Common Spatial Pattern with Feature Scaling (FSc-CSP) for Motor Imagery Classification

Yohanes de Britto Hertyasta Prathama[1(✉)], Mohd Ibrahim Shapiai[1,2],
Siti Armiza Mohd Aris[3], Zuwairie Ibrahim[4], Jafreezal Jaafar[5],
and Hilman Fauzi[1,6]

[1] Department of Electronic Systems Engineering, Malaysia-Japan International
Institute of Technology, Universiti Teknologi Malaysia, Jalan Sultan Yahya
Petra, 54100 Kuala Lumpur, WP Kuala Lumpur, Malaysia
hertyasta.prathama@gmail.com
[2] Centre for Artificial Intelligence and Robotics, Universiti Teknologi Malaysia,
Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, WP Kuala Lumpur, Malaysia
[3] Razak School of Engineering and Advanced Technology,
Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur,
WP Kuala Lumpur, Malaysia
[4] Faculty of Electrical and Electronic Engineering,
Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia
[5] Department of Computer and Information Sciences,
Faculty of Science and Information Technology,
Universiti Teknologi Petronas, Seri Iskandar, Malaysia
[6] Faculty of Electrical Engineering, Telkom University, Bandung, Indonesia

Abstract. Brain-Computer Interface (BCI) is a way to translate human thoughts into computer commands. One of the most popular BCI type is Electroencephalography (EEG)-based BCI, where motor imagery is considered one of the most effective ways. Previously, to extract useful information, various filters are introduced, such as spatial, temporal, and spectral filtering. A spatial filtering algorithm called Common Spatial Pattern (CSP) was developed and known to have excellent performance, especially in motor imagery for BCI application. In general, there are several approaches in improving CSP such as regularization approach, analytic approach, and frequency band selection. In general, the existing techniques for band selection is either to select or reject the band by ignoring the importance of the band. For example, Binary Particle Search Optimization Common Spatial Pattern (BPSO-CSP) was proposed to choose multiple possible best bands to be used in processing the data. In this paper, we propose an algorithm called Feature Scaling Common Spatial Pattern (FSc-CSP) to overcome the problem of feature selection. Instead of selecting features, the proposed algorithm employs a feature scaling system to scale the importance of each band by using Genetic Algorithm (GA) altogether with Extreme Learning Machine (ELM) as classifier, with 1 signifying the most important bands, declining until 0 for the unused bands, as opposed to the 1 and 0 selection system used in BPSO-CSP. Conducted experiments show that by employing feature scaling, better results can be achieved especially compared to vanilla CSP and feature selection with 100 hidden nodes in three from five BCI Competition III datasets IVa, namely *aa*, *aw* and *ay*, with around 5–8% better results compared to vanilla CSP and feature selection.

Keywords: EEG · Common Spatial Pattern · Motor imagery · Genetic Algorithm · Feature scaling

1 Introduction

Brain-computer interface (BCI) is a system that translates human thoughts into commands, which enables humans to interact with their surroundings, without the involvement of peripheral muscles and nerves. BCI based on electroencephalography (EEG) uses control signals from the EEG activity as input/commands. BCI creates an alternative non-muscular pathway for relaying a person's intentions to external devices such as computers, neural prostheses, and other different assistive peripherals [1]. Generally, BCI could be divided into three types; invasive, partially invasive, and non-invasive. Non-invasive BCI is the most risk-free, subject-wise, type of BCI, because it doesn't involve any surgery to be implemented to the subject. However, it has one trade-off: intervention from the scalp and skull decreases the resolution and provides noise to the brain signal, compared to the invasive and partially-invasive BCI. Some of the most popular modalities are Magnetic Resonance Imaging (MRI), Magnetoencephalography (MEG), and Electroencephalography (EEG).

Despite its worst spatial resolution compared to the other types of BCI, BCI through EEG is one of the most modern modality and most studied field of BCI, mainly because of its non-invasive, portable, and low-cost nature. But BCI through EEG is not without its weakness; it is susceptible to noises and artifacts, mostly from unwanted movements such as eye blinks, or poor contact surface [2]. This problem mostly solved by trying to reduce artifacts and noises by utilizing reference electrodes placed in locations where there is little cortical activity and attempting to filter out correlated patterns [3, 4]. Other precautions are done by applying filtering and defining the right information to the EEG signals [5].

EEG architecture basically can be divided into four blocks; signal acquisition, preprocessing, feature extraction, and classification. One of the most proponent algorithm of BCI feature extraction is Common Spatial Pattern (CSP) [6]. This algorithm focuses in maximizing the variance ratio of two classes of an EEG data [8]. However, CSP is not without its limitations; CSP needs a fine-tuning of its filtering band for each subject because it can only process one band and solely designed for two-classes use.

