Improving Classification Performance by Combining Feature Vectors with a Boosting Approach for Brain Computer Interface (BCI)

Rachel Rajan[(✉)] and Sunny Thekkan Devassy

GEC, Thrissur, Kerala, India
rachelrajan13@gmail.com, sunnythekkan@rediffmail.com

Abstract. In the classification of multichannel electroencephalograph (EEG) based BCI studies, the spatial and spectral information related to brain activities associated with BCI paradigms are usually pre-determined as default without speculation, which can lead to loses effects in practical applications due to individual variability across different subjects. Recent studies have shown that feature combination of each specifically tailored for different physiological phenomena such as Readiness Potential (RP) and Event Related Desynchronization (ERD) might benefit BCI making it robust against artifacts. Hence, the objective is to design a CSSBP with combined feature vectors, where the signal is divided into several sub bands using a band pass filter, and this channel and frequency configurations are then modeled as preconditions before learning base learners and introducing a new heuristic of stochastic gradient boost for training the base learners under these preconditions. Results showed that Boosting approach using feature combination clearly outperformed the state-of-the-art algorithms, and improved the classification performance, resulting in increased robustness.

Keywords: Brain computer interface · Motor imagery · Feature combination
Spatial-spectral precondition · Stochastic gradient boosting
Rehabilitation training

1 Introduction

Brain-computer interfaces (BCIs) provide 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 [7], and also a 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.

Rachel Rajan M. Tech student; S. Thekkan Devassy Asst. Professor.

In accordance with the topographic patterns of brain rhythm modulations, feature extraction using Common Spatial Patterns (CSP) algorithm [17] 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) [16] each corresponding to respective MI states (such as right hand or right foot motion). Apart from that, Readiness-potential (RP) [18] 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 1.5 s before action, which is immediately followed by late RP that occurs 500 ms 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 overfitting and selection of optimal frequency bands for CSP algorithm, various methods were proposed. To avoid overfitting problem, Regularized CSP (RCSP) [13] was proposed, in which the regularization information was added into the CSP learning procedure. The Common Spatio-Spectral Pattern (CSSP) [11] is an extension of CSP algorithm with time delayed sample. However, due to flexibility issues the Common Sparse Spectral-Spatial Pattern (CSSSP) [6] was presented, where its FIR filter consists of single time delay parameter. Since, these methods were computationally expensive, a Spectrally-weighted Common Spatial Pattern (SPEC-CSP) [19] was designed 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) [22] 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 CSSBP [12] with combined feature vectors is designed for BCI based paradigms, since the combination of features each corresponding to different physiological phenomena such as Readiness Potential (RP) and Event Related Desynchronization (ERD) can 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]. At first, the EEG signal is first divided into several sub bands using a band pass filter, then the channel and frequency bands are modeled as preconditions before classifying and a heuristic of stochastic gradient boost is used to train the base learners under these preconditions. The effectiveness and robustness of the designed algorithm along with feature combination is evaluated on widely used benchmark dataset BCI competition IV (IIa). The remaining part of the paper is organized as follows; a detailed design of proposed Boosting Algorithm is given in Sect. 2, performance comparison results shown in Sect. 3. Finally, conclusion is given in Sect. 4.

