Title: A Robust and Subject-Specific Sequential Forward Search Method for Effective Channel Selection In Brain Computer Interfaces

Authors: Onder Aydemir, Ebru Ergün

PII: S0165-0270(18)30397-2
DOI: https://doi.org/10.1016/j.jneumeth.2018.12.004
Reference: NSM 8204

To appear in: *Journal of Neuroscience Methods*

Received date: 5 September 2018
Revised date: 5 December 2018
Accepted date: 5 December 2018

A Robust and Subject-Specific Sequential Forward Search Method for Effective Channel Selection In Brain Computer Interfaces

Authors:

1- Onder Aydemir

Department of Electrical and Electronics Engineering

Karadeniz Technical University

61080, Trabzon, Turkey

onderaydemir@ktu.edu.tr

2- Ebru Ergün (Corresponding Author)

Department of Electrical and Electronics Engineering

Recep Tayyip Erdoğan University

53100, Rize, Turkey

ebru.yavuz@erdogan.edu.tr

Highlights:

- We proposed a robust and subject-specific RSS-SFSM for effective channel selection.
- The proposed method was successfully tested on two data sets consisting of 30 subjects in total.
- We improved average classification accuracy by 15.98%.
- The proposed method decreased computational complexity by an average of 71.53%.

Abstract

Background: The input signals of electroencephalography (EEG) based brain computer interfaces (BCI) are extensively acquired from scalp with a multi-channel system. However, multi-channel signals might contain redundant information and increase computational complexity. Furthermore, using only effective channels, rather than all channels, may enhance the performance of the BCI in terms of classification accuracy (CA).

New Method: We proposed a robust and subject-specific sequential forward search method (RSS-SFSM) for effective channel selection (ECS). The ECS procedure executes a sequential search among each of the candidate channels in order to find the channels which maximize the CA performance of the validation set. It should be noted that in order to avoid the problems of random selections in the validation set, we applied the ECS procedure for 100 times. Then, the total numbers of the selection of each channel present the effective ones. To demonstrate its reliability and robustness, the proposed method was applied to two data sets.

Results: The achieved results showed that the proposed method not only improved the average CA by 15.98%, but also decreased the considered number of channels and computational complexity by 71.53% on average.

Comparison with Existing Method(s): Compared with the existing methods, we achieved better results in terms of both the classification accuracy improvement and channel reduction rates.

Conclusions: Features extracted by Hilbert transform and sum derivative methods were effectively classified by support vector machine. In conclusion, the results obtained proved that the RSS-SFSM shows great potential for determining effective channel(s).

Keywords— Brain computer interface, channel selection, feature extraction, classification.

1. Introduction

A brain computer interface (BCI) is a communication platform which translates human intentions reflected by brain signals into control signals for an output device such as a robotic arm, wheelchair or various neuroprostheses (Aydemir and Kayikcioglu, 2014; Aydemir, 2016). Patients who suffer from motor disorders such as amyotrophic lateral sclerosis can use a BCI system to convert their thoughts into actions. Various brain imaging techniques can be used in a BCI system (Lotte et al., 2007). Among these techniques, electroencephalography (EEG) signals are widely used because they not only allow non-invasive signal recording approaches but also offer high temporal resolution.

For the appropriate BCI design, the common spatial pattern (CSP) has generally been used for extraction features (Miao et al., 2016; Kumar et al., 2017; Dong et al., 2016; Zhang et al., 2017), while a support vector machine classifier (Kumar et al., 2016; Kumar et al., 2017; Kumar et al., 2017; Zhang et al., 2018) has been preferred for classification algorithm. Studies have generally improved BCI performance by either developing a new feature extraction technique or a novel classification algorithm. Gaur et al. proposed a novel subject specific multivariate empirical mode decomposition based filtering method to classify the motor imagery based EEG signals into multiple classes. They used spatial covariance matrix features with Riemannian geometry classifier.

Their method achieved a mean Kappa value of 0.60 across nine subjects of the BCI competition IV dataset 2A (Gaur et al., 2018). In another study, Hamzah et al. proposed a method to analyze and classify different EEG patterns based on different motor movements performed by an individual. They used power spectral density values as features and classified them using a neural network method (Hamzah et al., 2016). Although multichannel EEG based BCI systems are preferred to achieve good performance (Yavuz et al., 2017; Tabar et al., 2016), most include unnecessary information and noise. Also, in many cases the number and location of necessary channels have not been connected to a result yet. Therefore, unnecessary channels should be eliminated by improving channel selection methods in order to achieve high performance for BCI system, especially in terms of classification accuracy (CA) and computational complexity.

In literature, a few channel selection methods have been proposed. Updated methods can be found in Torres-Garcia et al. (2016) and Lo et al. (2016). The studies in the literature are generally of two types: subject-independent method or subject-dependent method. While selected effective channels are common to all subject in subject-independent method, effective channels are customized for each individual in subject-dependent method. In addition to such studies, there are also several algorithms that search eliminating redundant channels. For example, Yang et al. applied a genetic algorithm (GA) for determining redundant information in a multichannel electrocorticography (ECOG) dataset which consisted of 64 channels. They used least-square approximation for pre-processing step, and then used multi-layer perceptron-based modeling for classification. While they improved test CA of ECOG based BCI dataset 13% by selecting 10 channels out of 64 (Yang et al., 2012), for the EEG based BCI dataset they provided 2% improvement by selecting 6 channels out of 32. In another BCI based channel selection study, Rayleigh coefficient maximization based GA was proposed in order to detect the effective channels by He et al. (2013). They applied their method on two motor imagery EEG datasets. For first dataset, while they calculated between 59.20-88.30% CAs for 5 subjects using all channels, they obtained between 75.20-98.50% CAs using effective channels. In the second dataset, they achieved between 68.70-90.50% CAs for 6 subjects using all channels, they calculated between 81.30- 97.40% CAs when they used effective channels. In another research, iterative multi-objective optimization was proposed for channel selection to reduce computational complexity (Handiru et al., 2016). They classified features, extracted by modified filter bank common spatial pattern, with a support vector machine (SVM). Finally, while they obtained 63.62% average CA with selected effective channels, they calculated 61.02% average CA using all channels for 85 subjects. In (Ghaemia et al., 2017) a binary gravitation search algorithm (IBGSA) was improved as a channel selection algorithm. First, EEG based BCI signal was filtered with a bandpass filter and then related features were extracted by wavelet transform. Extracted features, which were mean, mod, median, standard deviation and variance of the wavelet transform coefficients, were classified by SVM algorithm. Their results showed that 7 effective channels out of 22 provided CA performance as 76.24% in average. In another study, Qiu et al. proposed an improved sequential floating forward selection (ISFFS) algorithm to select effective channels for the common spatial pattern (Qiu et al., 2016). They tested their method on two data sets and they obtained 16% and 3.8% improvements compared to using all channels. Additionally, Gonzalez et al. (2014) proposed a particle swarm optimization (PSO) based method, which sought to optimize a fitness function product of an aggregation of two performance metrics, including the CA and the number of EEG channels. In the mentioned research, the EEG signals were represented by the temporal features, then the classification process is completed with the Fisher discriminant analysis As a conclusion, they improved the test CA of P300 based BCI dataset by selecting 5 channels out of 64.