Several approaches have been done to overcome the limitations of CSP, which can be categorized into five approaches: Improvement of estimation robustness [6, 7], regularization [9, 10], multi-subject implementation [10, 11], kernel-based algorithm [13, 14], and frequency band selection [14, 15]. There is an algorithm proposed to address the limitation of CSP which belongs to the frequency band selection approach called Binary Particle Swarm Optimization Common Spatial Pattern (BPSO-CSP) [16]. BPSO-CSP can use features from multiple bands it deems important, through the usage of feature selection. With this algorithm, the important band is defined as 1 and the unimportant or unused band is defined as 0. The classifier will then only process the bands marked as 1. However, despite its ability to solve the fine-tuning problem of CSP, BPSO-CSP only uses the selected bands and rejects the rest, despite the probability of usable information within the rejected band.

In this paper, we propose an alternative approach to BPSO-CSP, especially to overcome the selection problem of both algorithms. Instead of selecting or rejecting the feature, we scale the feature through importance of it between zero and one by using Genetic Algorithm (GA). Through this approach, the bands with less significance will have a chance to be processed due to the probability of having several usable information within them.

This manuscript is divided into part 1 as introduction, part 2 to present works related to this paper, part 3 to explain the research methodology, part 4 to present the findings and discussion, and part 5 serves as conclusion. Introduction explains about the issue of BCI through EEG, and more specifically, the potential of CSP-derived algorithms for its significance. And then, the related works discusses about the previously done studies in CSP, especially those with the significance regarding the topics of multi-band-related CSP. In the methodology part, the dataset used and the systematic methodology is explained. In the findings and discussion part, the result of experiment is presented, and by that result, opens a discussion on possible future improvements and current milestone on improvements. Finally, the conclusion summarizes the point of the whole proposal.

2 Related Works

2.1 Common Spatial Pattern

CSP produces a set of spatial filters that can be used to decompose multi-dimensional data into a set of uncorrelated components [6]. CSP maximizes the variance-ratio of two conditions or classes. In other words, CSP finds a transformation basis that maximizes the variances of one multi-channel signal and simultaneously minimizes the variances of the other multi-channel signal. For instance, CSP for BCI is employed to distinguish between the variance or power of the associated left-hand and right-hand motor imagery classification from brain signal. This features particularly useful tool for the discrimination of EEG data obtained during different mental states in BCI system. In general, CSP is considered among the best performing algorithm for BCI community. CSP is categorized as supervised learning method as it requires not only the training samples but also the information of the classes of the signal, e.g. left-hand signal.

CSP is an algorithm that maximizes the variance of one class and minimize the variance of other class to discriminate between two classes of EEG data. The multi-channel EEG signal separated into two mental tasks (A and B) can be described as two spatiotemporal signal matrices of X_A and X_B, with dimension ($N \times T$), where N signifies few number of channels and T signifies few number of samples. Then, the covariance matrices of these signals are calculated by

$$R_n = \frac{X_n X_n^t}{trace\left(X_n X_n^t\right)} \tag{1}$$

where n signifies the respective classes (A or B), M^t signifies the transpose of the matrix M, and *trace(M)* signifies the sum of the diagonal elements of matrix M.

The matrices X_A and X_B contain different mental task records; however, they share the same condition and therefore can be modeled as follows:

$$X_n = [C_n C_c]\begin{bmatrix} S_n \\ S_c \end{bmatrix} \tag{2}$$

where S_n and C_n signifies the specific source component and corresponding spatial pattern for each mental task, while S_c and C_c signifies the source component and spatial pattern for the common condition.

CSP algorithm is created with the purpose of designing two spatial filters, so the source component S_n can be extracted with

$$S_n = F_n X_n \tag{3}$$

where F_n refers to the spatial filters corresponding to each task. From these information, CSP applies principal component and spatial subspace analysis to the diagonalized covariance matrices using training data to estimate the spatial filters. A more detailed explanation and calculation of CSP can be found at its corresponding research [6].

CSP is known to be yielding bad results when applied to unfiltered EEG data or EEG data filtered with poor frequency band [15], which, in turn, leads a different problem of the need of fine-tuning in order to achieve optimum results. There are several CSP derived methods which have been proposed in the literature to deal with various limitations of CSP. It is generally divided into binary-class [11, 15, 17–19] and multi-class [20, 21]. Binary-class, as its name states, solves the problem of CSP with its own basic characteristics of maximizing and minimizing two covariance matrices for two classes, meanwhile the multi-class tries to implement CSP to data with more than two classes.

