

# Effect of Feature and Channel Selection on EEG Classification

Ahmed Al-Ani and Akram Al-Sukker

**Abstract**—In this paper, we evaluate the significance of feature and channel selection on EEG classification. The selection process is performed by searching the feature/channel space using genetic algorithm, and evaluating the importance of subsets using a linear support vector machine classifier. Three approaches have been considered: (i) selecting a subset of features that will be used to represent a specified set of channels, (ii) selecting channels that are each represented by a specified set of features, and (iii) selecting individual features from different channels. When applied to a Brain-Computer Interface (BCI) problem, results indicate that improvement in classification accuracy can be achieved by considering the correct combination of channels and features.

## I. INTRODUCTION

The analysis of EEG signals play an important role in a wide range of applications, such as psychotropic drug research, sleep studies, seizure detection and brain computer interface. Various EEG analysis methods have been proposed in the literature, and some of these methods achieved good results in specific applications [1]. However, automated EEG analysis is still a very challenging problem due to the poor resolution of EEG and its multi-channel nature.

For the particular problem of Brain-Computer Interface (BCI), there has been an extensive research on studying the EEG signals of subjects while performing different mental tasks. Three main aspects can be considered to improve the performance of BCIs, namely signal processing and feature extraction [2], [3], feature and channel selection [4], [5], [6], and classification technique [7], [8]. In this paper, we will focus on the feature and channel selection aspect.

Feature subset selection aims at reducing the feature set dimensionality through selecting a subset of features that performs the best under some classification system. This is essential to reduce computational cost and improve classification performance, especially when dealing with finite sample size. On the other hand, the appropriate choice of number of channels and their locations plays an important role in the analysis of EEG signals. In some cases, there is no clear agreement about the number and location of necessary channels to collect data. Using small number of channels without carefully choosing their locations may cause a loss of important information. Alternatively, including more channels to collect data will provide redundant information, which could degrade the system performance. Hence, both feature and channel selection are important to EEG analysis.

The application of feature and channel selection to BCI have been studied by a number of researchers. In [4],

frequency component selection has been considered for two channels (C3 and C4), and the authors concluded that the selection should be individual for each subject as well as for each channel. Six EEG channels have been considered in [5], and for each channel a set of features that represent the average power spectra of the standard EEG frequency bands have been used. The search was performed over 11 time windows and six frequency bands, while including all six channels in each case. The paper indicates that the subset that achieved the highest classification accuracy contained features from every time window and every frequency band. Channel selection has been addressed in [6], where 39 channels have been used, each represented by coefficients of an AR model of order 3. The importance of the channels were ranked for five subjects, and the results indicated that channels around the motor area of the cortex were relatively more important than other channels.

This work explores a more general approach to the problem in terms of analyzing both feature and channel selection. Firstly, we will search for the best subset of features that would represent a previously specified subset of channels. Secondly, we will use a fixed subset of features that represents each one of the channels and search for the best channel combination. Thirdly, both channel and feature spaces will be searched for the best subset of features from different channels. Finally, we will divide the scalp into two regions and compare the performance of their corresponding channels.

The paper is structured as follows: Section II describes the data. Explanation of the selection procedure is presented in Section III. In Section IV, experimental results are presented, and a conclusion is given in Section V.

## II. DESCRIPTION OF DATA

The data used here was obtained from the Department of Medical Informatics, University of Technology, Graz, Austria. EEG signals were recorded for three right-handed females with 56 Ag/AgCl Electrodes using monopolar montage, with reference electrode on the right ear. Fig. 1 illustrates the position of electrodes. The subjects were placed in an armchair and asked to imagine right or left finger movements according to stimuli on screen. A total of 8 seconds of data were recorded at 128 Hz sampling rate, 2 seconds before the stimuli and 6 after it. A total of 406 trials were used, 208 for left movement and 198 for right movement.