2 Proposed Algorithm

Under this section, a combination model of CSSBP (common spatial spectral boosting pattern) with feature combination is given in detail; it includes modeling the problem, and learning algorithm for the model. The model consists of five stages, data preprocessing which includes multiple spectral filtering by decomposing the signal into several sub bands using a band pass filter and spatial filtering, feature extraction using common spatial pattern (CSP), feature combination, training the weak classifiers, and pattern recognition with the help of a combinational model. The architecture of the designed
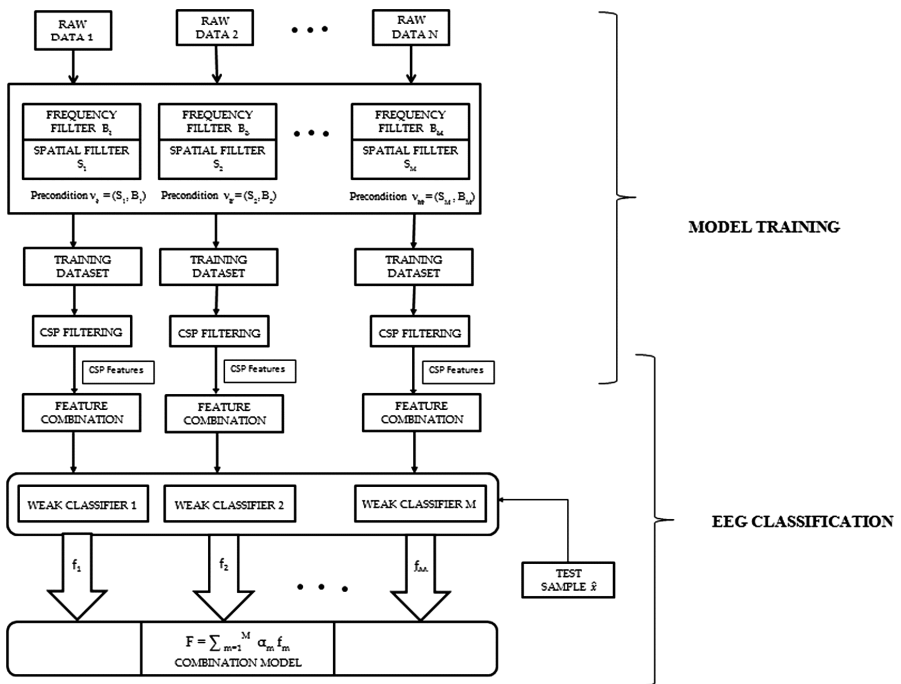

Fig. 1. Block diagram of proposed boosting pattern

algorithm is shown in Fig. 1. The EEG data is firstly spatial filtered and band pass filtered under multiple spatial-spectral preconditions.

Afterwards, the CSP algorithm is applied to extract features of the EEG training dataset and combine these feature vectors, then the 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.

2.1   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 ith 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,

$$\omega = \arg min_{\omega} \frac{1}{N} \left| Ei : F(x_i, \omega) \neq y_{i_{i=1}}^{N} \right| \tag{1}$$

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

Spatial Channel and Frequency Band Selection.   For channel selection, the aim is to select an optimal channel set S($S \subset U$), where U is the universal set including all possible channel subsets for set of channels C so that each subset Um in U satisfies $|Um| \leq |C|$ (here |.| is used to represent the size of the corresponding set), which produces an optimal combination classifier F on the training data by combining base classifiers learned under different channel set preconditions. Therefore, we get,

$$F(E_{train}; S) = \sum\nolimits_{S_m \in S} \alpha_m f_m(E_{train}; S_m) \tag{2}$$

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 work, 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.

For frequency band selection, the spectra denoted as G is simplified as a closed interval, where the elements are all integer points (e.g., G is Hz). Here G is split into various sub-bands B and D as given in [12, 14], which denotes a universal set composed

of all possible sub-bands. While selection of optimal frequency band, the objective is to obtain an optimal band set B (B ⊂ D), so that an optimal combination classifier on the training data is produced.

$$F\left(E_{train};B\right) = \sum\nolimits_{B_m \in B} \alpha_m f_m\left(E_{train};B_m\right) \tag{3}$$

Where f_m is mth weak classifier learned by sub-band B_m. In the simulation study, a fifth order zero phase forward/reverse FIR filter was used to filter the raw EEG signal E_i into sub bands B_m.

2.2 Model Learning Algorithm

Here, the models of channel selection and frequency selection are combined to form a two-tuple, $\vartheta_m = (S_m, B_m)$, 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

$$F\left(E_{train};\vartheta\right) = \sum\nolimits_{\vartheta_m \in \vartheta} \alpha_m f_m\left(E_{train};\vartheta_m\right) \tag{4}$$

Hence, for each spatial-spectral precondition $\vartheta_m \in \vartheta$, the training dataset E_{train} is filtered under ϑ_m. The CSP features are obtained by the filtered training dataset E_{train} and these features of individual physiological nature were combined using PROB method [1]. Let us denote the N features by random variables X_i, i = 1, … , N having class labels as Y ∈ {±1}. An optimal classifier f_i is defined for each feature i on the single feature space D_i hence reducing the misclassification rate. Let $g_{i,y}$ denote the density of $f_i\left(X_i|Y = y\right)$ for each i and labels say y = +1 or −1. Then f is the optimal classifier on the combined feature space D = (D₁, D₂, … , D_N), and X is the combined random variable X = (X₁, X₂, … , X_N), densities of f (X |Y = y) is given by g_y, hence under the assumption of equal class prior for x = $\left(x_1, x_2, … , x_N\right) \in$ D,

$$f_i\left(x_i;\gamma(\vartheta_i)\right) = 1 \leftrightarrow \hat{f}_i\left(x_i, \gamma(\vartheta_i)\right) := \log\left(\frac{g_{i,1}\left(x_i\right)}{g_{i,-1}\left(x_i\right)}\right) > 0 \tag{5}$$