The binary-class method can then be divided into five means, which specifies the solution. The first means is to enhance the estimation of the robustness [7], which also faces the inherent fine-tuning problem of CSP; the second one is applying regularization [9, 10], which has a susceptible selection of regularization parameter which may result in underfitting or overfitting of the solution; the third one is through addressing a multi-subject problem [11, 12], which causes differentiation of the acquired brain signal from multiple subject may differ much, leading to irreproducible results; the fourth one is through the kernel-based solution [13, 14], which subjects the result to the kernel selection and the parameter; and the last category is a multi-band solution [15, 16], which, mostly, suffers from information negligence due to the nature of band selection methods.

2.2   Binary Particle Swarm Optimization Common Spatial Pattern

BPSO-CSP is an improvement of CSP algorithm proposed by [16]. This methodology employs Binary Particle Swarm Optimization (BPSO) [22] to select the suitable bands

to process the brain signals. This algorithm selects the bands to be used according to acceptance and rejection of each band by BPSO.

The methodology of BPSO-CSP can be concluded into 4 steps: Initialization of each particles which correspond with a vector of 1 s and 0 s, where 1 means accepted band and 0 means rejected band; computing fitness values for each particle which is derived from the classification accuracy of each sub-band; updating the velocity and position of the iteration based on the previous best velocity and position (given that there has been a previous iteration); and mutation, in which a mutation operator is employed in order to get the optimal points out of the BPSO algorithm. The steps are repeated from updating until a certain threshold of maximum iteration which is pre-defined before is reached. In this paper, for comparison, we employ a similar algorithm, which is called Feature Selection. The procedure is shown in Fig. 1.
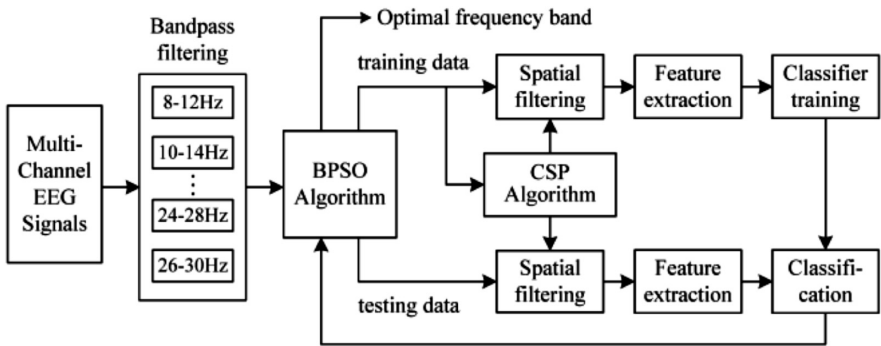

Fig. 1. BPSO-CSP framework [16].

3   Methodology

In brief, our proposed methodology of FSc-CSP is shown in Fig. 2. The data-set was preprocessed with Butterworth filter with sliding windows of 17 bands, next processed by CSP at respective bands, and then scaled by GA-ELM, which finally passed onto classifier for testing. Extreme Learning Machine (ELM) will be used as the classifier.
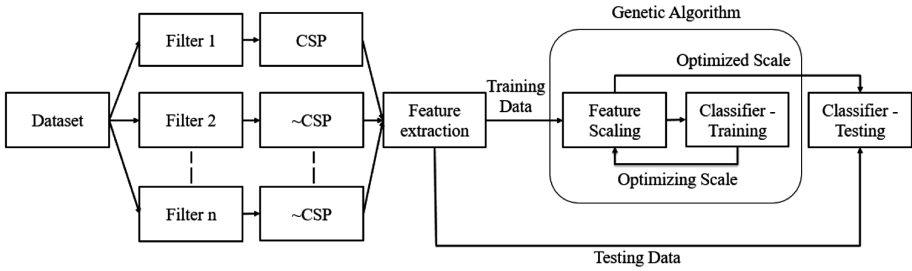
Fig. 2. FSc-CSP framework.

3.1 Preprocessing

The data were first preprocessed by using sliding window and Butterworth filters. Sliding Window is a method of choosing/splitting a lengthy band into several small bands. It works by splitting an individual length of bands, e.g. 4–40 Hz, into several bands based on windows, determined by variables called width and step. Width means the range of the band, e.g. width of 4 means the band is four values long, minus the starting value (e.g. a window that starts from 4 Hz with a width of 4 will have a length of 4–8 Hz). Step means the margin between the starting point of one window to the other, e.g. with a step of 2 and width of 4, with the band ranging from 4–12 Hz, the windows will consist of 4–8 Hz, 6–10 Hz, and 8–12 Hz. In this experiment, the actual length of the band is 4–40 Hz, with the width of 4 and step of 2, and 17 bands in total. This range encompasses the theta, alpha, mu and beta bands.