In this study, a robust and subject-specific method was proposed by detecting the most effective channels rather than using all channels. Additionally, compared to the literature the present study provides much better classification accuracy improvement rate (CAIR) and channel reduction rate (CRR). Thanks to this method, we focused not only on increasing CA with effective channels, but also highly reducing computational complexity, leading to a reduction in the time required for determining the class of trial compared to other methods. In order to verify the robustness and effectiveness of the proposed method, we used two kinds of EEG based BCI datasets. One of them is the imagination of opening and closing either a left- or right-hand dataset (dataset 1) and the second is cursor movement motor imagery dataset (dataset 2). It is worthwhile noticing that before we extracted features by Hilbert Transform (HT) for dataset 1 and HT+sum derivative (SDR) for dataset 2, we applied a moving-average filter (MAF) as a pre-processing step, which provided better CA performances for both the BCI datasets. We achieved 80.35% and 91.12% average CAs on the test datasets of dataset 1 and dataset 2 with SVM classifier, respectively. Additionally, computational complexity was decreased 88.89% and 54.16% in for dataset 1 and dataset 2, respectively. The achieved results verified that the proposed method provided faster and more accurate results by removing task-irrelevant channels (for a minimum subset of channels).

The remaining sections are organized as follows. In Section 2, the datasets are presented. Then, the proposed method, which includes pre-processing, feature extraction and classification stages, is clarified in Section 3. Then, the results and the performance comparison are provided in tables and figures in Section 4. Finally, Section 5 concludes the paper.

2. Dataset Description

In this study, we tested our proposed algorithm on the two different datasets introduced in the following subsections.

2.1. Dataset 1

Dataset 1 was collected from 29 healthy subjects (Shin et al., 2016) as acquired with 1000 Hz sampling frequency during imagination of opening and closing either left- (*class a*) or right-hand (*class b*) while the participants were grabbing a ball at a 1 Hz pace. Afterwards, the dataset was down sampled to 200 Hz. Signals were recorded by thirty EEG electrodes (1. AFp1, 2. AFp2, 3. AFF1h, 4. AFF2h, 5. AFF5h, 6. AFF6h, 7. F3, 8. F4, 9. F7, 10. F8, 11. FCC3h, 12. FCC4h, 13. FCC5h, 14. FCC6h, 15. T7, 16 T8, 17. Cz, 18. CCP3h, 19. CCP4h, 20. CCP5h, 21. CCP6h, 22. Pz, 23. P3, 24. P4, 25. P7, 26. P8, 27. PPO1h, 28. PPO2h, 29. POO1, 30. POO2 and Fz for ground electrode) located according to the International 10–5 system. The dataset was recorded in an ordinary bright room and experimental procedures were as shown in Figure 1. The experiment consisted of three sessions and each session had a rest of 60 seconds (sec) before the experiment. Each trial was started with 2 sec visual instruction. For the visual instruction, a black arrow pointing to either the left or right side appeared at the center of the screen for 2 sec. The arrow disappeared with a short beep sound and then a black fixation cross was displayed during the task period. Following the visual instruction, the task period consisted of one of two conditions: either first left and then right-hand motor imagery or vice versa. Additionally, each trial was ended with 15-17 sec resting time period. Finally, a session was ended with a 60 sec post-trial resting period. While each session consisted of 20 trials, a total of 60 trials were recorded in 3 sessions. Half of these trials were *class a* and rest of them were *class b*. As presented in Figure 1, at the beginning and end of the task period, a short beep 250 milliseconds in length was played. In this study, 50% of the trials were used for training and the remainder were used for testing in the classification process.

2.2. Dataset 2

For dataset 2, we used BCI Competition 2003 dataset Ia, which was recorded from a healthy subject at Tubingen University in Germany (Blankertz et al., 2003). That dataset was acquired with 256 Hz sampling frequency during imagination of cursor moved up (*class 1*) and down (*class 2*) on the computer screen. Signals were acquired with 6 EEG electrodes (1. A1, 2. A2, 3. FC3, 4. CP3, 5. FC4 and 6. CP4 referenced to Cz electrode) located according to the International 10–20 System. The length of a trial was 3.5 sec. The training set includes 268 trials (135 trials for class 1, 133 trials for class 2) and the test set consists of 293 trials (147 trials for *class 1*, 146 trials for *class 2*).

3. Proposed Method

In this study, we propose a robust and subject-specific sequential forward search method (RSS-SFSM) to remove task-irrelevant channels and reduce the calculation complexity. Our methods were inspired by sequential forward feature selection method, which is frequently used in literature for determined effective features among extracted ones (Xue et al., 2016). Figure 2 summarizes the proposed method, which is applied to the training set to determine an effective channel combination which provides the highest CA on validation set. In this process, first we extracted features from the training set. Then, we individually tested all channels performances and find the best channel, which achieves the highest cross validation accuracy (CVA). Afterwards, we randomly added and evaluated the remaining channel features over again to improve the previously obtained CVA. This procedure continued until all channel combinations are tested. It is worthwhile mentioning that in order to improve the robustness and avoid the problems of random selections in sub-training and validation sets, we applied the aforementioned channel selection procedure 100 times. Finally, the total numbers of the selection of each channel present the effectiveness of the channels. We determined the most effective channels by considering the channels which are selected more than half of the most selected channel.

In order to show the reliability and robustness of the proposed method, the algorithm was applied to two independent datasets. It should be also noted that all parameters were investigated using a training set with random sub-sampling cross-validation procedure. Specific details of the proposed method consist of pre-processing, feature extraction and classification as described below.

3.1. Pre-processing Stage

Pre-processing stage transforms the data into a format that will be more easily and effectively processed for the purpose of system. In this study, we used a moving average filter (MAF) for pre-processing, as given in Equation 1. MAF enables to analyze samples of a trial by creating series of averages of different subsets of the full trial and also helps to eliminate high-frequency noise (Azami et al., 2012).

$$\mathrm{m(n)} = \frac{1}{\mathrm{WS}}\left[\mathrm{p(n)} + \mathrm{p(n-1)} + \cdots \mathrm{p}\big(\mathrm{n} - (\mathrm{WS}-1)\big)\right] \tag{1}$$

In this equation $m(n)$, WS and $p(n)$ indicate filtered trial, windows size and raw trial, respectively. It should be noticed that the windows sizes were determined as 5 and 2, based on cross-validation analysis on the training sets of dataset 1 and dataset 2, respectively.