Where γ is the model parameter determined by ϑ_i and E_{train}, and incorporating independence between the features to the above equation results in an optimal decision function given by,

$$f(x;\gamma(\vartheta)) = 1 \leftrightarrow \hat{f}(x;\gamma(\vartheta)) = \sum\nolimits_{i=1}^{N} \hat{f}_i\left(x_i, \gamma(\vartheta_i)\right) > 0 \tag{6}$$

In this, the assumption is that, for each class the features are Gaussian distributed with equal covariance, i.e., $X_i|Y = yN\left(\mu_{i,y}, \sum i\right)$, with $w_i := \sum_i^{-1}\left(\mu_{i,1} + \mu_{i,-1}\right)$, then the classifier,

$$f(x;\gamma(\vartheta)) = 1 \leftrightarrow \hat{f}(x;\gamma(\vartheta)) = \sum_{i=1}^{N} [w_i^T x_i - \frac{1}{2}(\mu_{i,1} + \mu_{i,-1})^T w_i] > 0 \tag{7}$$

Then obtained weak classifier can be rewritten as $f_m(E_{train};\vartheta_m)$, which is trained using the boosting algorithm. Thus, the classification error defined earlier can be formulated as,

$$\{\alpha, \vartheta\}_0^M = min_{\{\alpha,\vartheta\}_0^M} \sum_{i=1}^{N} L\left(y_i, \sum_{m=0}^{M} \alpha_m f_m(x_i;\gamma(\vartheta_m))\right) \tag{8}$$

A Greedy approach [8] is used to solve (8), which is given in detail below,

$$F\left(E_{train}, \gamma, \{\alpha, \vartheta\}_0^M\right) = \sum_{m=0}^{M-1} \alpha_m f_m(E_{train};\gamma(\vartheta_m)) + \alpha_M f_M(E_{train};\gamma(\vartheta_M)) \tag{9}$$

Transforming the Eq. (9) into a simple recursion formula we get,

$$F_m(E_{train}) = F_{m-1}(E_{train}) + \alpha_m f_m(E_{train};\gamma(\vartheta_m)) \tag{10}$$

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

$$F_m(E_{train}) = F_{m-1}(E_{train}) + \arg min_f \sum_{i=1}^{N} L\left(y_i, [F_{m-1}(x_i) + \alpha_m f_m(x_i;\gamma(\vartheta_m))]\right) \tag{11}$$

The problem in (11) is solved by using a steepest gradient descent [9], and the pseudo-residuals are given by,

$$\begin{aligned} r_{\pi(i)m} &= -\nabla_F L\left(y_{\pi(i)}, F(x_{\pi(i)})\right) \\ &= -[\frac{\partial L\left(y_{\pi(i)}, F(x_{\pi(i)})\right)}{F(x_{\pi(i)})}]_{F(x_{\pi(i)})=F_{m-1}(x_{\pi(i)})} \end{aligned} \tag{12}$$

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

$$\gamma(\vartheta_m) = \arg min_{\gamma,\rho} \sum_{i=1}^{\hat{N}} [r_{\pi(i)m} - \rho f(x_{\pi(i)};\gamma_m(\vartheta_m))] \tag{13}$$

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

$$\alpha_m = \arg min_{\alpha} \sum_{i=1}^{N} L\left(y_i, [F_{m-1}(x_i) + \alpha f_m(x_i;\gamma(\vartheta_m))]\right) \tag{14}$$

Here, each weak classifier f_m is trained under a random subset $\{\pi(i)\}_{i=1}^{N}$ (without replacement) from the full training data set. This random subset is used instead of the full sample, to fit the base learner as shown in Eq. (13) and the model update is computed using Eq. (14) 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 is computed using local classification error and these copies of incorrectly classified samples are then added to the training data pool.