The Butterworth filter is a type of signal processing filter designed to have as flat frequency response as possible in the passband. In this experiment, third-order band-pass filter is used, meaning that the signal is filtered only at a certain range of length. The value for the low pass (lowest threshold of the length) and high pass (highest threshold of the length) of the bandpass filter is the value of the sliding window. For example, if the current window is 6–10 Hz, then the low pass is 6 Hz and the high pass is 10 Hz. The example representation of sliding window can be seen in Fig. 3 below.
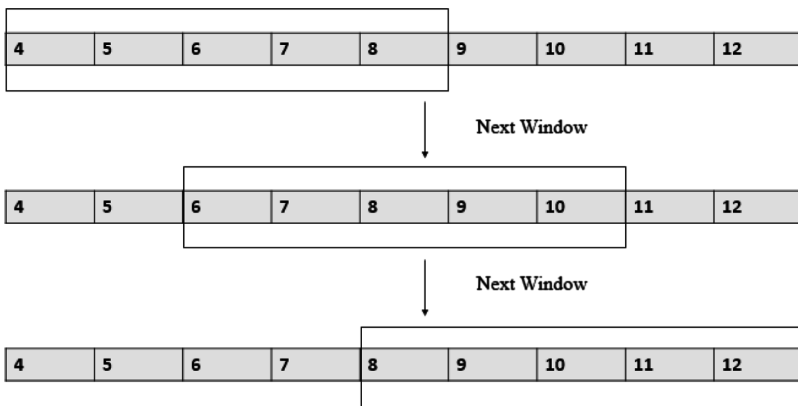

Fig. 3. Example representation of sliding window.

3.2   Feature Scaling for Proposed Framework

GA is an optimization method process that tries to mimic biological evolution. The algorithm creates a population of random solutions for the function, and at each generation, the algorithm is aimed to move toward an optimal solution of the function. The flowchart can be seen in Fig. 4. In this FSc-CSP algorithm, GA is used to find an optimized scale for the classifier. The scale signifies the importance of the combination; The higher the classifier accuracy with that scale, more important that scale is. The number of the scale is the same as the number of bands, and then the scale is replicated for each feature in that band.

As pictured in Fig. 4, the GA will first initialize a set of populations of the scale, and then these scales will be applied to the ELM classifier. And then GA will try to crossover all the members of the initial sets to produce children scales, in which will be passed again to the ELM classifier. Throughout the iterations, GA will try to optimize the scale so that the classifier will reach the best result. The scale that gives zero or closest to zero error percentage during training with ELM compared with other scale sets is the one that is selected as the best scale for the data. As for the process in ELM itself can be seen in Fig. 5.
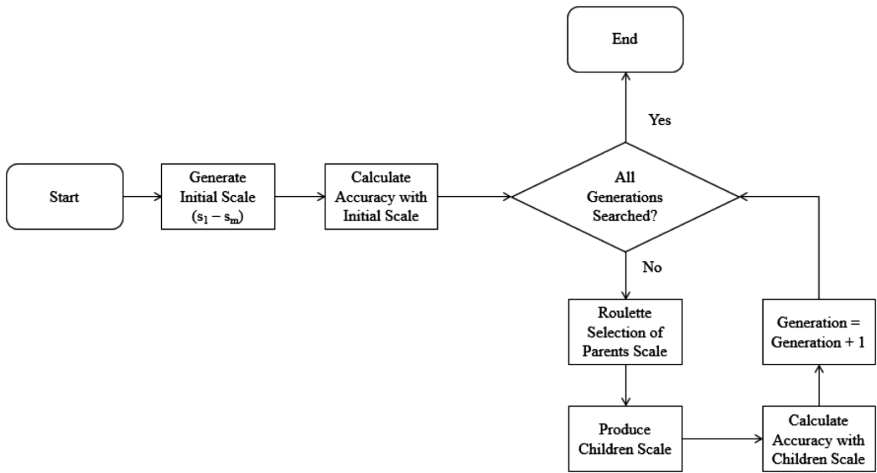

Fig. 4. Genetic Algorithm flowchart.

