

Optimising Frequency Band Selection with Forward-Addition and Backward-Elimination Algorithms in EEG-based Brain-Computer Interfaces

Haider Raza, Hubert Cecotti, Girijesh Prasad

Intelligent Systems Research Center, Ulster University, Londonderry, Northern Ireland, UK
raza-h@email.ulster.ac.uk, h.cecotti@ulster.ac.uk, g.prasad@ulster.ac.uk

Abstract— A major problem in a brain-computer interface (BCI) based on electroencephalogram (EEG) recordings is the varying statistical properties of the signals during inter- or intra-session transfers that often lead to deteriorated BCI performances. A filter bank CSP (FBCSP) algorithm typically uses all the features from all the bands to extract and select robust features. In this paper, we evaluate the performance of four methods for frequency band selection applied to binary motor imagery classification: forward-addition (FA), backward-elimination (BE), the intersection and the union of the FA and BE. These methods automatically select and learn the best discriminative sets of frequency bands, and their corresponding CSP features. The performances of the proposed methods are evaluated on binary motor imagery classification using a publicly available real-world dataset (BCI competition 2008 dataset 2A). It is found that the BE method provides the best improvement resulting in an average classification accuracy increase of the BCI system over the FBCSP algorithm, from 77.06% to 79.09%.

Index terms- Brain-computer interface, common spatial patterns, covariate shift, EEG, non-stationarity.

I. INTRODUCTION

A Brain-Computer Interface (BCI) allows a user to express his or her will without muscle exertion thanks to the translation of brain signals into computer commands [1]–[4]. With a non-invasive electroencephalography (EEG) based brain-computer interface (BCI) that operates online, in real-time, and in non-stationary environments, it is required to consider input features that are invariant to changes/shifts in data distribution or covariate shifts. The covariate shifts in the EEG maybe caused by various reasons such as changing user attention level, electrode connections, and/or user fatigue [5], [6]. There are notable covariate shifts observed in the EEG signals during trial-to-trial, and session-to-session transfers [7]–[12].

To date, the low accuracy of classification has been one of the main concerns of the developed BCI systems, which directly affects the decision made by the BCI output [13]. To enhance the performance of BCI systems, several feature extraction, feature selection, and feature classification techniques are proposed in the literature [5], [14]–[16]. A large variety of features have been used in BCI such as, band powers, power spectral density, and time frequency features.

However, the characteristics of the brain responses may change over time, resulting in shifts in feature distributions.

The common spatial patterns (CSP) algorithm is found to be very effective in creating optimal spatial filters that discriminate two classes of EEG signals in motor-imagery (MI) based BCI [13]. However, the realization of these spatial filters is reliant on the selected frequency band. The MI task classification accuracy based on the CSP features may generally be low, when EEG data are unfiltered or have been filtered with improperly selected frequency bands [17]–[20]. Hence, setting a broad frequency range or manually selecting a subject-specific frequency range is quite commonly practiced, while applying the CSP algorithm resulting in less than optimal performance.

To address the problem of subject-specific optimal frequency bands selection for the CSP, this paper comparatively investigates four approaches: Forward-addition (FA) and backward-elimination (BE) of frequency bands; their intersection (i.e., the bands of FA that also belongs to BE); and their union (i.e., the collection of the bands that are obtained from both FA and BE). For processing the EEG data in the MI based BCI, generally four stages are considered as given in Fig.1. In the first stage, the EEG data maybe band-pass filtered into multiple frequency bands. In the second stage, the CSP features maybe extracted from each of the bands (i.e., spatial filtering). In the third stage, the proposed algorithms operate to automatically select the most discriminative frequency bands and the corresponding CSP features. In the fourth stage, a classification algorithm is used to classify the CSP features. The experiments on the real-world data are used to show that the proposed algorithm can be used to select the subject-specific optimal frequency bands. Using the data from the BCI competition-IV 2A, we demonstrate the effectiveness of the proposed approach over the filter bank common spatial pattern approach [21], [22].

This paper proceeds as follows: Section II presents a common spatial pattern algorithm. Section III gives a brief description on filter bank common spatial pattern approach. Section IV consists of proposed backward-elimination and forward-addition algorithm. Section V presents the experimental evaluation. The results are then detailed in

Section VI. Section VII gives a brief discussion. Finally, the impact of the results is discussed in Section VII

II. COMMON SPATIAL PATTERNS (CSP) ALGORITHM

A spatial filter maximizes the variance of spatially filtered signals under one condition, while minimizing it for the other condition. Raw scalp EEG potentials are known to have poor spatial resolution due to volume conduction and smearing effect. If the signal of interest is weak while other sources produce strong signals in the same frequency range, then it is difficult to classify two classes of EEG measurements [23], [24]. The neurophysiological background of the MI based BCIs is that motor activity, both actual and imagined, causes an attenuation or increase of localized neural rhythmic activity called Event-Related Desynchronization (ERD) or Event-Related Synchronization (ERS). The Common-Spatial-Pattern (CSP) algorithm is highly successful in calculating spatial filters for detecting (ERD/ERS) [25].

As seen in Fig. 1, a pair of band-pass and spatial filters in the first and second stages performs spatial filtering of EEG signals that have been band-pass filtered in a specific frequency range. Thus, each pair of band-pass and spatial filter computes the CSP features that are specific to the band-pass frequency range.