2.3   Algorithm 1: Architecture of Proposed Boosting 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 Po $= E_{train} = \{x_i, y_i\}_{i=1}^N$,
(2)   for m $= 1$ to M.
(3)   Generate a random permutation

$$\{\pi(i)\}_{i=1}^{|P_{m-1}|} = \text{randperm}(i)_{i=1}^{|P_{m-1}|}$$

(4)   Select the first $\hat{N}$ elements $\{\pi(i)\}_{i=1}^{\hat{N}}$ as $(x_i, y_i)_{i=1}^{\hat{N}}$, from Po.
(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 $\{\vartheta_m\}_{m=1}^M$ is the preconditions under which these weak learners are learned.

(6)   Input $(x_i, y_i)_{i=1}^N$, and ϑ into a classifier using CSP, extract features and combine these feature vectors to generate family of weak learners.
(7)   Initialize P, $F_0(E_{train}) = \arg\min_\alpha \sum_{i=1}^N L(y_i, \alpha)$
(8)   Optimalize $f_m(E_{train}; \gamma(\vartheta_m))$ as defined in Eq. (10).
(9)   Optimalize α_m as defined in Eq. (11).
(10) Update Pm 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 $(x_i, y_i) \in T_{False}$ do.
C.   Select out all $(x_i, y_i) \in P_{m-1}$ as $\{x_{n(k)}, y_{n(k)}\}_{k=1}^K$.
D.   Copy $\{x_{n(k)}, y_{n(k)}\}_{k=1}^K$ with d(d ≥ 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

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

F.   end for.
(11) end for.

(12) for each $f_m\left(E_{train};\gamma(\vartheta_m)\right)$, use mapping $F \leftrightarrow \vartheta$, to obtain its corresponding precondition ϑ_m.

(13) Return $F, \left\{f_m\right\}_{m=1}^{M}, \left\{\alpha_m\right\}_{m=1}^{M}$, and $\left\{\vartheta_m\right\}_{m=1}^{M}$.

With the help of Early stopping strategy [23], the iteration time M is determined to avoid overfitting, using $\hat{N} = N$, doesn't introduce randomness, hence smaller $\dfrac{\hat{N}}{N}$ fraction, incorporates more overall randomness into the process. In this work, $\dfrac{\hat{N}}{N} = 0.9$ 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, $e = \dfrac{\left|T_{False}\right|}{N}$ is given by,

$$d = \max(1, \left\lfloor \frac{1-e}{e+\in} \right\rfloor) \tag{15}$$

Here, the parameter $\in$ is called as accommodation 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.

3   Result

The robustness of the designed algorithm was assessed on dataset obtained from BCI competition IV (IIa) dataset [2]. In order to remove artifacts obtained from eye and muscle movements, FastICA was employed [15]. For comparing the performance and efficiency of the designed algorithm, Regularized CSP (RCSP) [13] was used for feature extraction. In this, model parameter λ for RCSP, were chosen on the training set using a Hold Out validation procedure. In case of the four-class motor imagery classification task for dataset II, one-versus-rest (OVR) [21] strategy was employed for CSP. PROB method [1] was utilized for feature combination which incorporates independence between ERD and LRP features. 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 (a variant, $J = \dfrac{\left\|\mu_+ - \mu_-\right\|^2}{\sigma_+ + \sigma_-}$) [10], 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) [4] which minimizes the expected risk of misclassification rate was utilized for classification.