Since EEG is a time-varying and space-varying non-stationary signal, we used the wavelet transform to extract features from the data of each trial. According to [9], the

A. Al-Ani and A. Al-Sukker are with Faculty of Engineering, University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia  
ahmed@eng.uts.edu.au, alsukker@eng.uts.edu.au

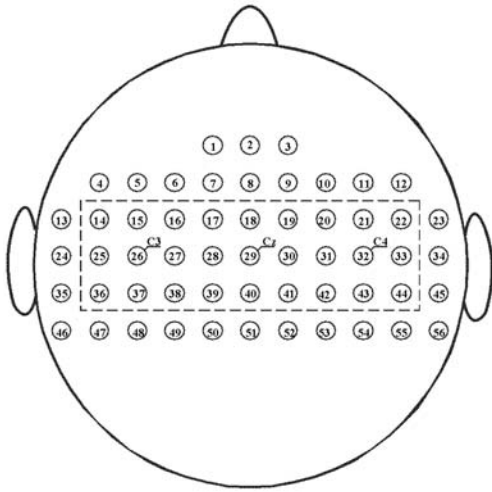


Fig. 1. Position of EEG electrodes

wavelet transform was found to provide good way to visualize and decompose EEG signals into measurable component events. The decomposition of the signal using this tool gives scaled and shifted version of the original "mother" wavelet. The first order Daubechi (db1) was chosen as the mother wavelet in this work.

### III. THE SELECTION PROCEDURE

All feature/channel selection methods need to use some sort of evaluation function together with a search procedure to find the optimal feature/channel subset. Evaluation functions estimate the importance of subsets in discriminating between classes, and can be divided into two main groups: filters and wrappers. Filters measure the relevance of subsets independently of any classifier, whereas wrappers use classifiers' performance as the evaluation measure. Because of this, Filter methods are faster than wrappers, but their performance is usually not as good as that of wrapper methods. We adopt a wrapper method in this work, where the importance of subsets are measured using the classification accuracy of a Linear Support Vector Machine (LSVM). The LSVM was found to be a reliable classifier, especially for binary classification problems that deal with limited number of training patterns.

Search procedures, on the other hand, are methods that only consider small portions of all possible subsets. several search procedure methods have been developed, which basically differ in their computational cost and the optimality of the subsets they find. One of the well-known search procedures is the Genetic Algorithm (GA), which is a combinatorial search technique based on both random and probabilistic measures. Subsets of features/channels are evaluated using a fitness function and then combined via cross-over and mutation operators to produce the next generation of subsets [10]. The GA employs a population of competing solutions, evolve over time, to converge to an optimal solution. Effectively, the solution space is searched in parallel, which helps in avoiding local optima. A GA-based

feature selection solution would typically be a fixed length binary string representing a feature/channel subset, where the value of each position in the string represents the presence or absence of a particular feature/channel.

In all the experiments described in the next section, a GA-based selection method is implemented using the average classification accuracy of a seven-fold cross-validation, obtained by a LSVM, as the fitness function. Trials from all three subjects will be used, i.e, the experiments are not subject dependent.

## IV. EXPERIMENTAL RESULTS

### A. Feature Selection

We extracted 13 features that represent the energy of different frequency bands extracted from the discrete wavelet packet transform. Two packet trees (shown in Fig. 2) have been used for this purpose to obtain the following frequency bands: 0 – 4, 4 – 8, 8 – 12, 12 – 16, 16 – 20, 20 – 24, 24 – 28, 28 – 32, 0 – 8, 8 – 16, 16 – 24, 24 – 32 and 32 – 64 Hz.

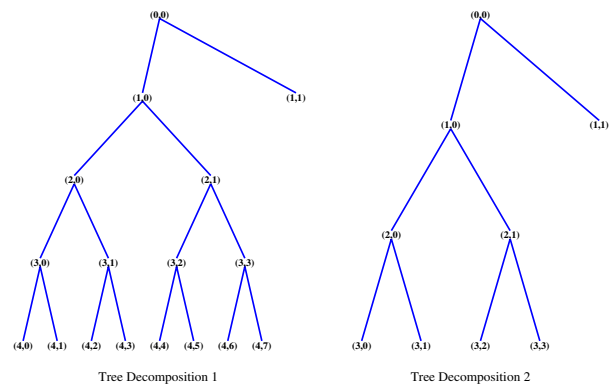


Fig. 2. Wavelet packet trees

In this experiment, only channels 26 and 32, which represent channels C3 and C4 respectively, are used. In order to select the best frequency bands that would represent both channels, we varied the desired number of features to be selected between 2 and 10. The selection procedure described in the previous section is used to select the best subsets. The GA-based selection is performed using the following parameter settings: population size = 30, number of generations = 25, probability of crossover = 0.8, and probability of mutation = 0.05. The obtained binary strings are constrained to have the number of '1's matching the predefined number of desired features. The obtained results are shown in Table I.