3.2. Feature Extraction Stage

Feature extraction stage plays an important role in constructing the classifier. Therefore, after pre-processing section, we extracted features from all channels of the datasets. It is noteworthy that in BCI based pattern recognition challenges; there is not a stable feature extraction method which was performed prior to classification any kind of EEG based BCI signals. Therefore, in this study, we applied HT and SDR methods. The Hilbert transform shifts the phase by $\pi/2$ without changing the amplitude of the signal (Pachori et al., 2011). According to this definition, the Hilbert transform ($\hat{p}(t)$) of the input signal ($p(t)$) is calculated as follows:

$$H\{sn(t)\} = \hat{p}(t) = p(t) * \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{p(\tau)}{t-\tau} d\tau \tag{2}$$

where $\hat{p}(t)$ consists of the imaginary and real part as follows:

$$\hat{p}(t) = a \pm jb \tag{3}$$

The average and standard deviation of the real part, standard deviation of the imaginary part of $\hat{p}(t)$ are calculated as follows:

$$HT_1 = \frac{a_1 + a_2 + \cdots + a_N}{N} \tag{4}$$

$$HT_2 = \sqrt{\frac{(a_1 - \bar{a})^2 + (a_2 - \bar{a})^2 + \cdots + (a_N - \bar{a})^2}{N-1}} \tag{5}$$

$$HT_3 = \sqrt{\frac{(b_1 - \bar{b})^2 + (b_2 - \bar{b})^2 + \cdots + (b_N - \bar{b})^2}{N-1}} \tag{6}$$

where $\bar{a}$ and $\bar{b}$ respectively show mean of the a and b coefficients and N indicates the length of a trial.

Additionally, the mathematical presentation of SDR is given in Equation 7.

$$SDR = \sum_{n=1}^{N} p[n] - p[n-1] \tag{7}$$

The random sub-sampling cross-validation procedure on the training set demonstrated that while only the HT_1 feature could be used for classifying dataset 1 trials, $HT_1 + HT_2 + HT_3 + SDR$ features could be used for classifying dataset 2 trials with higher performance.

3.3. Classification Stage

In this study, the proposed method was tested using an SVM classifier. SVM is a high-performing algorithm and frequently used for classification problem in data mining (Cortes et al. 1995; Fu et al. 2014). It determines the most appropriate hyper plane on the training data set to separate the class labels into subgroups known as $\{-1, +1\}$. Training set is given as $T = \{\{x_1, y_1\}, \dots \dots .\{x_n, y_n\}\}$ where $\mathrm{y} \in (-1,1)$ represents the class labels and n specifies the data size. A number of hyper planes can be drawn to separate these data. SVM separates the two classes of points by a hyper plane,

$$\mathrm{h} \cdot \mathrm{y_i} + \mathrm{r} = 0 \tag{8}$$

where h signifies the weight vector, r defines the bias and $i = 1, 2, \dots, n$. The purpose of this algorithm is to select the hyper plane that best divides the data classes. Inequalities for the optimal hyper plane are given in Equation 9. Along with optimal hyper plane, the support vectors which have an equal distance to this plane and boundary planes are shown in Figure 3.

$$\begin{aligned} \mathrm{h} \cdot \mathrm{y_i} + \mathrm{r} &\geq +1 \\ \mathrm{h} \cdot \mathrm{y_i} + \mathrm{r} &\leq -1 \end{aligned} \tag{9}$$

4. Results

In this study, we proposed RSS-SFSM for determining effective channels among all channels used for recording motor imagery EEG signals. RSS-SFSM was applied to training parts of the datasets by cross-validation analysis over 100 runs. Based on training CAs results, the effective channels were determined. The results of average training CAs and standard deviations were shown in Figure 4 for the dataset 1. In this figure, the x-axis represented subjects as $S1, S2, \ldots\ldots, S29$ and y-axis shows the average CAs which were calculated for 100 runs. Moreover, training CA averaged over 29 subjects was calculated as 86.12%$\pm$ 5.52. Based on cross-validation analysis over 100-times run, selected channels were also given in Figure 5 for dataset 1 and Figure 6 for dataset 2. In this figure, the horizontal axis represents the channels of the datasets, while the vertical axis expresses the number of the selected channels.

For the following experiment, after we selected channels based on 100 runs on training sets, effective channel selection stage was passed. To do this, first of all, the threshold value was determined. Per Figure 5 and Figure 6, we could recognize the threshold value by considering the channels which are selected more than half of the number of the best selected channels after 100 runs of training stages. Secondly, effective channels were assigned for each subject based on this threshold value. In Figure 7, dark-colored squares illustrate the determined effective channels for each subject for dataset 1. In this figure, the horizontal axis shows the channels and the vertical axis shows the subjects. With a similar approach, A1 and A2 electrodes were determined as effective channels for the dataset 2. It should be mentioned that although determining the effective channels took some time in the training stage, which were approximately 874.25 and 41.9 minutes for dataset 1 and dataset 2, respectively, using effective channels decreased computational complexity in the testing stage and provided a fast and accurate EEG-based BCI system. As shown in Table 1, compared to using all channels, the proposed method also decreased the computational complexity 88.89% (average of 29 subjects) and 54.16% for dataset 1 and dataset 2, respectively. Determining the class of a test trial only took 0.33 and 1.12 milliseconds for dataset 1 and dataset 2, respectively. All the runtime experiments were conducted on a PC with Intel®CoreTMi5 CPU at 1.70 GHz and 4 GB RAM.

Based on the determined effective channels, test results are presented in Figure 8 as radar charts. The average test CA, sensitivity (SE) and specificity (SP) over all subjects were achieved as 80.35%, 81.38% and 79.31% for the dataset 1, respectively. For the dataset 2, test CA, SE and SP were calculated as 91.12%, 97.95% and 84.24%, respectively. In order to clarify the improvement of proposed method, we calculated the test CA, SE and SP values using all channels, which were given in Figure 9 as radar charts for dataset 1. In Figure 8 and Figure 9, while circles represented CAs, SEs and SPs from 0.00% to 100.00%, each line dividing these circles signified S1,S2,…..,S29 expressed all subjects. The averages of these metrics were obtained as 57.93%, 72.42%, and 43.45%. For dataset 2, we obtained the test CA, SE and SP as 81.57%, 76.19%, and 86.98%, respectively. As a result, it can be said that the CA performances of dataset 1 and dataset 2 were improved by 22.42% and 9.55% with effective channels, respectively. Thus, the average improvement for dataset 1 and dataset 2 was calculated as 15.98%. These results indicate that the RSS-SFSM algorithm is capable of selecting effective channels.

In order to determine its effectiveness, the performance of the proposed RSS-SFSM is compared with other studies reported in literature. Table 2 compares the channel selection methods in terms of average test CA with effective channels, average number of effective channels, CAIR and CRR (Ghaemi et al., 2017). The CAIR and CRR were calculated as given in Equation 10 and Equation 11, respectively.

$$\mathrm{CAIR} = \left(\frac{\text{Average CA with effective channels}}{\text{Average CA with all channels}} - 1\right)\mathrm{x}100 \tag{10}$$

$$\mathrm{CRR} = \left(1 - \frac{\text{Average number of effective channels}}{\text{Total number of channels}}\right)\mathrm{x}100 \tag{11}$$

While the average CA with effective channels, CAIR and CRR were respectively obtained as 80.35%, 38.70% and 90.93% for dataset 1, they were respectively achieved as 91.12%, 11.70% and 66.67% for dataset 2. Thus, the averages of these values were respectively calculated as 85.73%, 25.20% and 78.80% for both datasets. Based on these results, it can be said that the proposed method outperforms other existing methods in terms of average CAIR and CRR values. Compared to the IBGSA method, which achieved the closest rates, the proposed method has better performance levels at 8.32% and 9.65% with respect to the average CAIR and CRR values, respectively. On the other hand, although ISFFS method has only better performance in terms of the CRR for dataset 1, the proposed method offers 13.6% better performance in average value of CRR.