Here, the most optimal channel using [20] for all four MI movements i.e., left hand, right hand, foot and tongue were CP4, Cz, FC2, and C1. The 2-D topoplot maps of peak amplitudes of boosting based CSSP filtered EEG in each electrode for subject S1 is shown in Fig. 2.
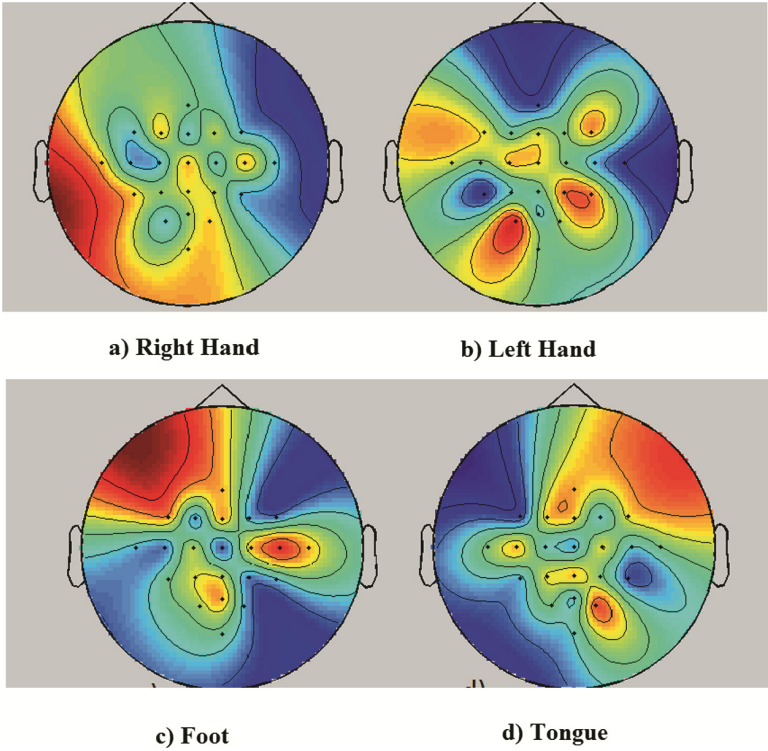
Fig. 2. 2-D topoplot maps of peak amplitude of Boosting based CSSP filtered EEG in each channel for subject S1 in BCI competition IV (II a) dataset.

To compute the spatial weight for each channel, the quantitative vector, $L = \sum_{S_i \in S} \alpha_i S_i$ [17] was used where S_i is the channel sets and α_i are their weights. The spectral weights were computed as given in [12] and then projected onto the frequency bands. In addition, the temporal information were also obtained and visualized. The training dataset are preprocessed under the spatial-spectral pre-condition $\vartheta_m \in \vartheta$, 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 $C_i \in C$. Then the P_{mC_i} is averaged over all set of preconditions $\vartheta_m \in \vartheta$, computed as $P_{C_i} = (\frac{1}{|\vartheta|}) \sum_{\vartheta_m \in \vartheta} \alpha_m P_{mC_i}$ where α_m is the corresponding weight for the mth condition, which is then visualized using a 2-D topoplot map. From the topoplot, it can be observed that the left hand and right hand movement resulted in activation over the right and left hemisphere of the brain, the foot movement activated the central cortical area and tongue showed activation in the motor cortex region.

The classification results of the test dataset for the proposed method and the other competing method i.e., Regularized CSP (RCSP) is detailed as follows. 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 cohen's kappa values for all 9 subjects in the BCI IV(IIa) dataset is shown in Fig. 3. In case of dataset 2, the CSSBP outperformed the RCSP algorithm and showed highest average cohen's kappa value [3]. From the kappa values, it can be seen that when feature vectors are combined in RCSP algorithm, there was a significant improvement in kappa values in all subjects (except for subjects S4 and S6).
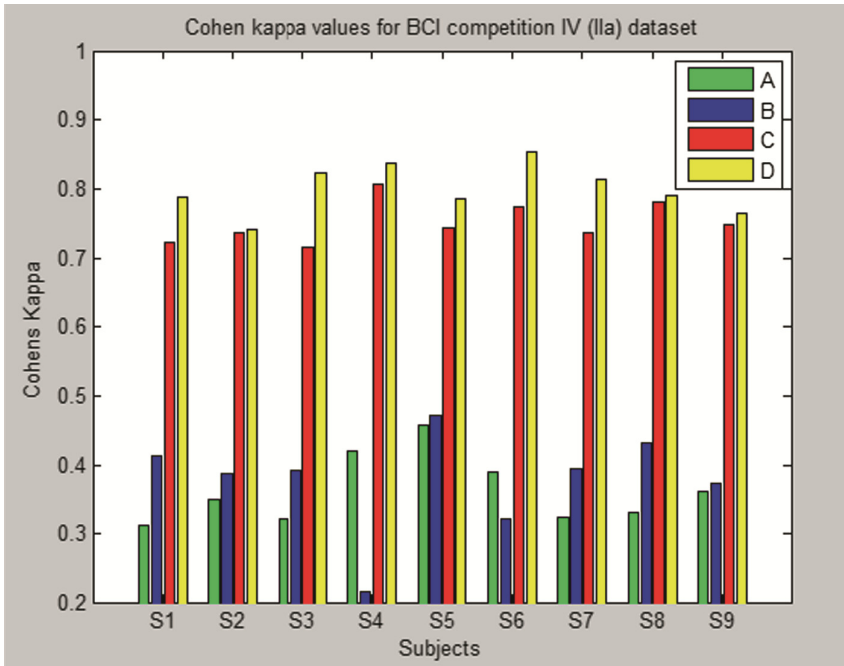
Fig. 3. Cohen's kappa values for all the 9 subjects in BCI IV (II a) dataset, where A is RCSP, B is RCSP with combined feature vectors, C is Boosting based CSSP (CSSBP), and D is Boosting based CSSP (CSSBP) with combined feature vectors.

