

Design and Implementation of an Online Brain Computer Interface System

M. Raza Fraz¹, Zulqarnain Zahid¹, Shair Shah²

¹National University of Sciences and Technology, Islamabad, Pakistan;

² University of West of England, Bristol, United Kingdom;

Abstract

Brain Computer Interface is the communication channel between the brain and the computer for recording of electrical activity along the scalp produced by the firing of neurons within the brain. The brain signals which are also known as Electroencephalography (EEG) can be used to direct and control some external activity. This work reports a methodology for acquisition and detection and of EEG signals, and extraction of useful information in order to differentiate the signals related to particular type of movement. A modified Common Spatial Pattern (CSP) algorithm has been used at preprocessing stage. Logarithmic transform along with the information theoretic feature extraction has also been used for feature extraction. KNN, SVM and Artificial Neural Networks are employed for classification. The proposed methodology is tested on publically available data sets and the results are found to be comparable with the published approaches.

Keywords—Brain Computer Interface, Common Spatial Pattern, Electroencephalograph (EEG), K-Nearest Neighbor, Support Vector Machine (SVM), Neural Network.

1. INTRODUCTION

BRAIN COMPUTER INTERFACE (BCI) is a communication channel between the brain and the computer which is used to direct and control several types of external activities like controlling some machinery without involving any kind of physical movements. BCI is an emerging field now days and it has vast applications in the field of medical sciences. Patients suffering from severe mental impairments which restrict their movements can make use of this machine to make their mobility possible and easy without dependence on others. This paper focuses on the detection of Brain electrical activity produced by firing of neurons within the brain i.e. the EEG signals and then the overall prediction of the imagined movement.

EEG signals are recorded by measuring the electrical activity of brain using electrodes placed along the scalp. This electrical activity is basically due to electrical fringing of neurons. So

BCI uses EEG signals and works on the surface information. As this is a non-invasive method, the signals recorded are subjected to surrounding noise. Therefore proper filtering and amplification of these signals is required. The first step is of signal acquisition from the surface so they are highly distorted by noise and ocular artifacts. The raw EEG data is preprocessed to get the required brain signals. It's very important to extract the discriminative features from signal, to classify the activity related tasks. After pre-processing and feature extraction classification is done on the basis of feature vector, which is then used to generate control commands to control any external machinery (e.g. wheel chair).

The algorithm frequently used for preprocessing of the data before applying feature extraction and classification is Common Spatial Patterns (CSP) algorithm. CSP was first proposed in 1991 to detect the abnormalities in the EEG signals. Later in 2000, it started to be used for detecting the event related de-synchronizations (ERD) [1]. The imagination of a limb movement modifies electrical activity of brain. The change in this electrical activity is different for imagination of different movements. For example in case of real or imagined hand movement, event related de-synchronizations can be observed in the electrical activity of brain [2]. In case of one-sided hand movement imagination changes the EEG signals recorded from contra and ipsilateral central areas. The patterns of electrical activity so obtained will be different for those obtained in case of left hand movement or any other limb movement. This difference in the electrical activity is not visible if we just look at the signal by naked eye. But with the help of certain algorithm, the two movements can be separated. This forms the basis of CSP algorithm [3, 4].

First we decompose raw EEG data into spatial patterns for the two classes. Then we calculate spatial patterns in such a way so as to maximize the ratio of the variance of data conditioned on one class to the variance of data conditioned on the other class. In this way, spatial filters are designed to extract those components of the EEG data that differs maximally (in terms of variance) between conditions. These spatial patterns are then used to extract features on the basis of log transformation. We proposed a modification in original CSP algorithm that is spatial filters are extracted on the basis of mutual information between the classes. To make selection of the subset of the spatial filters optimum in terms of minimum classification error we are proposing a theoretic framework that is information theoretic feature extraction. In this way not only two classes but multiple classes can be discriminated from each other and classified. We also proposed optimization of certain parameter along with the choice for the optimum classifier which leads to improved accuracies which are further illustrated in detail in this paper.

The paper is organized as follows. In Section II, datasets used and their processing is explained. Extractions of spatial features are explained in Section III. The next section describes the methods used for their classification and proposes the most suitable classifier with CSP. Parameter optimization for the improvement of classification accuracies and results are presented in Section V. Finally, Section VI is dedicated to the presentation of the conclusions of this work.

2. DATA ACQUISITION AND SIGNAL PROCESSING

2.1 The Datasets

1.1.1 BCI competition IV

First dataset used (dataset#1 BCI competition IV)[5] was provided by Berlin institute of Technology. These datasets were recorded from healthy subjects in whom motor Imagery was performed without Feedback. For each subject two classes of motor imagery were selected from the three classes“ left hand, right hand, and foot.