To approve the effectiveness of the proposed feature extraction methods, we also tested the most popular motor imagery related features, including statistical (Kevric and Subasi, 2017), Fourier transform (FT) (Wang and Veluvolu, 2017) and CSP (Zhang et al., 2017; Aydemir, 2016) based techniques. While we calculated the mean of absolute values, the average power, the standard deviation, the ratio of the absolute mean, the skewness and the kurtosis values of each trials for the statistical features, frequency-domain band power was obtained as feature for the FT. The test CA results of these methods are given in Table 3. Although the CAIR and the CRR values appeared to be relatively high, the CAs were quite poor. While the statistical, the FT and the CSP methods respectively achieved 63.80%, 61.73% and 63.91% CAs for dataset 1, they were obtained as 71.33%, 64.16% and 59.09% for dataset 2, respectively. Compared to these results, it can be said that the proposed features are proved successfully and effectively to classify motor imagery EEG datasets with RSS-SFSM.

5. Conclusion

Investigation of the most effective channels is crucial to develop a high performance BCI system in terms of classification accuracy and computational complexity. In this study, we have proposed a robust and subject-specific sequential forward search method for effective channel selection in brain computer interfaces. To validate its success, the proposed method was applied to two different BCI datasets. The results achieved indicate that the proposed method not only decreases the computational complexity, but also improves the CA performance of BCI datasets.

Based on the results of 100 runs on the training sets as given in Figure 5 and Figure 6, it is clearly seen that while some channels were frequently selected since they were contributed to improve CA, some of them were rarely or never selected. For example, while the 1^{st} channel was selected 50 times, the 3^{rd}, 10^{th}, 15^{th}, 20^{th}, 26^{th} and 29^{th} channels were never selected for Subject 2 in the dataset 1. On the other hand, for dataset 2, while the 1^{st} channel was almost always selected, 4^{th} and 5^{th} channels were rarely selected. Therefore, it can be concluded that channel selection is a necessary stage for eliminating redundant channels.

Based on the determined effective channels shown in Figure 7, we achieved 80.35% average test CA for dataset 1, which was 22.42% better than the result of all channel. Individually, the highest CAs were achieved for Subject 17, Subject 21 and Subject 28 at 90%, as given in Figure 8. In contrast to this, the lowest CAs were calculated for Subject 9, Subject 12 and Subject 16 as 73.34%. It is worthwhile mentioning that while the number of effective channels were determined as at least 2 for 15 subjects, a maximum of 5 effective channels were determined for only 1 subject (Subject 27). In average 2.72 channels out of 30 were determined to be effective. On the other hand, for dataset 2, we obtained 91.12% test CA, which was 9.55% better than the result of all channels. For dataset 2, only two channels (A1 and A2) out of 6 were determined to be effective.

Another good attribution of the proposed method was its simplicity in the feature extraction stage. The feature vectors were very basically calculated, which were HT_1 for dataset 1 and $HT_1 + HT_2 + HT_3 + SDR$ for dataset 2. Additionally, although 4^{th}, 10^{th} or 19^{th} electrodes were never selected, it should be noted that only 3^{rd}, 16^{th}, 20^{th} or 25^{th} electrodes were generally selected as effective channels for subjects for the dataset 1. It is worthwhile mentioning that F7 electrode, which was located in the left side of brain, was selected for all subjects as effective.

Compared with the result of Shin et al. 2016, the reported results showed that the proposed method achieved 14.75% CA improvement over the best accuracy result of the dataset 1. In conclusion, the obtained results proved that the proposed RSS-SFSM shows great potential for determining effective channel(s) which classify the BCI trials more quickly and accurately.

Acknowledgement

The authors acknowledge Recep Tayyip Erdogan University for allowing them to use the Samsung Laboratory.

References

Aydemir, O. and Kayikcioglu, T. 2014. 'Decision tree structure based classification of EEG signals recorded during two dimensional cursor movement imagery', *Journal of neuroscience methods*, 229, 68-75.

Aydemir, O. 2016. 'Common spatial pattern-based feature extraction from the best time segment of BCI data', *Turkish Journal of Electrical Engineering & Computer Sciences*, 24(5), 3976-3986.

Azami, H., Mohammadi, K., & Bozorgtabar, B. 2012. 'An improved signal segmentation using moving average and Savitzky-Golay filter', *Journal of Signal and Information Processing*, 3(01), 39.

Blankertz, B., Mueller, K.R., Curio, G., Vaughan, T.M., Schalk, G., Wolpaw, J.R., Schloegl, A., Neuper, C., Pfurtscheller, G., Hinterberger, T., Schroeder, M., Birbaumer, N., 2004. 'The BCI competition 2003'. *IEEE Trans. Biomed*. Eng. 51, 1044–1051.

Cortes, C., & Vapnik, V. 1995. 'Support-vector networks', *Machine learning*, 20(3), 273-297.

Dong, E., Li, C., Li, L., Du, S., Belkacem, A. N., & Chen, C. 2017. 'Classification of multi-class motor imagery with a novel hierarchical SVM algorithm for brain–computer interfaces'. *Medical & biological engineering & computing*, 55(10), 1809-1818.

Fu, K., Qu, J., Chai, Y., & Dong, Y. 2014. 'Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM', *Biomedical Signal Processing and Control*, 13, 15-22.

Gaur, P., Pachori, R. B., Wang, H., & Prasad, G. 2018. 'A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry', *Expert Systems with Applications*, 95, 201-211.

Ghaemi, A., Rashedi, E., Pourrahimi, A. M., Kamandar, M., & Rahdari, F. 2017. 'Automatic channel selection in EEG signals for classification of left or right hand movement in Brain Computer Interfaces using improved binary gravitation search algorithm', *Biomedical Signal Processing and Control*, 33, 109-118.

Gonzalez, A., Nambu, I., Hokari, H., & Wada, Y. 2014. 'EEG channel selection using particle swarm optimization for the classification of auditory event-related potentials'. The Scientific World Journal, 2014. Article ID 350270, 1-11.

Handiru, V. S., & Prasad, V. A. 2016. 'Optimized Bi-Objective EEG Channel Selection and Cross-Subject Generalization With Brain–Computer Interfaces', *IEEE Transactions on Human-Machine Systems*, 46(6), 777-786.

Hamzah, N., Norhazman, H., Zaini, N., & Sani, M. 2016. ‘Classification of EEG signals based on different motor movement using multi-layer Perceptron artificial neural network’. *J Biol Sci*, 16(7), 265-71.

He, L., Hu, Y., Li, Y., & Li, D. 2013. ‘Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG’, *Neurocomputing*, 121, 423-433.

Kevric, J., and Subasi, A. 2017. ‘Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system’, *Biomedical Signal Processing and Control*, 31, 398-406.

Kumar, S., Sharma, R., Sharma, A., & Tsunoda, T. 2016. ‘Decimation filter with common spatial pattern and fishers Discriminant analysis for motor imagery classification’. In Neural Networks (IJCNN), 24-29 July 2016; Vancouver, BC, Canada: IEEE. pp. 2090-2095.

Kumar, S., Mamun, K., & Sharma, A. 2017. ‘CSP-TSM: optimizing the performance of Riemannian tangent space mapping using common spatial pattern for MI-BCI’. *Computers in biology and medicine*, 91, 231-242.