CSP is a technique to analyze multi-channel data based on the recording from two classes. It is a data-driven supervised decomposition of signals parameterized by a projection matrix $W \in \mathbb{R}^{C \times C}$, where C is the number of selected channels. W projects the single trial EEG signal $E \in \mathbb{R}^{C \times T}$ in the original sensor space to $Z \in \mathbb{R}^{C \times T}$, which lies in the surrogate sensor space, as follows:

$$Z = WE \quad (1)$$

where E is a $C \times T$ EEG measurement data of a single trial from C channels, and T is the number of time points per channel. The rows of the W are the spatial filters and the columns of W^{-1} are the common spatial patterns. The spatial filtered signal Z given in eq. (1) maximizes the difference in the variance of the two classes. A CSP analysis is performed in order to obtain an effective discrimination of mental states that are characterized by ERD/ERS effects. However, the variances corresponding to only a small number of spatial filters are generally used. The m first and m last rows of Z i.e. $Z_t, t \in \{1 \dots 2m\}$ from the feature vector x_t given in eq. (2) is provided as an input to a classifier, where $m=1$. The CSP features of a single trial are then given by:

$$x_t = \log \left(\frac{\text{var}(Z_t)}{\sum_{i=1}^m \text{var}(Z_t(:, i)) + \text{var}(Z_t(:, C + 1 - i))} \right) \quad (2)$$

Then, the CSP based features from all the ten frequency bands are combined to form the input features for a single classifier.

III. FILTER BANK COMMON SPATIAL PATTERNS

The Filter Bank Common Spatial Pattern (FBCSP) [22] is an approach for combining multiple frequency bands. It generally includes four step-by-step stages of EEG data processing as

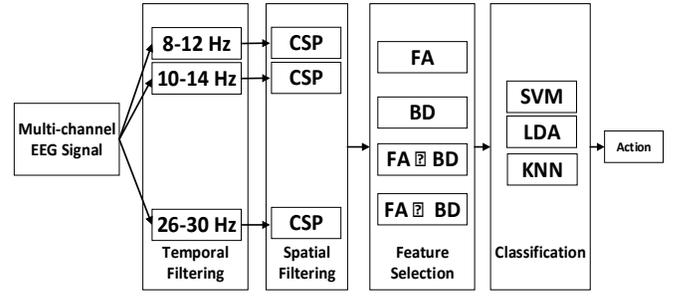


Fig 1. Architecture of Filter Bank Common Spatial Pattern (FBCSP), along with an optimal frequency band selection approach.

discussed previously. According to Fig 1, in the first stage, the filter bank comprises numerous 4th-order Butterworth band-pass filters to filter the data into multiple frequency bands. The second stage performs the spatial filtering on each of the bands using the CSP algorithm. Thus, each pair of the band-pass and spatial filter yields CSP features that are specific to a frequency range of the band-pass filter. In the third stage, a feature selection is carried out to select the best discriminative pairs of frequency bands and corresponding CSP features. In the fourth stage, a classification algorithm is used to classify the CSP features.

IV. PROPOSED FORWARD-ADDITION AND BACKWARD-ELIMINATION METHOD

The proposed optimal frequency band selection for the MI based EEG pattern classification is defined as: given a set of frequency bands, select a subset of size S that minimizes the classification error. To achieve this, we present two approaches for the optimal frequency band selection: the forward-addition (FA) and the backward-elimination (BE).

A. Forward-Addition

In the FA algorithm (see Algorithm I) at the step 1, a set of multiple frequency bands $F = \{F_1, \dots, F_N\}$ is provided as an input, where N is the number of frequency bands in the filter bank. At the step 4, a null set $OF = \{\emptyset\}$ is defined to store the features of the band that has maximum accuracy at each iteration. Another variable NF is defined at the step 5 to store the number of frequency bands, which remains static. Then at the steps 6 & 7, two *for* loops are defined running from 1 to NF , and 1 to the length of F , where F is a dynamic variable, which updates at each iteration. The inner loop is responsible to pick the CSP features for each band and merge it with the features of the set OF into F_{Tr} , wherein OF is initially empty. Once the features are prepared at the step 8, the training accuracy is computed at the step 9. At the step 11, the band that provides the maximum accuracy is selected and added to the set OF and removed from the set F simultaneously. Similarly, through the inner loop, the process is iterated until the set F becomes empty. Then, at the step 15, a set of bands which has a maximum training accuracy is selected and stored into a variable $FOFS$ and provided as output that contains an optimal set of frequency bands for a subject. The complete pseudo code of the FA algorithm is given under Algorithm 1.