In Fig. 5, features encompass all the features from every bands, with m representing the number of bands, and n representing the number of features. s_1 until s_m is the scale the GA is trying to find in order to reach the optimal scale combination, which is valued between zero to one depending on the importance of the band, and replicated for the number of n for each m. For example, if there are 17 bands and each band has 8 features, then the scale will be replicated 8 times for each band. h represents the number of the predetermined hidden nodes. The criteria for the optimal scale combination tuning is based on the training accuracy of a certain scale combination. The combination of the scaling which yields the best training accuracy will then be passed into testing to validate the result of the algorithm.
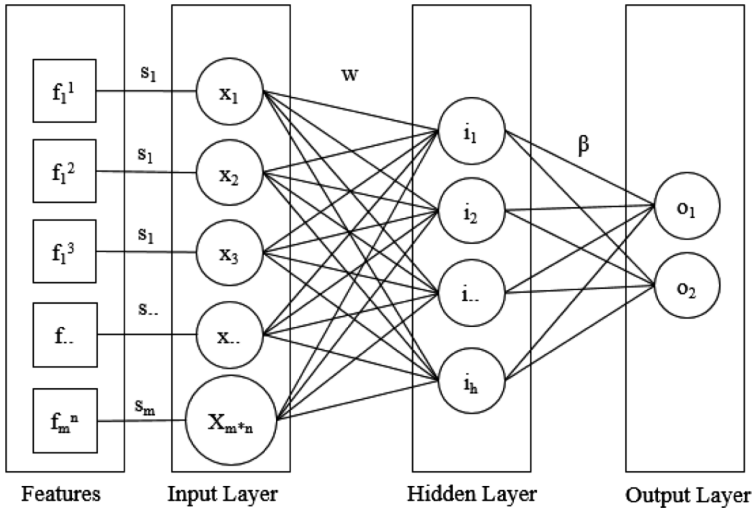
Fig. 5. Extreme learning machine.

4 Experiments Setup, Results and Discussion

4.1 Experiments Setup

To test our method, we use dataset IVa *aa*, *al*, *av*, *aw*, and *ay* from BCI Competition III [23]. For the Genetic Algorithm, we use 50 Generations and 50 Populations. The band is split into 17 sub-bands, ranging from 4–40 Hz, with a width of 4 and a step of 2. We generate ten random samples from each dataset, and split them as follows: 60% of the sample is used as the training sample and 40% of the sample is used as the testing sample. Detailed explanation can be seen in Tables 1 and 2.

Table 1. Dataset partitioning with 60:40 ratio.

Dataset	# of trials	# of class 1 data	# of class 2 data
Subject *aa*	168	80	88
Subject *al*	224	112	112
Subject *av*	84	42	42
Subject *aw*	56	30	26
Subject *ay*	28	18	10

Table 2. Parameter setup for conducted experiment.

Parameter setup	
# of GA generations	50
# of GA populations	50
# of hidden nodes	5/10/68/100
# of runs	10
# of sub-bands	17
# of features	8
# of window steps	2
# of window width	4
Window range	4–40 Hz
Filter type	Third-order butterworth
ELM activation function	Sigmoid
# of input nodes	136 (Bands × Features = 17 × 8)
# of output nodes	2

4.2   Performance Evaluation

The performance of the algorithms is evaluated through the Mean Squared Error (MSE) of the algorithm, which can be represented as

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i' - Y_i\right)^2 \tag{4}$$

where n represents the number of prediction, Y' represents the predictions itself and Y is the predictor. The evaluation is done per hidden nodes basis to attune the underfitting or overfitting problem of neural network.

4.3   Results and Discussion

In this test, we conduct tests using vanilla CSP, with a bandwidth of 8–12 Hz, or the alpha band, and Feature Selection to represent the BPSO-CSP, to give a comparison to our proposed algorithm. Each algorithm is tested using different numbers of hidden nodes for the classifier. The number of hidden nodes is decided randomly, except for 68, which is half of the total number of training and testing data of dataset *aa*. ELM is used as the classifier. 10-fold cross validation is used to validate each data. These tests were done in order to validate the result of our proposed algorithm.

Table 3 shows the testing performance of FSc-CSP algorithm. The performance of feature scaling compared to Feature Selection yields some performance improvement for at least three data sets, namely *aa*, *aw* and *ay* for 100 hidden nodes configuration, while for a lower number of hidden nodes, Feature Selection yields better performance, albeit with lower accuracy compared to FSc-CSP with higher hidden nodes.

There is an anomaly, however, with dataset *al* as the only dataset which performance with both algorithms yield not an objectively bad result, but compared to the

Table 3. Testing performance validation of the FSc-CSP.