Whereas the proposed method improved the kappa values compared to the above algorithm and moreover when feature vectors were combined, it outperformed CSSBP with single feature when compared with combined feature vectors. The statistical analysis was done using IBM SPSS ver. 23., it showed significant difference between designed method and the other methods used for comparison in a Mann-Whitney U test. For all the cases, the designed method outperformed for level of significance $p < 0.05$, as shown in Fig. 4.
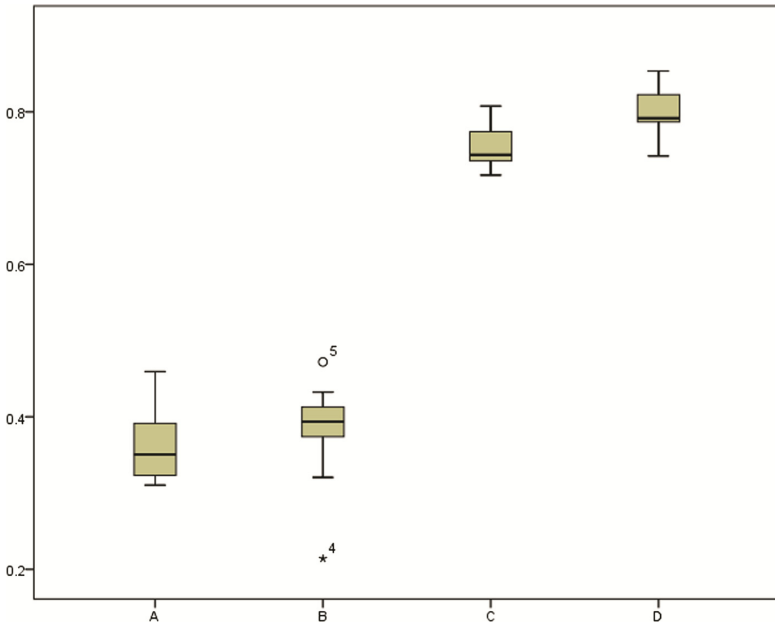
Fig. 4. Boxplots of RCSP and Boosting Approach, where A is RCSP, B is RCSP with combined feature vectors, C is CSSBP, and D is CSSBP with combined feature vectors for BCI IV (IIa) dataset ($p < 0.05$).

4   Conclusion

In this work, a boosting based common spatial-spectral pattern (CSSBP) algorithm with feature combination has been designed for multichannel EEG classification. 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 CSSBP clearly outperformed the other method use for comparison. In addition, combining the widely used feature vectors ERD and readiness potentials (RP) significantly improved the classification performance compared to CSSBP 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. Feature combination of different brain tasks in feedback environment, where the subject is trying to adapt with the feedback scenario might cause the learning process complex and time consuming, so for that this process needs to investigate further in future online BCI experiments.