Kumar, S., Sharma, A., & Tsunoda, T. 2017. ‘An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information’, *BMC bioinformatics*, 18(16), 545.

Kumar, S., & Sharma, A. 2018. ‘A new parameter tuning approach for enhanced motor imagery EEG signal classification’, *Medical & biological engineering & computing*, 1-14.

Lo, C. C., Chien, T. Y., Chen, Y. C., Tsai, S. H., Fang, W. C., & Lin, B. S. 2016. ‘A wearable channel selection-based brain-computer interface for motor imagery detection’, *Sensors*, 16(2), 213.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. 2007. ‘A review of classification algorithms for EEG-based brain–computer interfaces’, *Journal of neural engineering*, 4(2), R1.

Miao, M., Wang, A., & Liu, F. 2017. ‘A spatial-frequency-temporal optimized feature sparse representation-based classification method for motor imagery eeg pattern recognition’, *Medical & biological engineering & computing*, 55(9), 1589-1603.

Qiu, Z., Jin, J., Lam, H. K., Zhang, Y., Wang, X., & Cichocki, A. 2016. ’Improved SFFS method for channel selection in motor imagery based BCI’, *Neurocomputing*, 207, 519-527.

Pachori, R. B., & Bajaj, V. 2011. ‘Analysis of normal and epileptic seizure EEG signals using empirical mode decomposition’, *Computer methods and programs in biomedicine*, 104(3), 373-381.

Shin, J., von Lühmann, A., Blankertz, B., Kim, D. W., Jeong, J., Hwang, H. J., & Müller, K. R., 2017. ‘Open access dataset for EEG+ NIRS single-trial classification’, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10), 1735-1745.

Tabar, Y. R. and Halici, U. 2016. ‘A novel deep learning approach for classification of EEG motor imagery signals’, *Journal of neural engineering*, 14(1), 016003.

Torres-García, A. A., Reyes-García, C. A., Villaseñor-Pineda, L., & García-Aguilar, G. 2016. ‘Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification’, *Expert Systems with Applications*, 59, 1-12.

Xue, B., Zhang, M., Browne, W. N., and Yao, X. 2016. ‘A survey on evolutionary computation approaches to feature selection’, *IEEE Transactions on Evolutionary Computation*, 20(4), 606-626.

Wang, Y. and Veluvolu, K. C. 2017. ‘Evolutionary algorithm based feature optimization for multi-channel EEG classification’, *Frontiers in neuroscience*, 11, 28.

Yang, J., Singh, H., Hines, E. L., Schlaghecken, F., Iliescu, D. D., Leeson, M. S., & Stocks, N. G. 2012. ‘Channel selection and classification of electroencephalogram signals: an artificial neural network and genetic algorithm-based approach’, *Artificial intelligence in medicine*, 55(2), 117-126.

Yavuz, E. and Aydemir, Ö. 2017. 'Determination of effective channels in brain-computer interface applications', In Signal Processing and Communications Applications Conference (SIU), 15-18 May 2017; Antalya, Turkey: IEEE. pp. 1-4.

Zhang, Y., Wang, Y., Jin, J., & Wang, X. 2017. 'Sparse Bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification'. *International journal of neural systems*, 27(02), 1650032.

Zhang, Y., Wang, Y., Zhou, G., Jin, J., Wang, B., Wang, X., & Cichocki, A. 2018. 'Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces'. *Expert Systems with Applications*, 96, 302-310.