The obtained results indicate that there is a small variation between the classification accuracy of feature subsets with the desired number of features ranging between 3 and 9. As using a small number of features is more desirable, a subset of 3 features that represent the energy values of the frequency bands 4 – 8, 8 – 16 and 16 – 24 Hz will be used in all the remaining experiments.

TABLE I  
PERFORMANCE OF THE SELECTED FEATURE SUBSETS

Desired No. of features	Classification Accuracy	Sensitivity class1/ class2	Specificity class1/ class2
2	68.00%	66.47/69.85	72.91/79.37
3	70.25%	69.78/72.35	75.11/80.47
4	70.75%	68.80/73.92	74.39/80.77
5	71.25%	70.16/73.16	75.24/80.66
6	71.25%	67.85/75.40	74.33/82.23
7	70.50%	71.18/70.46	75.59/79.14
8	70.75%	72.42/70.30	76.25/79.14
9	71.00%	69.71/73.04	74.77/80.77
10	69.25%	69.06/69.53	74.75/78.77

It is worth mentioning that the classification accuracy of features 3 and 6 (frequency bands 8 – 12 and 20 – 24 Hz) is found to be 60.00%, which is clearly outperformed by the results shown in Table I. Those two features represent frequency bands that are similar to the ones suggested in [4].

### B. Channel Selection

In this section, we used the chosen subset of 3 features to represent each of the 56 channels shown in Fig. 1. The GA-based selection is performed using the same parameters described earlier, but because we are dealing with a bigger search space, we chose to increase the population size and number of generations to 50. The aim here is to investigate the following:

- Are channels C3 and C4 the best two channels?
- Will it be useful to include more than two channels?

Firstly, we calculated the classification accuracy of all possible subsets of two channels (1540 subsets), and found that channels C3 and C4 give the best result. This supports the literature in identifying the importance of channels C3 and C4 in analyzing motor imagery data [4].

In the second experiment, we varied the desired number of channels between 2 and 50 and performed channel selection. The training and testing classification accuracy of the selected subsets are shown in Fig. 3. The results clearly show the advantage of including more than two channels, where a maximum classification accuracy of 83.25% (for the test set) could be achieved using 16 channels. This represents an improvement of 13% compared to the accuracy of channels C3 and C4, and a reduction in the error rate of 43.70%<sup>1</sup>.

It can also be noticed that the testing classification accuracy starts to decline after selecting more than 20 channels, and the decline becomes more rapid when selecting 40 or more channels. On the other hand, the classification accuracy of the training set continues to have some increase when more channels are included. The main reason for the improvement in the training set and the deterioration in the test set is that as the number of channels increases, there will be more parameters for the classifier to estimate, which

<sup>1</sup>The equation of the error reduction rate is  $(Er_{b2} - Er_{b16})/Er_{b2} \times 100\%$ , where  $Er_{b2}$  is the error rate of the best 2 channels (29.75%), and  $Er_{b16}$  is the error rate of the best 16 channels (16.75%)