# of HN	Type	aa	al	av	aw	ay
5	FSc-CSP	**0.52**	0.60	0.48	0.50	0.48
	Feature selection	0.50	0.59	**0.52**	**0.56**	**0.50**
	Vanilla CSP	0.48	**0.71**	0.47	0.45	0.43
10	FSc-CSP	0.48	0.65	**0.49**	0.48	0.48
	Feature selection	0.47	0.60	0.45	**0.52**	0.48
	Vanilla CSP	**0.52**	**0.74**	0.47	0.48	0.48
68	FSc-CSP	**0.53**	0.60	0.53	**0.68**	0.48
	Feature selection	0.49	0.70	0.53	0.57	**0.50**
	Vanilla CSP	0.52	**0.84**	**0.57**	0.52	0.47
100	FSc-CSP	**0.58**	0.64	0.53	**0.61**	**0.55**
	Feature selection	0.53	0.68	**0.55**	0.53	*0.55*
	Vanilla CSP	0.53	**0.82**	0.54	0.52	0.48

Note: **Bold** means the best result for the dataset compared to
other algorithms with respective hidden nodes configuration

vanilla CSP, the performance is much worse, and compared to other datasets, should
we put vanilla CSP results aside and only put FSc-CSP and Feature Selection into
consideration, dataset al yields a reversed result: FSc-CSP yields better performance for
lower hidden nodes, while Feature Selection yields better performance for higher
hidden nodes.

In Fig. 6, we present an example of scale distribution of FSc-CSP for 100 Hidden
Nodes. It is seen that the algorithm can distinct between the most important bands
with the less important bands, evident in the results. Some bands, namely band 11
(24–28 Hz), 12 (26–30 Hz) and 13 (28–30 Hz), are deemed less important due to the
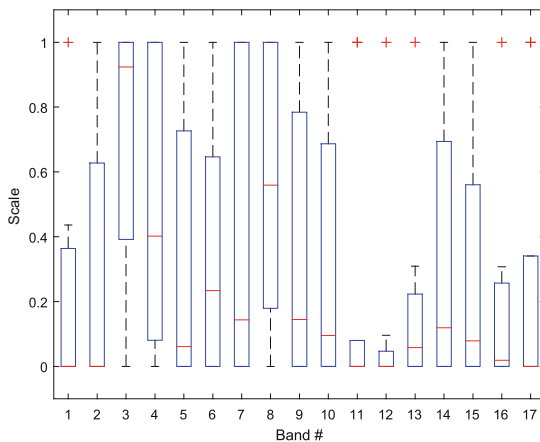

Fig. 6. Dataset *aa* scale distribution for 100 hidden nodes.

range of the scale. The outliers from these bands are occasions where the algorithm deem the band as important enough, however, compared to significant bands like band 3, where the most evident range of importance is equal or above 0.4, these bands seem to contain not so many useful information.

In Fig. 7(a) and (b), we present the power spectrum distribution of two of the most significant bands with the closest range, namely band 3 (8–12 Hz) and band 4 (10–14 Hz). From the figures, it is seen that both bands can distinguish between the hand imagery and foot imagery, which is evident from the gradient difference in both sides. We can also see that the 10–14 Hz band, which was deemed less important by the scale in Fig. 7, albeit able to distinct the signals, the spread of the gradient is less distinct and has more overlapping gradients compared to 8–12 Hz.
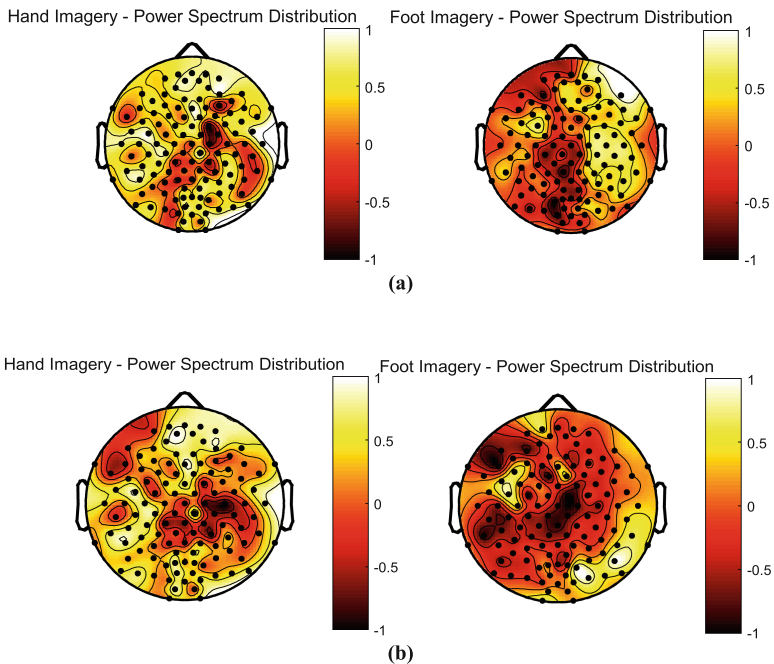

Fig. 7. EEG Power spectrum for (a) band 8–12 Hz (b) band 10–14 Hz.

In Fig. 8(a) and (b), we present the eigenvalue pairing of both bands, and it is seen that the results are in accordance with the scale. 8–12 Hz can distinguish both classes in a good manner, evident in EV Pair 1 and 2 for class 1 and EV Pair 5 for class 2. As for the 10–14 Hz band, it can distinguish for EV Pair 1 and 2 for the first class, but not so much for the second class.