Figure captions

Figure 1. Experimental procedure for dataset 1

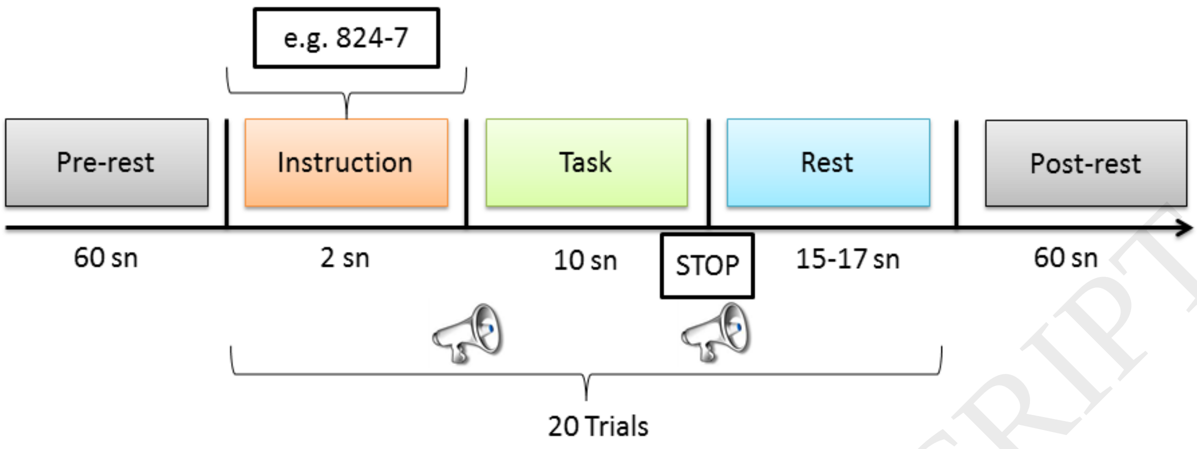

Figure 2. Flow chart of training process

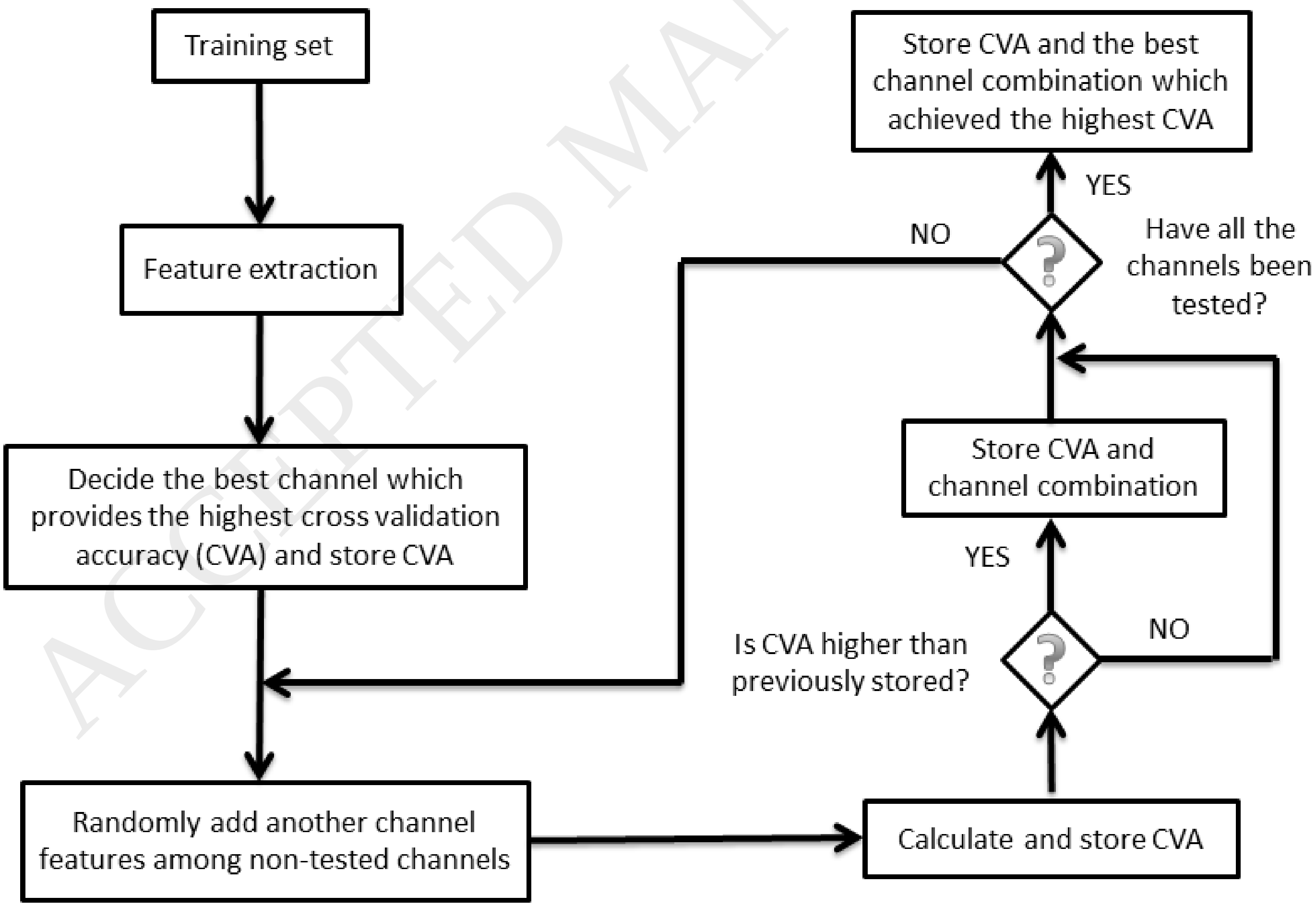

Figure 3. Border planes and optimal hyper plane for a 2-class problem

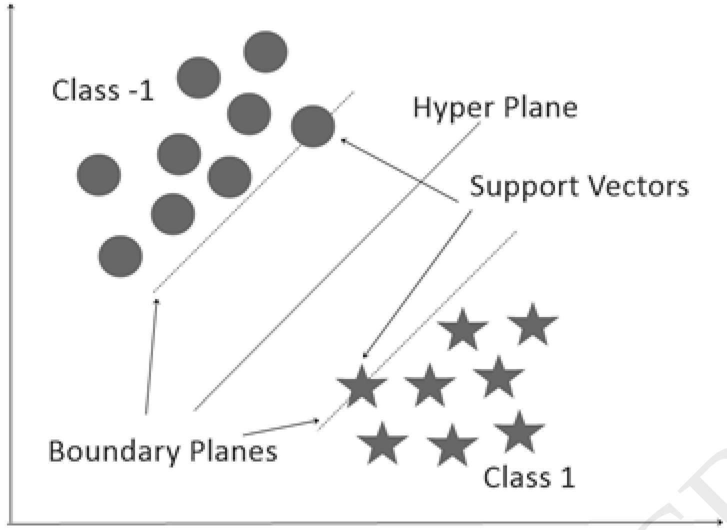

Figure 4. The average training results of dataset 1

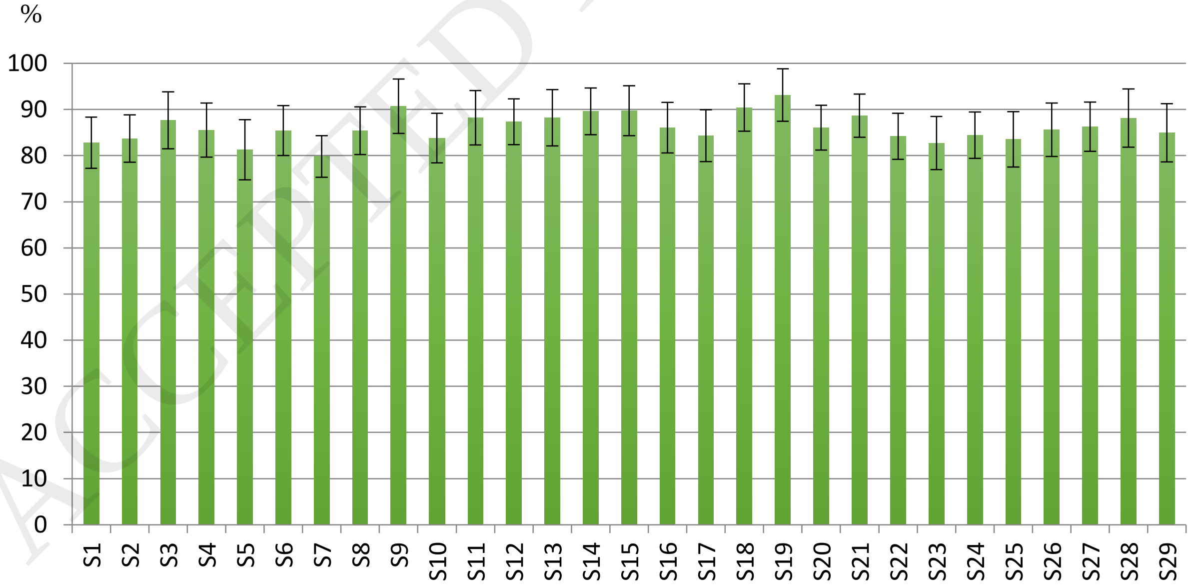

Figure 5. Effective channel distribution of dataset 1

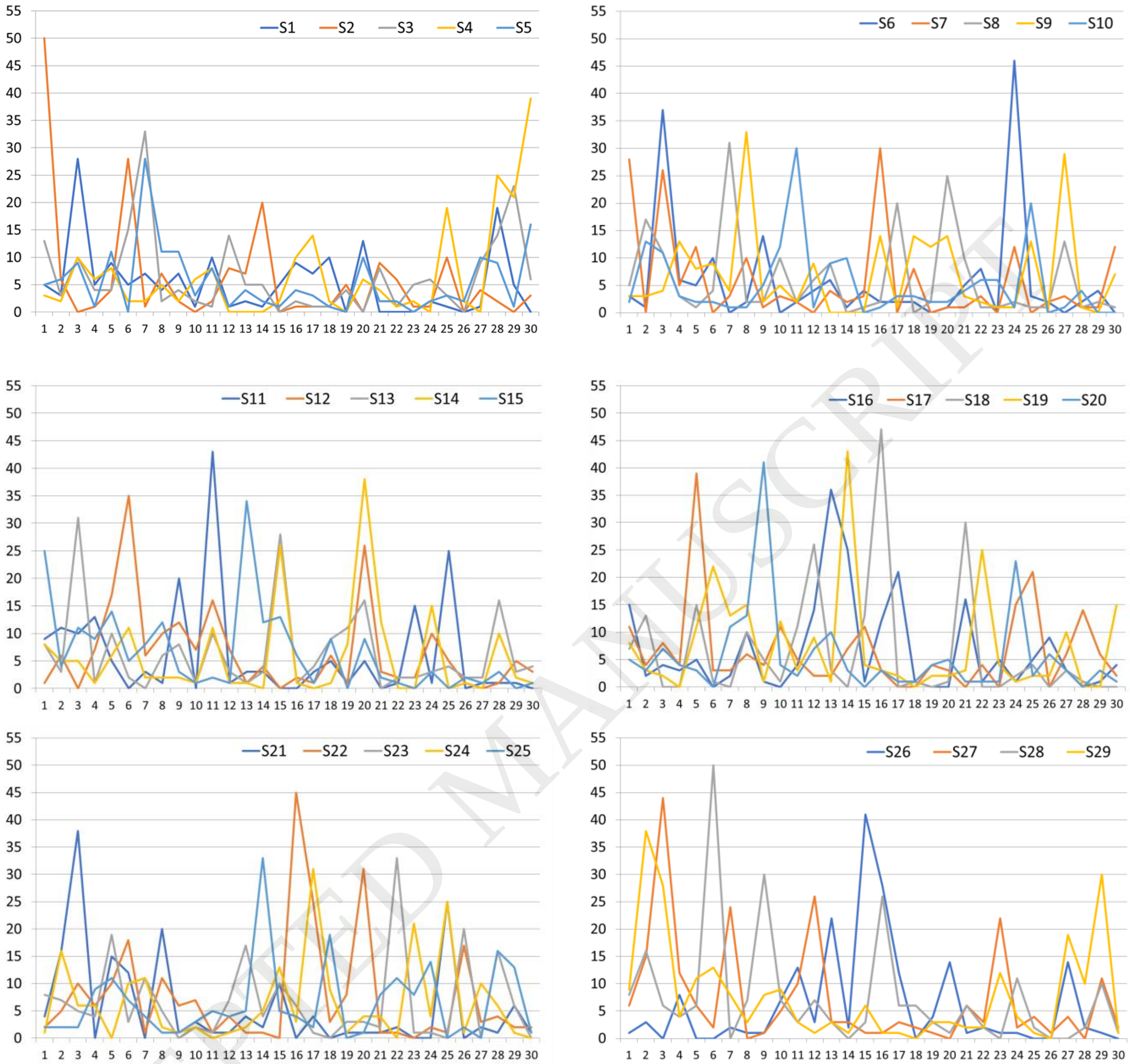

Figure 6. Effective channel distribution of dataset 2

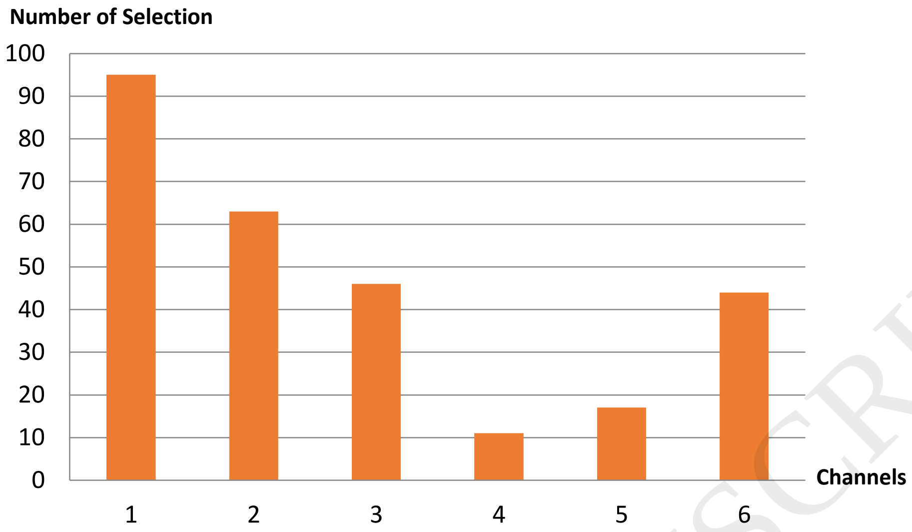

Figure 7. Determined effective channels of dataset 1

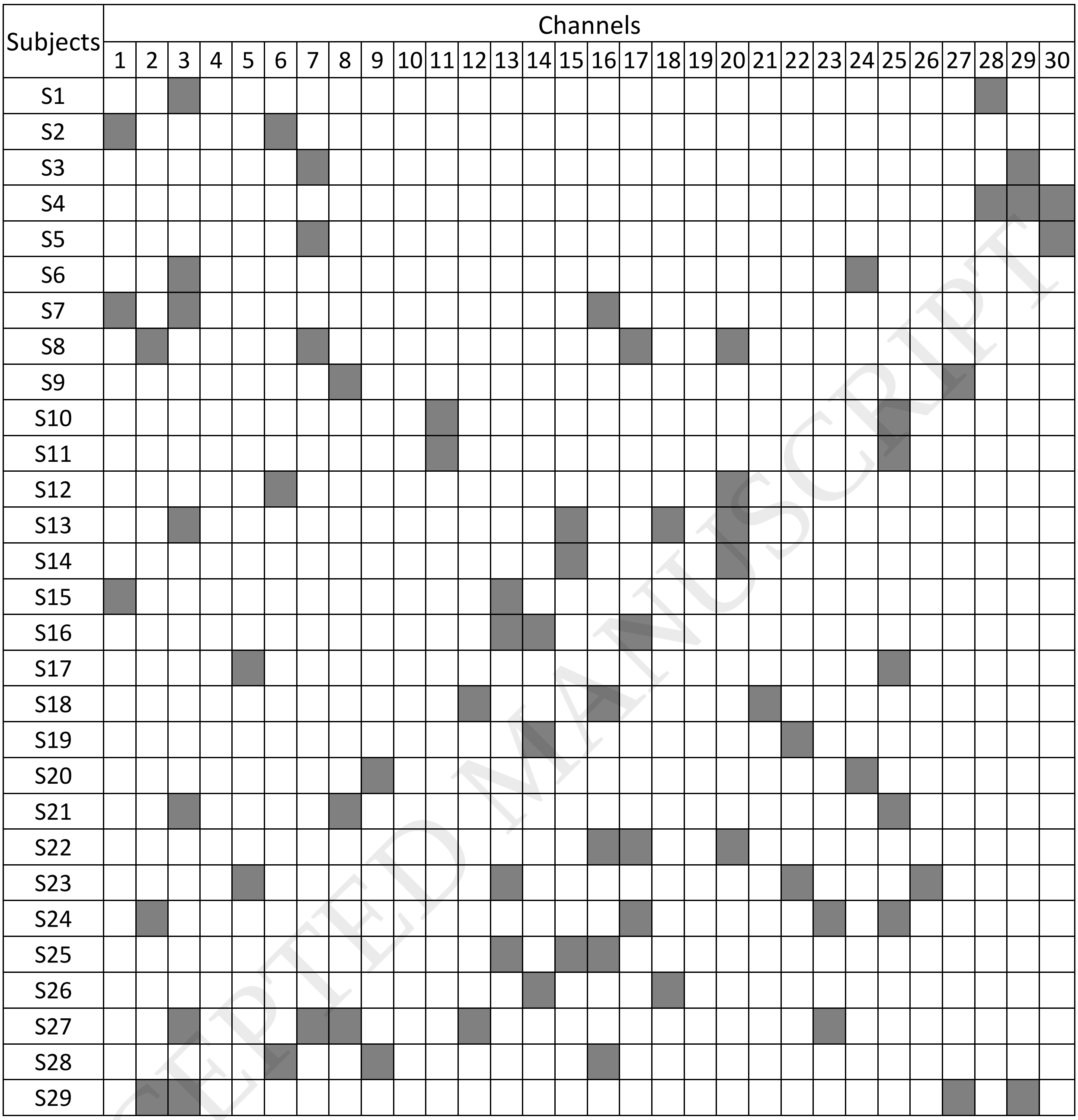

Figure 8. Test CAs, SEs and SPs with effective channels for dataset 1