References

1. Blankertz, B., Curio, G., Müller, K.-R.: Classifying single trial EEG: towards brain computer interfacing. In: Dietterich, T.G., Becker, S., Ghahramani, Z. (eds.) Advances in Neural Information Processing Systems 14, pp. 157–164. MIT Press, Cambridge (2002)
2. Brunner, C., Leeb, R., Muller-Putz, G., Schlogl, A., Pfurtscheller, G.: BCI competition 2008-Graz data set A, Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology (2008). http://www.bbci.de/competition/iv/
3. Cohen, J.: A coefficient of agreement for nominal scales. Educ. Psychol. Measur. **20**(1), 37–46 (1960)
4. Dornhege, G., Blankertz, B., Curio, G., Müller, K.-R.: Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. IEEE Trans. Biomed. Eng. **51**(6), 993–1002 (2004). https://doi.org/10.1109/TBME.2004.827088
5. Dornhege, G., Blankertz, B., Curio, G., Müller, K.-R.: Combining features for BCI. In: Proceedings of the 15th International Conference on Neural Information Processing Systems (NIPS 2002), pp. 1139–1146. MIT Press, Cambridge (2002). http://dl.acm.org/citation.cfm?id=2968618.2968760
6. Dornhege, G., Blankertz, B., Krauledat, M., Losch, F., Curio, G., Müller, K.R.: Combined optimization of spatial and temporal filters for improving brain-computer interfacing. IEEE Trans. Biomed. Eng. **53**(11), 2274–2281 (2006). https://doi.org/10.1109/TBME.2006.883649
7. Jerry, J., et al.: Brain-computer interfaces in medicine. Mayo Clin. Proc. **87**(3), 268–279 (2012). https://doi.org/10.1016/j.mayocp.2011.12.008
8. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**(5), 1189–1232 (2001)
9. Friedman, J.H.: Stochastic gradient boosting. Comput. Stat. Data Anal. **38**(4), 367–378 (2002). https://doi.org/10.1016/S0167-9473(01)00065-2
10. Gu, Q., Li, Z., Han, J.: Generalized Fisher Score for Feature Selection. CoRR abs/1202.3725 (2012). http://arxiv.org/abs/1202.3725
11. Lemm, S., Blankertz, B., Curio, G., Müller, K.R.: Spatio-spectral filters for improving the classification of single trial EEG. IEEE Trans. Biomed. Eng. **52**(9), 1541–1548 (2005). https://doi.org/10.1109/TBME.2005.851521
12. Liu, Y., Zhang, H., Chen, M., Zhang, L.: A boosting-based spatial-spectral model for stroke patients; EEG analysis in rehabilitation training. IEEE Trans. Neural Syst. Rehab. Eng. **24**(1), 169–179 (2016). https://doi.org/10.1109/TNSRE.2015.2466079
13. Lotte, F., Guan, C.: Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. IEEE Trans. Biomed. Eng. **58**(2), 355–362 (2011). https://doi.org/10.1109/TBME.2010.2082539
14. Novi, Q., Guan, C., Dat, T.H., Xue, P.: Sub-band common spatial pattern (SBCSP) for brain-computer interface. In: 2007 3rd International IEEE/EMBS Conference on Neural Engineering, pp. 204–207 (2007). https://doi.org/10.1109/CNE.2007.369647

15. Mishra, P., Singla, S.: Artifact removal from biosignal using fixed point ICA algorithm for pre-processing in biometric recognition. Measur. Sci. Rev. **13**(1), 7–11 (2013). https://doi.org/10.2478/msr-2013-000

16. Pfurtscheller, G., et al.: Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin. Neurophysiol. **110**(11), 1842–1857 (1999). https://doi.org/10.1016/S1388-2457(99)00141-8

17. Ramoser, H., Muller-Gerking, J., Pfurtscheller, G.: Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans. Rehab. Eng. **8**(4), 441–446 (2000). https://doi.org/10.1109/86.895946

18. Shibasaki, H., Hallett, M.: What is the Bereitschaftspotential? Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol. **117**(11), 2341–2356 (2006). https://doi.org/10.1016/j.clinph.2006.04.025

19. Tomioka, R., Dornhege, G., Nolte, G., Blankertz, B., Aihara, K., Müller, K.-R.: Spectrally weighted common spatial pattern algorithm for single trial EEG classification. Mathematical Engineering (Technical reports) (2006)

20. Wang, Y., Gao, S., Gao, X.: Common spatial pattern method for channel selection in motor imagery based brain-computer interface. In: 27th Annual Conference 2005 IEEE Engineering in Medicine and Biology, pp. 5392–5395 (2005). https://doi.org/10.1109/IEMBS.2005.1615701

21. Wu, W., Gao, X., Gao, S.: One-Versus-the-Rest (OVR) algorithm: an extension of common spatial patterns (CSP) algorithm to multi-class case. In: 27th Annual Conference 2005 IEEE Engineering in Medicine and Biology, pp. 2387–2390 (2005). https://doi.org/10.1109/IEMBS.2005.1616947

22. Wu, W., Gao, X., Hong, B., Gao, S.: Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL). IEEE Trans. Biomed. Eng. **55**(6), 1733–1743 (2008)

23. Zhang, T., Yu, B.: Boosting with early stopping: convergence and consistency. Ann. Statist. **33**(4), 1538–1579 (2005)