In Fig. 9(a) and (b), the results of CSP filtering are presented for Channel C5 and C6. Both channels were selected because of its position as the center position of both left and right side of the head, thus providing the most representative signal for the
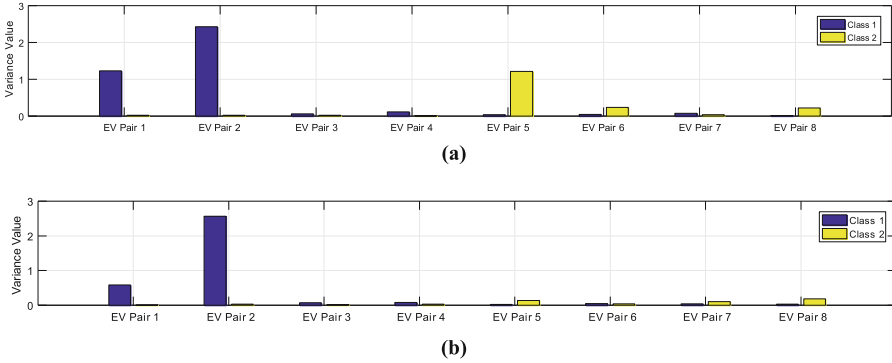
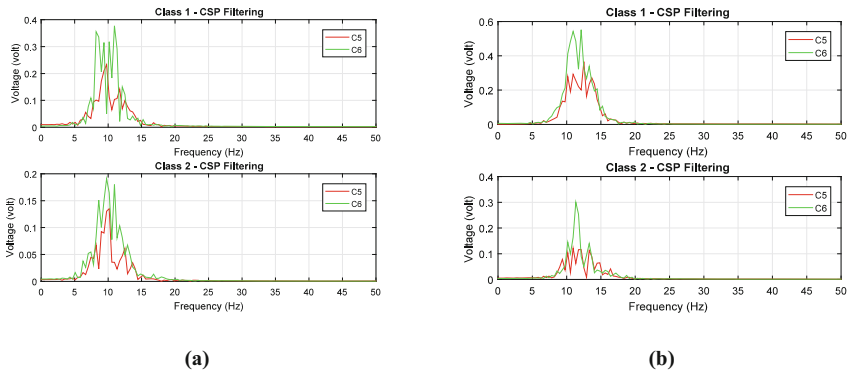**Fig. 8.** Eigenvalue pairing for (a) band 8–12 Hz (b) band 10–14 Hz.

Fig. 9. CSP class distinction for (a) band 8–12 Hz and (b) band 10–14 Hz.

distinction of left and right brain signal. Figure 9(a) shows the result of CSP filtering for 8–12 Hz, which shows good distinction with some spike overlaps in Class 1, but a clean distinction in Class 2. A good distinction can also be seen in Fig. 9(b); however, 10–14 Hz has more overlaps compared to the 8–12 Hz, which means the scale of the features can predict the importance of the bands.

5 Conclusion

In this paper, we present a new algorithm developed to overcome the fine-tuning problem of CSP. The fine-tuning problem is one of the most significant problem faced in BCI research due to its time-consuming nature and its fine-tuned nature causes difficulty for other researchers to replicate the results for benchmarking purposes.

We propose a Feature Scaling Common Spatial Pattern (FSc-CSP) algorithm, which is benchmarked using datasets from BCI Competition III datasets IVa. In the experiment, we also include the vanilla CSP algorithm and From the results, it is seen

that FSc-CSP could provide better result spreads compared to Feature Selection. Feature Selection tends to have better results at a very low hidden nodes number compared to FSc-CSP, whilst FSc-CSP has a better spread across different numbers of hidden nodes compared to Feature Selection, even at the lower ones, as evidenced by the result of dataset *aa*. FSc-CSP is also proven to be able to determine the importance of each band through the scaling system.

The only exception for both algorithm is in dataset *al*, where in all hidden nodes spread, dataset *al* is consistent in providing best result by using only one specific band range, which is 8–12 Hz or the alpha wave.

In the future works, several improvements can be made based on these findings; for example, an improvement of the algorithm through the improvement of the vanilla CSP algorithm itself to overcome the limitation of the vanilla CSP algorithm, or the improvement of feature scaling algorithm to include the probability of using single sub-band, as evidenced by the result of dataset *al* where best results can only be seen by using single sub-band.