ALGORITHM 1

Algorithm I: Forward Addition

1. **INPUT:** A set of band $F = \{F_1, \dots, F_N\}$
 2. **OUTPUT:** FOSF
 3. **FUNCTION:** [FOSF]: $f_Forward_Addition$
 4. $OF = \{\emptyset\}$
 5. $NF = \text{length}(F)$
 6. **FOR** $i = 1: NF$
 7. **FOR** $j = 1: \text{length}(F)$
 8. $F_{Tr} = \text{Add}(F_j, OF)$
 9. $\text{Acc}(j) = \text{Compute}(\text{CV Accuracy on } F_{Tr})$
 10. **END**
 11. $[\text{Ind}(i), \text{Max_CV_Acc}(i)] = \text{Max}(\text{Acc})$
 12. $\text{Add}(F_{\text{Ind}(i)})$ to OF and Update OF
 13. $\text{Remove}(F_{\text{Ind}(i)})$ from F and Update F
 14. **END**
 15. $\text{FOSF} = \text{Max}(\text{Max_CV_Acc})$
 16. **Return FOSF**
-

ALGORITHM II

Algorithm II: Backward Elimination

1. **INPUT:** A set of band $F = \{F_1, \dots, F_N\}$
 2. **OUTPUT:** BOSF
 3. **FUNCTION:** [BOSF]: $f_Backward_Elimination$
 4. $NF = \text{length}(F)$
 5. **FOR** $i = 1: NF$
 6. **FOR** $j = 1: \text{length}(F)$
 7. $F_{Tr}(j) = F - F_j$
 8. $\text{Acc}(j) = \text{Compute}(\text{CV Accuracy on } F_{Tr}(j))$
 9. **END**
 10. $[\text{Ind}(i), \text{Max_CV_Acc}(i)] = \text{Max}(\text{Acc})$
 11. $C = \text{differece}(F, F_{Tr}(i))$
 12. $\text{Remove}(C)$ from F and Update F
 13. **END**
 14. $\text{BOSF} = \text{Max}(\text{Max_CV_Acc})$
 15. **Return BOSF**
-

B. Backward-Elimination

In the BE algorithm (see Algorithm II) at the step 1, a set of multiple frequency bands $F = \{F_1, \dots, F_N\}$ is provided as an input, where N is the number of frequency bands in the filter bank. At the step 4, a variable NF is defined to store the number of total frequency bands that remains static. Then at the steps 5 & 6, two *for* loops are defined running from 1 to NF , and 1 to the length of F . For the inner loop at the step 7, a set of CSP features are included into the variable $F_{Tr}(j)$. The set $F_{Tr}(j)$ is comprised of all the bands from set F , except a band F_j . On the set $F_{Tr}(j)$, the training accuracy is computed using a 10-fold cross-validation technique at the step 8 and stored into the variable $\text{Acc}(j)$. Now at the step 10, a set of bands that has maximum accuracy is selected along with its index into $\text{Ind}(i)$. At the step 11, the difference between the set F and the selected maximum accuracy band set $F_{Tr}(i)$ is measured and stored into variable C , which gives the information about the band that is not involved in obtaining the maximum accuracy. Now, the variable C contains the band information that causes worst performance and has been deleted from the set F at the step 12. This procedure repeats until all the set F becomes empty. At the step 14, a set of band

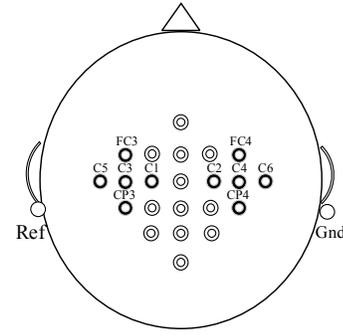


Fig.2. Electrode montage corresponding to the international 10-20 system: Dataset 2A, among all 22 EEG channels, total 10 channels are selected as shown in black filled hollow circles.

that results into maximum accuracy is selected and stored into a variable BOSF. The complete pseudo code of the BE algorithm is given in Algorithm II.

V. EXPERIMENTAL EVALUATION

A. Dataset Description

The BCI Competition IV dataset 2A [26] is used in the investigation of the proposed method. It comprised of EEG data collected from nine subjects, namely [A01-A09], that were recorded during two sessions on separate days. The data consists of 25 channels, and includes 22 EEG channels, and 3 mono-polar EOG channels with sampling frequency of 250 Hz. Among the 22 EEG channels, 10 channels are selected for this study, which are responsible for capturing most of the MI activities. The selected channels are shown in Fig. 2. The data was collected on four different MI tasks: left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). Each session consists of six runs separated by short breaks, each run comprised of 48 trials (12 for each class). The total numbers of 288 trials are in each session. Only the class 1 and the class 2 for left hand and right hand MIs were considered in this study. The MI data from the session-I was used to select the optimal frequency bands and train the classifier, and the MI data from the session-II was used for evaluation purposes. Each trial involves a paradigm of 7.5 seconds, for more detail refer to [26].

B. Data Processing and Feature Extraction

1) Band-Pass Filtering and segment selection

The first stage of signal processing employs a filter bank that decomposes the EEG signals into multiple frequency bands. A total of 10 band-pass filters are used, namely [B1-B10]: [8-12], [10-14], [12-16], [14-18], [16-20], [18-22], [20-24], [22-26], [24-28], [26-30] Hz. These frequency ranges are used because mu (μ) [8-12] Hz and beta (β) [14-30] Hz rhythms cover a stable frequency response over the range of [8-30] Hz. In the next sections, we consider a time segment of 3 s after the cue onsets for the dataset.