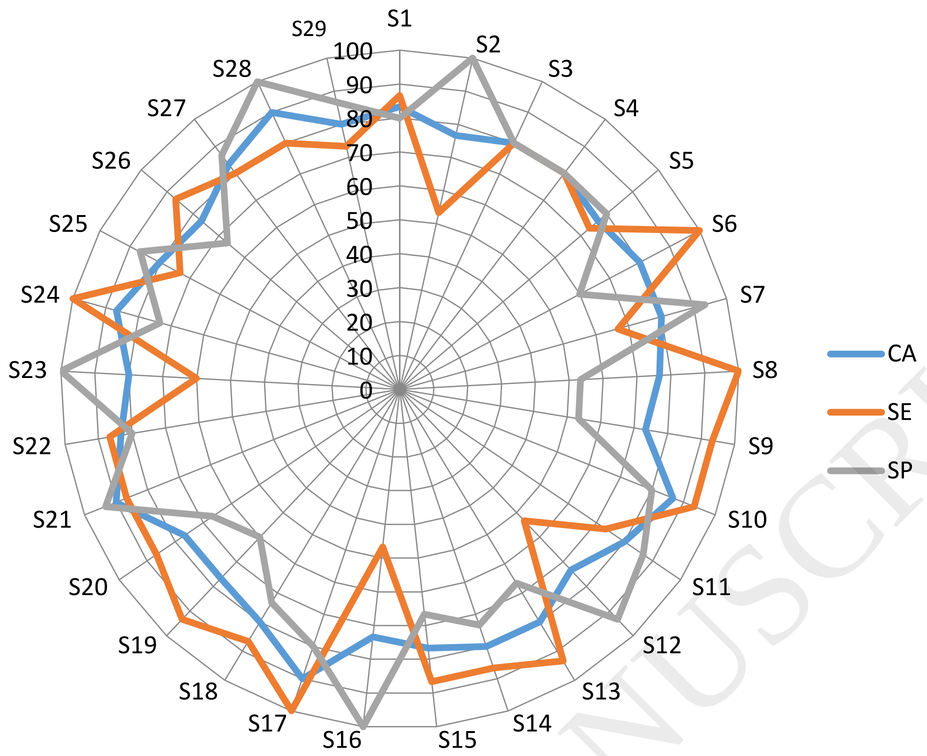

Figure 9. Test CAs, SEs and SPs with all channels for dataset 1

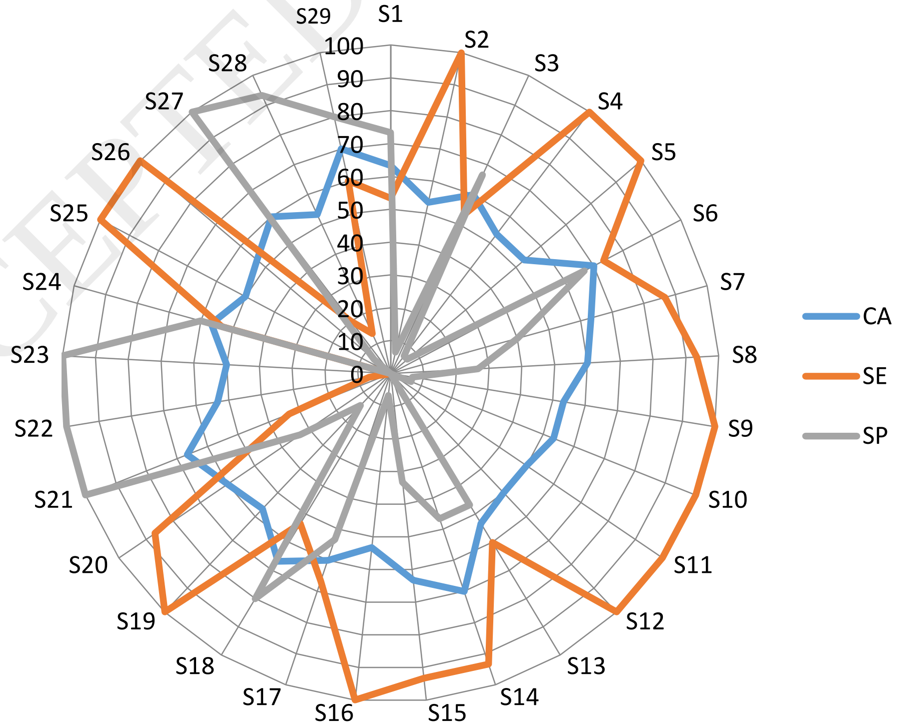

Table captions

Table 1. Computational complexity of feature extraction stage

	With effective channels	With all channels	Reduction Rate
Dataset 1	0.02sec.	0.18 sec.	88.89%
Dataset 2	0.33sec.	0.72sec.	54.16%

Table 2. Performance comparison

Method	Dataset	Average number of effective channels	Average CA with effective Channels (%)	CAIR	CRR
GA (Yang et al. 2012)	Dataset 1	6.90	71.42	23.29	77.00
	Dataset 2	4.00	83.25	2.06	33.33
	Average	5.45	76.84	12.68	55.17
PSO (Gonzalez et al. 2014)	Dataset 1	6.04	67,52	16,55	79.87
	Dataset 2	4.00	89.76	10.04	33.33
	Average	5.02	78.64	13,29	56,60
ISFFS (Qiu et al. 2016)	Dataset 1	2.43	66.54	14.86	91,90
	Dataset 2	3.69	86.81	6.42	38.50
	Average	3.06	76.68	10.64	65.20
IBGSA (Ghaemi et al. 2017)	Dataset 1	3.76	73.54	26.95	87.47
	Dataset 2	2.95	87.12	6.80	50.83
	Average	3.36	80.33	16.88	69.15
Proposed	Dataset 1	2.72	80.35	38.70	90.93
	Dataset 2	2.00	91.12	11.70	66.67
	Average	2.36	85.73	25.20	78.80

Table 3. The results of commonly used feature extraction methods

Method	Dataset	Average number of effective channels	Average CA with effective Channels (%)	CAIR	CRR
Statistical	Dataset 1	2.72	63.80	15.12	90.94
	Dataset 2	1.00	71.33	4.51	83.34
	Average	1.86	67.57	9.82	87.14
FT	Dataset 1	2.86	61.73	8.69	90.47
	Dataset 2	3.00	64.16	9.30	50.00
	Average	2.93	62.95	9.00	70.24
CSP	Dataset 1	2.79	63.91	9.03	90.70
	Dataset 2	2.00	59.09	13.92	66.67
	Average	2.40	61.50	11.48	78.69