1.1.2 Dataset IVa BCI Competition III

Second dataset used (dataset IVa BCI Competition III) [6] was provided by Berlin BCI group: Fraunhofer First, Intelligent Data Analysis Group (Klaus Robert Muller, Benjamin Blankertz) and campus Benjamin Franklin of the Charite University Medicine Berlin .The datasets are for two class motor imagery (right hand and foot) taken from 5 subjects. Training samples provided are less as compared to the testing samples. Data was recorded from 118 EEG channels at 1000Hz sampling rate .We used the down sampled to 100Hz version of the dataset.

1.1.3 Dataset 2a BCI competition IV

Dataset 2a, BCI competition IV [12], which is a four class dataset, was provided by Institute for Knowledge Discovery, Graz University of Technology, Austria and Institute for Human-Computer Interfaces, Graz University of Technology, Austria. This includes; Class-I (Right hand movement), Class-II (Left hand movement), Class-III (Both feet movement), and Class-IV (tongue).

2.2 Organization of Raw EEG data

Before the implementation of CSP [7] raw EEG data is filtered between 8-30Hz in order to remove the artifacts caused due to eye movements also known as EOG artifacts[8]. Filtration greatly improved the accuracies obtained. Data is filtered using the Butterworth filter of order 10. Two types of filtration is done i.e. Low pass filtering and Band-Pass filtering.

For analysis raw EEG data is arranged into E-matrices. Each E-matrix is of size (N x T), where “N” is the number of channels and “T” is the number of EEG samples per channel in a specific interval of time “T”. So, in this way the raw EEG data is first arranged into structures.

2.3 Co-variance Matrix Formation

Normalized spatial co-variance is obtained by implementing the following equation

$$C = \frac{EE^T}{\text{trace}(EE^T)} \quad (1)$$

Where E^T is the transpose of the “E matrix” and “trace (EE^T)” is the sum of diagonal elements of EE^T. This C is calculated for both classes data individually. For example for the first dataset “ C_l ” is for left hand imagination data and “ C_r ” is for the data related to the imagination of right hand class. They are averaged over each class. For the first dataset case “ C_l ” and “ C_r ” are both of order (59 x 59). The overall composite spatial co-variance is given by

$$C_c = \bar{C}_l + \bar{C}_r \quad (2)$$

Now “ C_c ” is factorized into eigenvalues and eigenvectors.

$$C_c = U_c \lambda_c U_c^T \quad (3)$$

Where “ U_c ” is the matrix containing Eigenvectors and λ_c is the diagonal matrix containing the eigenvalues. all the eigenvalues are arranged in the descending order.

2.4 Whitening Transform Projection Matrix and Spatial Patterns Calculation

After this we applied the whitening transformation

$$P = \sqrt{\lambda_c^{-1}} U_c^T \quad (4)$$

This whitening transform equalizes the variances in the space that is created by U_c . If we calculate $P C_c P^T$ it will result in 1. This will show that up till now our method is correct and we have successfully maximized the variance. After this we have to find

$$S_l = P \bar{C}_l P^T \quad (5)$$

Similarly

$$S_r = P \bar{C}_r P^T \quad (6)$$

Another test to check the procedure is if we take $S_l = B \lambda_l B^T$ and $S_r = B \lambda_r B^T$ than $\lambda_l + \lambda_r = I$ (identity matrix). Now as we know the sum of the corresponding eigenvalues is always 1, so this test indicates that when λ_l is maximum λ_r is minimum. This indicates that eigenvalues for one class will be maximum at a point whereas the other class will have eigenvalues minimum at that

same point. So we are successful in maximizing the co-variance between the two classes. Now B was the eigenvectors. This distribution in covariance makes it possible to classify the eigenvectors belonging to the two different classes. Now in order to find the feature vectors we have to find the projection of “P” (whitening transform), onto the first and last eigenvectors in B that are corresponding to largest λ_l and λ_r . The projection matrix is given by

$$W = (B'P') \quad (7)$$

2.5 Joint approximate diagonalization

In CSP for two class, diagonalization of two covariance matrices is done. Now for M number of classes eigenvectors are combined in a W matrix such that

$$W^T R_{x|c_i} W = D_{c_i} \quad (8)$$

Where, “R” is the covariance matrix, c_i represents to which class it belongs.

This “W” matrix is then used for diagonalizing of covariance matrices of multiple classes. Once this transformation is found some columns of this “W” matrix are selected as spatial filters. Up till now it remains ambiguous that which columns are selected as spatial patterns, which will give the most optimum result. For this purpose, the eigenvalues are computed of all the covariance matrices, then eigenvectors which corresponds to the largest eigenvalues are selected. In a case that any eigenvector is selected more than one times that is it corresponds to more than one largest eigenvalues than the eigenvector with next eigenvalue is selected.