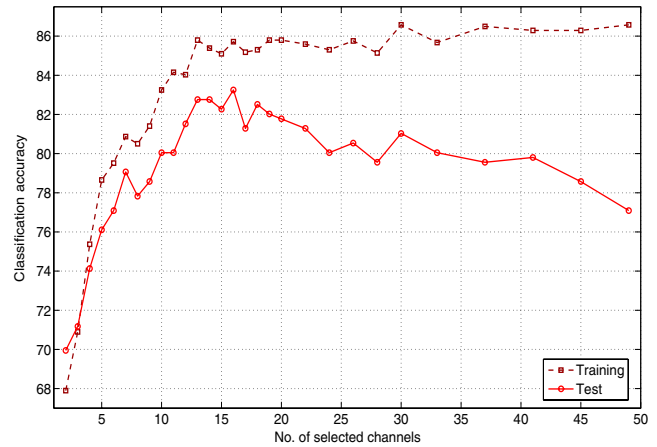


Fig. 3. Classification accuracy of the selected channels for the training and test sets

makes it harder to generalize, especially when dealing with limited number of patterns.

### C. Feature/Channel Selection

In the previous section, we used three features to represent each channel, and hence, when two channels are selected, six features are actually used to classify each trial. We will now investigate the importance of selecting individual features from different channels, which may be referred to as combined feature/channel selection. This will increase the search space, as there will be 168 variables, compared to 56 for the channel selection. However, the GA-based selection parameters are kept the same in order not to increase the computational time.

In this experiment, the desired number of selected features was varied between 6 and 150, i.e., the equivalent of [2, 50] channels each represented with three features. It has been found that the selected features are extracted from far more channels than the ones used when implementing channel selection, i.e., more channels are exploited. For instance, when the desired number of features was set to 12, the best subset was found to consist of features extracted from 11 different channels.

The obtained results, shown in Fig. 4, indicate that the selection of individual features from different channels is particularly useful when selecting 30 features or less (equivalent to 10 or less channels with 3 feature each). An error reduction rate of up to 16.4% could be achieved. When increasing the desired number of selected features, the performance of channel and feature/channel selection gets closer, but for most of the cases feature/channel selection achieved a higher classification accuracy. Similar to channel selection, the performance of feature/channel selection started to decline when selecting large number of features, which is due to the generalization problem. These findings support the fact that feature/channel selection is more general than channel selection, and only due to the increased complexity of the search, we could find in few cases that channel selection achieved better results.

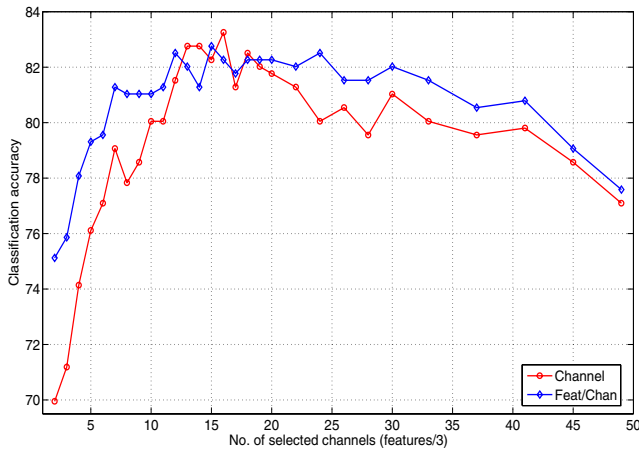


Fig. 4. Classification accuracy of the selected channel and feature/channel subsets

#### D. Motor Versus Non-motor Channels

In this experiment, we divided the 56 channels into two sets. The first set consists of 27 channels in and around the motor area (inside the dashed rectangle in Fig. 1). The remaining 29 channels, which will be referred to as the non-motor channels, are used to form the second set. Similar to the previous two sections, each channel is represented by the chosen 3 features described in section IV-A.

Individual features from different channels were selected from each set by varying the desired number of features between 6 and 78. The obtained results, shown in Fig. 5, confirm that motor channels are more important than non-motor channels. The figure also indicates that non-motor channels can provide useful information about the classification task, which is evident by achieving classification accuracy well above the random guess probability of 50%. It has also been found that the classification accuracy of the best subset of 6 features extracted from the motor channels was 73.89% compared to 75.12% for features extracted from all channels. Nevertheless, for most cases, the performance of the selected features from motor channels was found to be better than that of features extracted from all channels, with a maximum classification accuracy of 84%. This is because of the increased complexity of the search when considering all channels, which makes it difficult for the genetic algorithm to find the optimal solution.