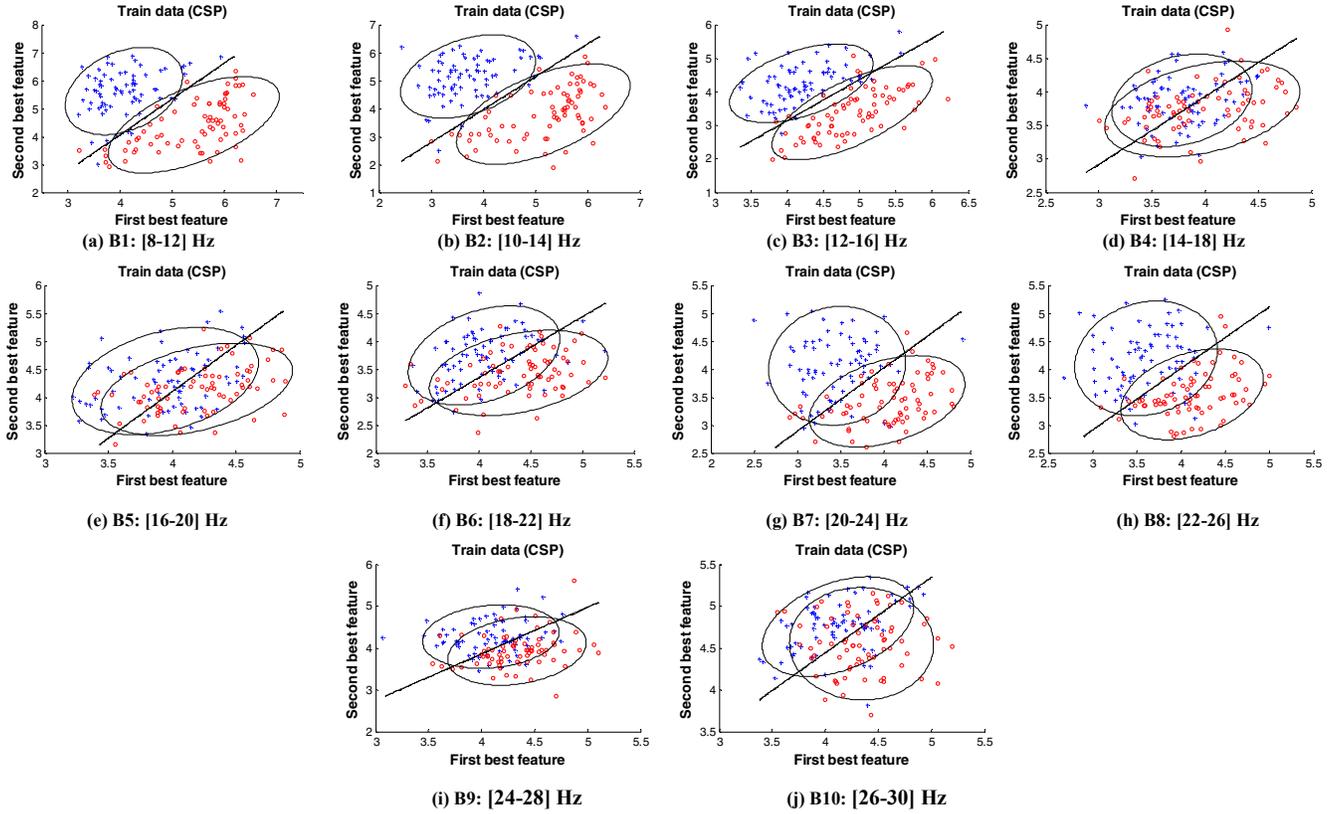


Fig. 3. Distribution of the two best features obtained by CSP for Subject A03. The figure number (a-j), represent the CSP features for each band given as follows: B1-B10. The red circle denotes the features of the left hand MI and blue crosses denote the features of the right hand MI. The black line represents the linear separation, for illustration purpose only.

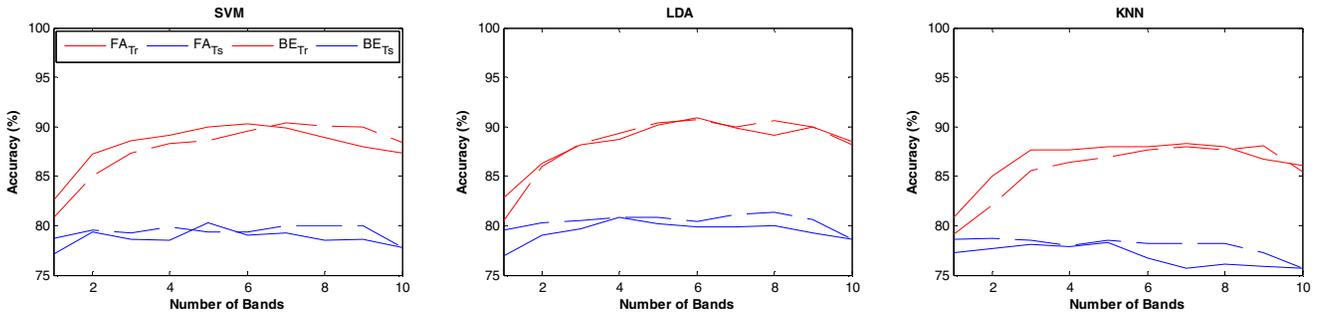


Fig. 4: Evolution of the accuracy as the function of the number of frequency bands for SVM, LDA, and KNN, for forward-addition (FA) and Backward-Elimination (BE), for training (Tr) and the test (Ts).

