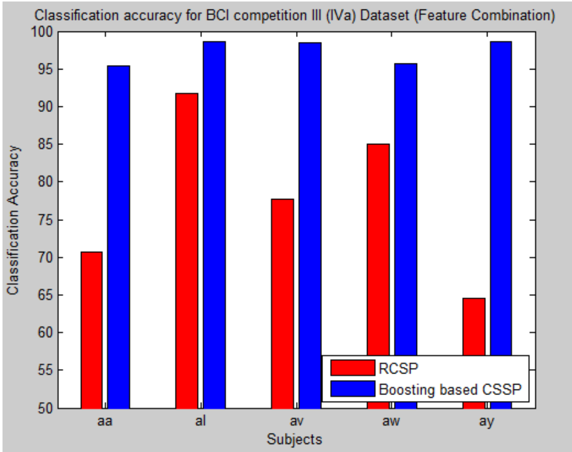

Fig. 7. Classification performances obtained after feature combining for Regularized CSP (RCSP) and Boosting based CSSP.

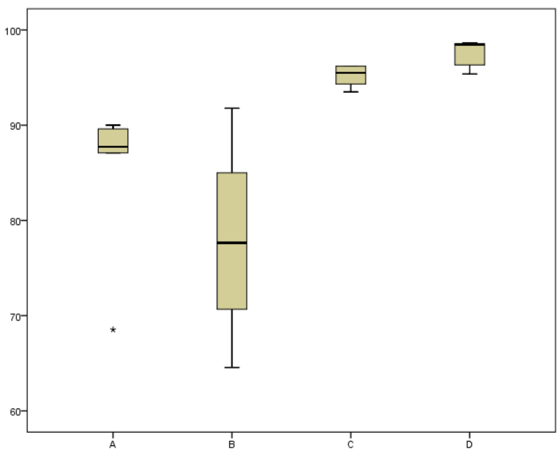

Fig. 8. Boxplots of RCSP and Boosting based CSSP, where A is RCSP, B is RCSP with combined feature vectors, C is Boosting based CSSP, and D is Boosting based CSSP with combined feature vectors.

Here, the channel and frequency configurations are divided into multiple spatial-spectral preconditions by using a sliding window strategy. Under these preconditions, the weak learners are trained using a boosting approach. The motive is to select the most contributed channel groups and frequency bands related to neural activity. From the results, it can be seen that the boosting approach clearly outperformed the other method use for comparison. In addition, combining the widely used feature vectors ERD and readiness potentials significantly improved the classification performance and resulted in increased robustness. The PROB method was utilized which incorporates independence between ERD and LRP features enhanced the performance. This can also be used to better explore the neurophysiological mechanism of underlying brain activities.

Acknowledgment

The authors would like to thank Fraunhofer First, Intelligent Data Analysis Group, and Campus Benjamin Franklin of the Charite' - University Medicine Berlin (http://www.bbci.de/competition/iii), for providing the dataset online.

References

[1] Shih, Jerry J., Dean J. Krusienski, and Jonathan R. Wolpaw. Brain-Computer Interfaces in Medicine. *Mayo Clinic Proceedings 87.3* (2012): 268279. PMC. Web. 21 Apr. 2017.

[2] H. Ramoser, J. Muller-Gerking and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," in *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441-446, Dec 2000.doi: 10.1109/86.895946

[3] G. Pfurtscheller, F.H. Lopes da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles, *Clinical Neurophysiology*, Volume 110, Issue 11, 1 November 1999, Pages 1842-1857, ISSN 1388-2457, http://doi.org/10.1016/S1388-2457(99)00141-8.

[4] Hiroshi Shibasaki, Mark Hallett, What is the Bereitschaftspotential?, *Clinical Neurophysiology*, Volume 117, Issue 11, November 2006, Pages 2341-2356, ISSN 1388-2457, http://doi.org/10.1016/j.clinph.2006.04.025.

[5] Guido Dornhege, Benjamin Blankertz, Gabriel Curio, and Klaus-Robert Mller. 2002. Combining features for BCI. *In Proceedings of the 15th International Conference on Neural Information Processing Systems (NIPS'02)*, Suzanna Becker, Sebastian Thrun, and Klaus Obermayer (Eds.). MIT Press, Cambridge, MA, USA, 1139-1146.

[6] Fabien Lotte, and Cuntai 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, February 2011.

[7] S. Lemm, B. Blankertz, G. Curio and K. R. Muller, "Spatio-spectral filters for improving the classification of single trial EEG," in *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 9, pp. 1541-1548, Sept. 2005. doi: 10.1109/TBME.2005.851521.

[8] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio and K. R. Muller, "Combined Optimization of Spatial and Temporal Filters for Improving Brain-Computer Interfacing," in *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2274-2281, Nov. 2006. doi: 10.1109/TBME.2006.883649

[9] Ryota Tomioka, Guido Dornhege, Guido Nolte, Benjamin Blankertz, Kazuyuki Aihara, and Klaus-Robert Mueller, "Spectrally Weighted Common Spatial Pattern Algorithm for Single Trial EEG Classification", *Mathematical Engineering Technical Reports (METR-2006-40)*, July 2006.

[10] W. Wu, X. Gao, B. Hong and S. Gao, "Classifying Single-Trial EEG During Motor Imagery by Iterative Spatio-Spectral Patterns Learning (ISSPL)," in *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 6, pp. 1733-1743, June 2008. doi: 10.1109/TBME.2008.919125.

[11] J. H. Friedman, Stochastic gradient boosting, Computational Statistics *Data Anal.*, vol. 38, no. 4, pp. 367378, 2002.

[12] J. H. Friedman, Greedy function approximation: A gradient boosting machine, *Ann. Statistics*, pp. 11891232, 2001.

[13] T. Zhang and B. Yu, Boosting with early stopping: Convergence and consistency, *Ann. Statistics*, pp. 15381579, 2005.

[14] G. Dornhege, B. Blankertz, G. Curio and K. R. Muller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," in *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 993-1002, June 2004. doi: 10.1109/TBME.2004.827088

[15] Quanquan Gu, Zhenhui Li, and Jiawei Han. 2011. Generalized Fisher score for feature selection. *In Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI'11)*, Fabio Cozman and Avi Pfeffer (Eds.). AUAI Press, Arlington, Virginia, United States, 266-273.

[16] B. Blankertz, G. Curio, and K.-R. Mller, Classifying single trial EEG: Towards brain computer interfacing, *in Advances in Neural Inf. Proc. Systems (NIPS 01)*, T. G. Diettrich, S. Becker, and Z. Ghahramani, Eds., vol. 14, 2002, pp. 157164.

[17] Yijun Wang, Shangkai Gao and Xiaornog Gao, ”Common Spatial Pattern Method for Channel Selelction in Motor Imagery Based Brain-computer Interface,” *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Shanghai, 2005, pp. 5392-5395.doi: 10.1109/IEMBS.2005.1615701