#### V. CONCLUSIONS AND FUTURE WORKS

We investigated in this paper the importance of channel and feature selection on EEG classification and particularly in relation to the brain computer interface problem. We have used the discrete wavelet packet transform to process the signals and have shown that a good compromise between performance and computational cost could be achieved by using three features that represent the energy of frequency bands 4–8, 8–16 and 16–24 Hz. It has also been found that motor imagery channels play an important role in the analysis of BCI, however, including other channels as well could

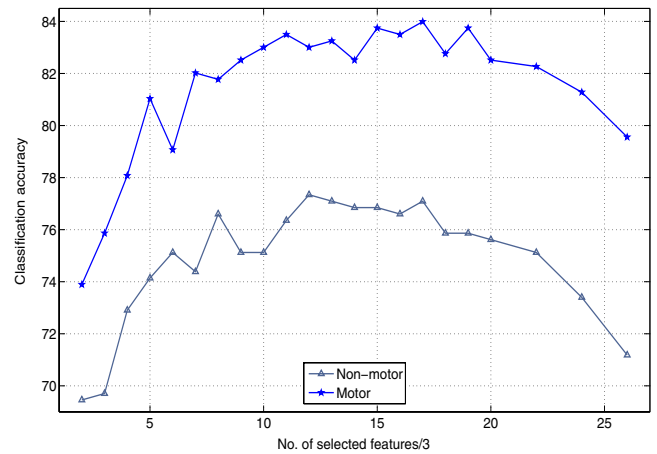


Fig. 5. Classification accuracy of selected features from motor and non-motor channels

provide additional information that would help in improving the performance in certain cases. The obtained results also prove the advantage of selecting individual features from different channels, especially when selecting limited number of features.

This paper have underlined the importance of channel and feature selection on EEG classification. We are planning to further validate the findings on a larger dataset and apply this concept to other EEG problems.

#### VI. ACKNOWLEDGMENT

The authors would like to thank the Department of Medical Informatics, University of Technology, Graz, Austria for providing the data.

#### REFERENCES

- [1] W. Xu, C. Guan, C.E. Siong, S. Ranganatha, M. Thulasidas and J. Wu, High accuracy classification of EEG signal, *Intl. Conference on Pattern Recognition*, 2004, pp 391-394.
- [2] P. Sykacek, S. Roberts, M. Stokes, E. Curran, M. Gibbs and L. Pickup, Probabilistic Methods in BCI Research, *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 11, 2003, pp 192-195.
- [3] N.F. Ince, A. Tewfik and S. Arica, Classification of Movement EEG with Local Discriminant Bases, in *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, 2005, pp V414-416.
- [4] M. Pregenzer and G. Pfurtscheller, Frequency component selection for an EEG-based brain to computer interface, *IEEE Trans. on Rehabilitation Engineering*, vol. 7, 1999, pp 413-419.
- [5] D. Garrett, D.A. Peterson, C.W. Anderson and M.H. Thaut, Comparison of linear, nonlinear, and feature selection methods for EEG signal classification, *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 11, 2003, pp 141-144.
- [6] T.N. Lal, M. Schrder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer and B. Schalkopf, Support Vector Channel Selection in BCI, *IEEE Trans. on Biomedical Engineering*, vol. 51, 2004, pp 1003-1010.
- [7] R. Palaniappan, R. Paramesran, S. Nishida and N. Saiwaki, A New BrainComputer Interface Design Using Fuzzy ARTMAP, *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 10, 2002, pp 140-142.
- [8] B.O. Peters, G. Pfurtscheller and H. Flyvbjerg, Automatic Differentiation of Multichannel EEG Signals, *IEEE Trans. on Biomedical Engineering*, vol. 48, 2001, pp 111-116.
- [9] V.J. Samar, Wavelet Analysis of Neuroelectric Waveforms, *Brain and language*, vol 66, 1999, pp 1-6.
- [10] J. Yang and V. Honavar, Feature subset selection using a genetic algorithm, *IEEE Trans. on Intelligent Systems*, vol. 13, 1998, pp 44-49.