2) Spatial Filtering

The second stage employs a spatial filter that maximizes the variance of band-pass filtered signals under one condition, while minimizing it for the other condition. At this step all the filtered data from the frequency bands [B1-B10] are spatially filtered using a CSP technique, which maximizes the difference between two classes. Fig. 3 shows the features obtained by the CSP technique for the subject A03. Each of the sub-figures from (a)-(j), represents a CSP feature corresponding to a frequency band. The blue crosses and red circles denote the features of the left hand and right hand motor imagery, respectively. The black line represents the separation plane between the features of two classes

obtained from each frequency band, this separation plane is plotted for illustration purpose only.

C. Classifier and Evaluation

In order to evaluate the performance of the method, we have considered the classification accuracy as the measure of index. The classification accuracy is given in percentage (%). The experiments are performed using support vector machine (SVM), linear discriminate analysis (LDA) and K-nearest neighbor (KNN) based pattern classifiers. The classification results using FA and BE has been obtained. Moreover, band combinations obtained from the intersection of FA & BE (i.e. $FA \cap BE$), and union of FA & BE (i.e. $FA \cup BE$) are also evaluated. In the intersection, the common frequency bands from each subject for both FA and BE are used to evaluate the

TABLE I
FREQUENCY BANDS SELECTED USING FA FOR SVM CLASSIFIER

Subjects	BANDS
A01	B02, B03, B04, B05, B06, B07, B09, B10
A02	B01, B02, B03, B04, B07, B10
A03	B02, B03, B04, B06
A04	B01, B02, B03, B04, B06, B07, B08, B09, B10
A05	B02, B03, B04, B05, B09, B10
A06	B02, B03, B04, B07, B09, B10
A07	B01, B03, B06, B08, B10
A08	B01, B02, B05, B07, B09
A09	B01, B02, B04, B05, B06, B09

TABLE II
FREQUENCY BANDS SELECTED USING BE FOR SVM CLASSIFIER

Subjects	BANDS
A01	B01, B02, B04, B05, B06, B08, B09, B10
A02	B01, B03, B05, B07, B08, B10
A03	B02, B03, B04, B06, B07, B08, B09, B10
A04	B01, B02, B04, B05, B06, B07, B08, B09, B10
A05	B02, B04, B05, B07, B08, B09, B10
A06	B01, B02, B03, B04, B05, B06, B07, B08, B10
A07	B01, B02, B03, B04, B05, B06, B07, B09, B10
A08	B01, B02, B03, B05, B06, B07, B08, B10
A09	B02, B04, B05, B06

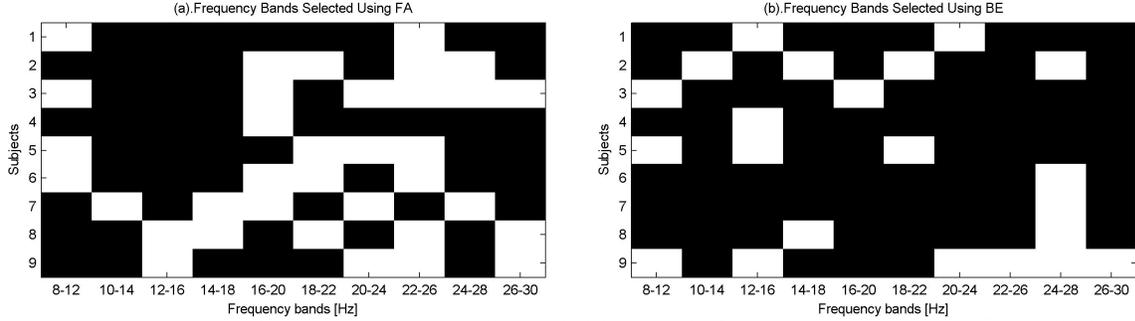


Fig. 5. The images illustrate the frequency bands selected using the (a) FA and (b) BE algorithm using SVM classifier, the black region represents the selected frequency bands.