References

1. Falzon, O., Camilleri, K.P., Muscat, J.: The analytic common spatial patterns method for EEG-based BCI data. J. Neural Eng. **9**, 45009 (2012)
2. Szafir, D.J.: Non-invasive BCI through EEG (2009)
3. Ludwig, K.A., Miriani, R.M., Langhals, N.B., Joseph, M.D., Anderson, D.J., Kipke, D.R.: Using a common average reference to improve cortical neuron recordings from microelectrode arrays. J. Neurophysiol. **101**, 1679–1689 (2009)
4. Adam, A., Ibrahim, Z., Mokhtar, N., Shapiai, M.I., Mubin, M.: Evaluation of different peak models of eye blink EEG for signal peak detection using artificial neural network. Neural Netw. World. **26**, 67–89 (2016)
5. Adam, A., Shapiai, M.I., Mohd Tumari, M.Z., Mohamad, M.S., Mubin, M.: Feature selection and classifier parameters estimation for EEG signals peak detection using particle swarm optimization. Sci. World J. **2014**, 1–13 (2014)
6. Ramoser, H., Müller-Gerking, J., Pfurtscheller, G.: Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans. Rehabil. Eng. **8**, 441–446 (2000)
7. 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**, 1541–1548 (2005)
8. Asyraf, H., Shapiai, M.I., Setiawan, N.A., Wan Musa, W.S.N.S.: Masking covariance for common spatial pattern as feature extraction. J. Telecommun. Electron. Comput. Eng. **8**, 81–85 (2016)
9. Lu, H., Plataniotis, K.N., Venetsanopoulos, A.N.: Regularized common spatial patterns with generic learning for EEG signal classification. In: Proceedings of 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009, pp. 6599–6602 (2009)

10. Thang, L.Q., Temiyasathit, C.: Regularizing multi-bands common spatial patterns (RMCSP): a data processing method for brain-computer interface. In: International Symposium on Medical Information and Communication Technology (ISMICT), May 2015, pp. 180–184 (2015)
11. Kang, H., Choi, S.: Bayesian multi-task learning for common spatial patterns. In: 2011 International Pattern Recognition in NeuroImaging, vol. 1, pp. 61–64 (2011)
12. Devlaminck, D., Wyns, B., Grosse-Wentrup, M., Otte, G., Santens, P.: Multisubject learning for common spatial patterns in motor-imagery BCI. Comput. Intell. Neurosci. **2011**, 8 (2011)
13. Samek, W., Binder, A., Muller, K.R.: Multiple kernel learning for brain-computer interfacing. In: 2013 Proceedings of Annual International Conference on IEEE Engineering in Medicine and Biology Society, pp. 7048–7051 (2013)
14. Albalawi, H., Song, X.: A study of kernel CSP-based motor imagery brain computer interface classification. In: 2012 IEEE Signal Processing in Medicine and Biology Symposium, SPMB 2012 (2012)
15. Novi, Q., Guan, C., Dat, T.H., Xue, P.: Sub-band common spatial pattern (SBCSP) for brain-computer interface. In: Proceedings of 3rd International IEEE EMBS Conference on Neural Engineering, pp. 204–207 (2007)
16. Wei, Q., Wei, Z.: Binary particle swarm optimization for frequency band selection in motor imagery based brain-computer interfaces. Biomed. Mater. Eng. **26**, S1523–S1532 (2015)
17. Zhang, H., Chin, Z.Y., Ang, K.K., Guan, C., Wang, C.: Optimum spatio-spectral filtering network for brain-computer interface. IEEE Trans. Neural Netw. **22**, 52–63 (2011)
18. Lu, H., Eng, H.L., Guan, C., Plataniotis, K.N., Venetsanopoulos, A.N.: Regularized common spatial pattern with aggregation for EEG classification in small-sample setting. IEEE Trans. Biomed. Eng. **57**, 2936–2946 (2010)
19. Benoît, F., van Heeswijk, M., Miche, Y., Verleysen, M., Lendasse, A.: Feature selection for nonlinear models with extreme learning machines. Neurocomputing **102**, 111–124 (2013)
20. Grosse-Wentrup, M., Buss, M.: Multiclass common spatial patterns and information theoretic feature extraction. IEEE Trans. Biomed. Eng. **55**, 1991–2000 (2008)
21. Gouy-Pailler, C., Congedo, M., Brunner, C., Jutten, C., Pfurtscheller, G.: Multi-class independent common spatial patterns: exploiting energy variations of brain sources. In: 2008 Proceedings of 4th International Brain-Computer Interface Workshop and Training Course, pp. 20–25 (2008)
22. Yuan, X., Nie, H., Su, A., Wang, L., Yuan, Y.: An improved binary particle swarm optimization for unit commitment problem. Expert Syst. Appl. **36**, 8049–8055 (2009)
23. BCI Competition III Dataset IVa. http://www.bbci.de/competition/iii/desc_IVa.html