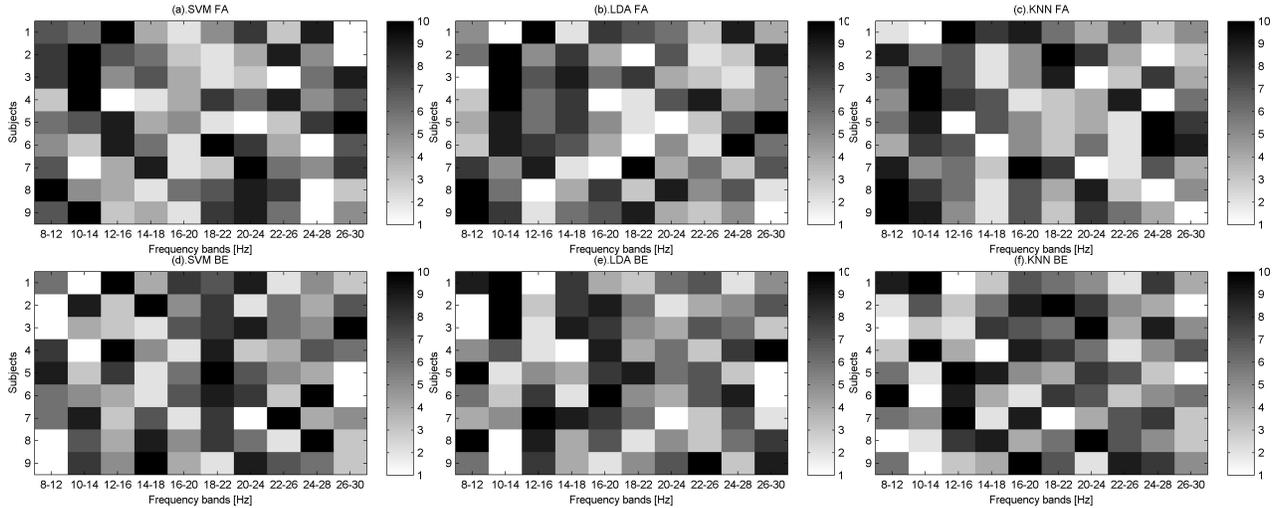


Fig. 6. Representation of the subject's specific band that lead to better classification of the motor imagery responses. A dark shade indicates that a band is useful to include in the feature set, and a light shade indicates that the band has a less discriminant power for the classification.

classification accuracy. In union, all distinct bands from both FA and BE are used to evaluate the classification accuracy. The proposed method is compared with traditional FBCSP algorithm [22]. A two-sided Wilcoxon signed rank test is used to assess the statistical significance of the improvement at a confidence level of 0.05.

The BCI performance over the FA and BE has been monitored for all the subjects on the training and test data, and for each classifier the mean accuracy as a function of the number of operating frequency bands are plotted. Moreover, gray scale images are also used to depict the usefulness of the

subject-specific frequency bands for both FA and BE algorithms.

VI. RESULTS

The subject-specific optimal frequency bands for each subject obtained from the FA and BE are given in Table I and II, respectively. Additionally, in Fig. 5, a black and white images are used to represent the bands selected for the FA and BE algorithms, respectively using the SVM classifier. The black shade represents the selected band, and the white shade depicts not selected band. Next, the results produced by the

TABLE III
CLASSIFICATION ACCURACY (%) USING SVM CLASSIFIER

Subjects	FBCSP	FA	BE	FA \cap BE	FA \cup BE
A01	90.97	93.06	94.44	84.72	90.28
A02	61.11	63.19	65.28	61.11	61.11
A03	90.97	93.06	94.44	93.06	87.50
A04	69.44	69.44	70.14	68.06	69.44
A05	68.11	70.83	70.13	54.17	63.19
A06	65.28	65.28	64.58	65.97	65.28
A07	68.75	65.28	68.06	63.89	68.75
A08	90.25	92.36	93.75	91.67	93.75
A09	88.65	89.58	90.97	90.97	92.36
<i>Mean</i>	77.06	78.01	79.09	74.85	76.85
<i>Std</i>	12.73	13.51	13.74	15.14	13.73
<i>p-Value</i>		0.296	0.023	0.546	1

TABLE IV
CLASSIFICATION ACCURACY (%) USING LDA CLASSIFIER

Subjects	FBCSP	FA	BE	FA \cap BE	FA \cup BE
A01	90.97	90.28	90.28	79.86	90.28
A02	63.19	63.19	63.19	61.81	63.19
A03	90.97	96.53	93.75	95.83	93.75
A04	69.44	69.44	70.14	70.14	69.44
A05	70.14	70.83	72.92	52.78	63.19
A06	68.06	70.14	65.97	63.89	68.06
A07	72.92	69.44	75.00	61.81	72.92
A08	90.20	90.97	91.67	94.44	93.75
A09	88.89	92.36	92.36	91.67	92.36
<i>Mean</i>	78.31	79.24	79.48	74.69	78.55
<i>Std</i>	11.63	12.91	12.41	16.21	13.63
<i>p-Value</i>		0.312	0.101	0.410	0.812

TABLE V
CLASSIFICATION ACCURACY (%) USING KNN CLASSIFIER

Subjects	FBCSP	FA	BE	FA \cap BE	FA \cup BE
A01	90.28	89.58	90.97	77.78	88.89
A02	59.72	52.08	57.64	52.78	59.72
A03	93.06	90.97	93.75	95.14	94.44
A04	69.44	63.89	69.44	68.75	69.44
A05	64.58	65.28	63.89	48.61	59.72
A06	60.42	61.11	56.25	56.94	60.42
A07	60.42	71.53	63.89	62.50	64.58
A08	90.06	93.06	92.36	93.06	93.06
A09	88.89	90.28	90.97	90.28	90.97
<i>Mean</i>	75.21	75.31	75.46	71.76	75.69
<i>Std</i>	14.90	15.71	16.17	17.99	15.68
<i>p-Value</i>		1	0.679	0.347	0.687

proposed method are compared against the FBCSP method on the BCI competition IV dataset 2A are given in Table III, IV, and V for the SVM, LDA, and KNN, respectively. For each table, the mean and standard deviation are given to compare the performance.

To monitor the accuracy of the system versus the number of operating frequency bands, the mean accuracy as a function of the number of operating frequency bands for the FA and BE over training and test data are plotted in the Fig 4. For all the three classifiers, the figures clearly show that the curves at the starting position have low classification accuracies when the number of operating frequency bands are less and it gradually increases as the number of frequency band increases, and then it start decreasing. These plots clearly justify the purpose of the study for choosing the subject-specific optimal frequency

bands. It is clear that using the large number of frequency bands does not always result into high accuracy. Moreover, the greyscale images in Fig 5. (a-f) represent weights of the subject-specific optimal frequency bands for both the FA and BE algorithms. The darkest gray shaded regions correspond to the highest weight and have more importance than the white shaded region, and any fractional value in between. These figures clearly state that choosing the subject-specific frequency bands is a better choice and it may increase the accuracy. Hence, the FA and BE algorithms are proposed to be used to select the subject-specific optimal frequency bands and the results are given below.

In Table III for the SVM classifier, the results for the FA and BE have shown an improvement over traditional FBCSP approach, wherein the mean accuracy of the FA has increased from 77.06% to 78.01% and the mean accuracy for BE has increase to 79.09%, with a p-value of 0.023. However, the intersection and union of the bands from the FA and BE methods have failed to improve the performance.

In Table IV for the LDA classifier, the results for the FA and BE have shown an improvement over traditional FBCSP approach, where the mean value for FA accuracy is increased from 78.31% to 79.24% and for BE the mean value is increased to 79.48%. The intersection of (FA \cap BE) has not shown any improvement, whereas the union (FA \cup BE) has shown a slight improvement in the mean accuracy from 78.31% to 78.55%. However, all the algorithms for the LDA failed to reject the null hypothesis.

In Table V for the KNN classifier, the results for the FA and BE have shown an improvement over the traditional FBCSP approach, wherein the mean value for FA accuracy has increased from 75.21% to 75.31% and for the BE, mean value is increased to 75.46%. The intersection has not shown an improvement, whereas the union (FA \cup BE) has shown a slight improvement in the mean accuracy from 75.21% to 75.69%. However, all the methods for the KNN failed to reject the null hypothesis.

VII. DISCUSSION

The experimental results demonstrated the effectiveness of the FA and BE algorithms. The results showed that the features selected using proposed method have outperformed the traditional FBCSP algorithm, although the obtained bands were not guaranteed to be a globally optimal solution.

In particular, the plots given in Fig. 4 for the mean accuracy as a function of the number of operating frequency bands clearly demonstrate that using all frequency bands from the filter-bank is not the best option. Next, it is important to note that performance plots from all the classifiers depict that initially when the number of frequency bands are less, the accuracy is low and as the number of frequency band increases, the accuracy starts increasing gradually to a certain limit, and after that it starts decreasing steadily. Using both the FA and BE algorithms, the optimal bands are in the middle of the B1 to B10.

Additionally, to view the subject-specific importance of the each frequency band, the gray scale images are used. The gray scale images depict the usefulness of the subject-specific frequency bands for both FA and BE algorithms. For both the FA and BE algorithms, the bands around μ rhythm (i.e., [8-13]

Hz) have much more importance as dark shaded regions appear for most of the subjects with the SVM classifier. Similar type of behavior can also be seen for other classifiers (i.e., LDA and KNN).

Regarding the computational complexity, the time consuming part of the methods is to find the optimal subject specific frequency bands, by computing the 10-fold cross-validation training accuracy. The additional fact is that when the number of EEG channels is increased, the FA and BE methods require more time to find out the optimal filters. Next, the numbers of filters have an adverse effect on the time complexity, as the number of filter increases in the filter bank, the time complexity will also increase.

VIII. CONCLUSION

In this paper, forward-addition (FA) and backward-elimination (BE) algorithms are compared in the band selection task of a filter bank CSP approach to BCI design. The FA and BE employ an iterative feature selection approach to select discriminative CSP features from a bank of multiple band-pass filters and spatial filters, and a classification algorithm to classify the selected features. The proposed method addresses the problem of selecting an appropriate subject-specific operational frequency bands for extracting discriminating CSP features. It is shown to be capable of learning subject-specific patterns from the high-dimensional EEG measurements and yields relatively high classification accuracies. The results clearly support the conclusion that band selection has a significant impact on the performance of a filter bank CSP based BCI. Further investigation using finer frequency ranges may offer scope for further performance improvement.

ACKNOWLEDGEMENT

H.R. was supported by Ulster University Vice-Chancellor's research scholarship (VCRS). G.P. and H.C. were supported by the Northern Ireland Functional Brain Mapping Facility project (1303/101154803), funded by InvestNI and the University of Ulster. G.P. and H.R. were also supported by the UKIERI DST Thematic Partnership project "A BCI operated hand exoskeleton based neurorehabilitation system" (UKIERI-DST-2013-14/126).

REFERENCES

- [1] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A Review of Classification Algorithms for EEG-Based Brain-Computer Interfaces," *J. Neural Eng.*, vol. 4, no. 2, pp. 1–13, Jun. 2007.
- [2] K. Müller, M. Krauledat, and G. Dornhege, "Machine learning techniques for brain-computer interfaces," *J. Biomed. Eng.*, vol. 49, pp. 11–22, 2004.
- [3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-Computer Interfaces for Communication and Control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–91, Jun. 2002.
- [4] B. Blankertz, G. Curio, K. Müller, N. Group, and K. B. Franklin, "Classifying Single Trial EEG: Towards Brain Computer Interfacing," in *Advances in Neural Information Processing Systems*, 2002, pp. 157–164.
- [5] M. Arvaneh, G. Cuntai, K. Kai, and Q. Chai, "Optimizing Spatial Filters by Minimizing Within-Class Dissimilarities in Electroencephalogram-Based Brain-Computer Interface," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 24, no. 4, pp. 610–619, 2013.
- [6] Y. Li, H. Kambara, Y. Koike, and M. Sugiyama, "Application of Covariate Shift Adaptation Techniques in Brain-Computer Interfaces," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 6, pp. 1318–24, Jun. 2010.
- [7] H. Raza, G. Prasad, Y. Li, and H. Cecotti, "Covariate shift-adaptation using a transductive learning model for handling non-stationarity in EEG based brain-computer interfaces," in *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2014, pp. 230–236.
- [8] H. Raza, G. Prasad, Y. Li, and H. Cecotti, "Toward Transductive Learning Classifiers for Non-Stationary EEG," in *Engineering in Medicine and Biology Society (EMBC), 2014 35th Annual International Conference of the IEEE*, 2014.
- [9] H. Raza, G. Prasad, and Y. Li, "Dataset Shift Detection in Non-stationary Environments Using EWMA Charts," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2013, pp. 3151–3156.
- [10] H. Raza, G. Prasad, and Y. Li, "EWMA Based Two-Stage Dataset Shift-Detection in Non-stationary Environments," in *Artificial Intelligence Applications and Innovations (AIAI)*, 2013, pp. 625–635.
- [11] H. Raza, G. Prasad, and Y. Li, "EWMA model based shift-detection methods for detecting covariate shifts in non-stationary environments," *Pattern Recognit.*, vol. 48, no. 3, pp. 659–669, Aug. 2015.
- [12] H. Raza, G. Prasad, and Y. Li, "Adaptive learning with covariate shift-detection for non-stationary environments," in *IEEE 14th UK Workshop on Computational Intelligence (UKCI)*, 2014, pp. 1–8.
- [13] B. Blankertz and R. Tomioka, "Optimizing spatial filters for robust EEG single-trial analysis," *Signal Processing Magazine, IEEE*, pp. 41–56, 2008.
- [14] D. Coyle, G. Prasad, and T. M. McGinnity, "Faster Self-Organizing Fuzzy Neural Network Training and A Hyperparameter Analysis for A Brain-Computer Interface," *IEEE Trans. Syst. Man, Cybern.*, vol. 39, no. 6, pp. 1458–71, Dec. 2009.
- [15] C. Vidaurre, R. Cabeza, R. Scherer, and G. Pfurtscheller, "A Fully On-line Adaptive BCI," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1214–1219, 2006.
- [16] S. Shahid and G. Prasad, "Bispectrum-based feature extraction technique for devising a practical brain-computer interface," *J. Neural Eng.*, vol. 8, no. 2, p. 025014, Apr. 2011.
- [17] B. Rivet, H. Cecotti, E. Maby, and J. Mattout, "Impact of spatial filters during sensor selection in a visual P300 brain-computer interface," *Brain Topogr.*, vol. 25, pp. 55–63, 2012.
- [18] H. Cecotti, B. Rivet, M. Congedo, C. Jutten, O. Bertrand, E. Maby, and J. Mattout, "A robust sensor-selection method for P300 brain-computer interfaces," *J. Neural Eng.*, vol. 8, p. 016001, 2011.
- [19] B. Rivet, H. Cecotti, R. Phlypo, O. Bertrand, E. Maby, and J. Mattout, "EEG sensor selection by sparse spatial filtering in P300 speller brain-computer interface," in *Engineering in Medicine and Biology Society (EMBC), 30th Annual International Conference of the IEEE*, 2010, pp. 5379–5382.
- [20] H. Cecotti, B. Rivet, M. Congedo, C. Jutten, O. Bertrand, E. Maby, and J. Mattout, "Suboptimal sensor subset evaluation in a p300 brain-computer interface," in *18th European Signal Processing Conference (EUSIPCO-2010)*, pp. 1–5.
- [21] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter Bank Common Spatial Pattern (FBCSP)," in *Proc. Int'l Joint Conf. Neural Networks (IJCNN)*, 2008, pp. 2390–2397.
- [22] 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," *Front. Neurosci.*, vol. 6, p. 39, Jan. 2012.
- [23] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller, "Towards Adaptive Classification for BCI," *J. Neural Eng.*, vol. 3, no. 1, pp. R13–23, Mar. 2006.
- [24] S. R. Liyanage, C. Guan, H. Zhang, K. K. Ang, J. Xu, and T. H. Lee, "Dynamically Weighted Ensemble Classification for Non-Stationary EEG Processing," *J. Neural Eng.*, vol. 10, no. 3, p. 036007, Apr. 2013.
- [25] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.
- [26] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the BCI Competition IV," *Front. Neurosci.*, vol. 6, p. 55, Jan. 2012.