3. FEATURE EXTRACTION

3.1 Information Theoretic Feature Extraction

To make selection of the subset of the spatial filters optimum in terms of minimum classification error we are proposing a theoretic framework named as “information theoretic feature extraction”. [9].

First of all covariance matrices of each class is calculated as.

$$C_i = \frac{E_i E_i'}{\sum_N E_i E_i'} \quad (9)$$

Where $i=1, \dots, M$ and N is the number of class

These Covariance matrices are then joined in a single matrix R, which is then computed to get transformation “W”. First of all an update matrix “W” is computed using A Fast algorithm for Joint diagonalization with Non-orthogonal Transformations generally known as “FFDIAG” . For the computation of this update matrix following equations are used.

$$Z_{ij} = \sum_k C_i^k C_j^k \quad (10)$$

$$Y_{ij} = \sum_k D_j \frac{E_{ij}^k + E_{ji}^k}{2} \quad (11)$$

$$W_{ij} = \frac{Z_{ij}Y_{ji} - Z_{ii}Y_{ij}}{Z_{jj}Z_{ii} - Z_{ij}^2} \quad (12)$$

$$W_{ji} = \frac{Z_{ij}Y_{ij} - Z_{ji}Y_{ji}}{Z_{jj}Z_{ii} - Z_{ij}^2} \quad (13)$$

Then normalization on this update matrix is performed, for this first scale “W” by power of two so that its norm is less than one. The orthogonality of update matrix is ensured by exponentially update of a matrix.

$$V_{n+1} = e^{W_n} V_n \quad (14)$$

Where “W_n” is a skew matrix, i-e $W = -W^T$. Once we get this update matrix, there comes the selection of spatial filters for optimum results.

Then Gaussian mutual information is computed.

$$I_g = \log(w' R_x w) \quad (15)$$

Then estimation of negentropy is performed and subtracted from mutual information, and then “N” number of columns of transformation matrix with highest mutual information is selected as most optimum spatial filter to get the final transformation matrix. This transformation matrix is then used to calculate features.

3.2 Logarithmic transform

The features are calculated by decomposition of trails of “E” and can be given as

$$Z = WE \quad (16)$$

Where, “W” was the projection matrix.

Now for each class EEG sample matrix we are going to select only small number of signals say ‘m’ that are most important for discrimination between the two classes. As mentioned earlier the discrimination is achieved on the basis of maximized covariance. We are going to select “Zn” signals only that will play the most important role in maximizing the covariance, where $n=1, \dots, 2m$. These will be associated to largest eigenvalues (λ_l, λ_r) . We will be taking the m first and m last rows of Z so making the total dimension 2m. The feature vectors can be calculated by the following equation.

$$f_p = \log\left(\frac{\text{var}(Z_n)}{\sum_{i=1}^{2m} \text{var}(Z_i)}\right) \quad (17)$$

4. . CLASSIFICATION

Three different types of classifiers were used to evaluate the classification accuracies for two, three and four class data. The projection matrix " $W_{N \times D}$ " calculated from the training data is used to plot the features for the testing data. Based on the graphs obtained for features the kNN gives the best classification accuracies as compared to the SVM and Neural Network Classifier [10]. Results of different classifiers on dataset IVa of BCI competition III are shows in the *Figure 2*

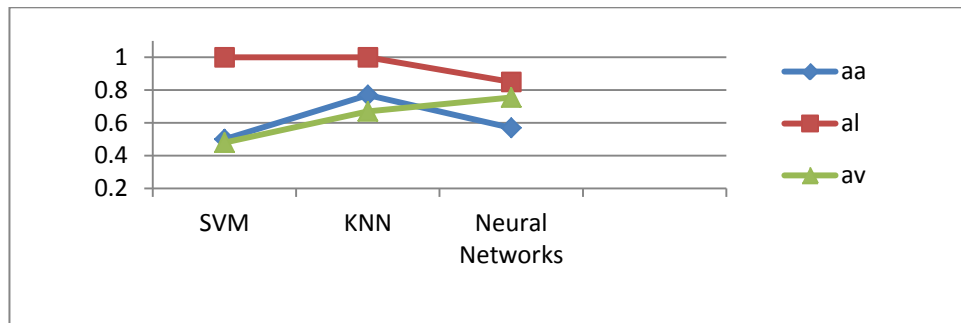


Figure 2: Classification Accuracies using Three Different Classifiers

Hence the KNN classifier was used for two, three and four class classifications and the maximum number of neighbors was selected as five. Parameter optimization is also done to further improve the classification accuracies. All these aspects are further discussed in the subsequent section.

5. RESULTS

Three different datasets taken from the BCI competitions are used for classification of movement related features.

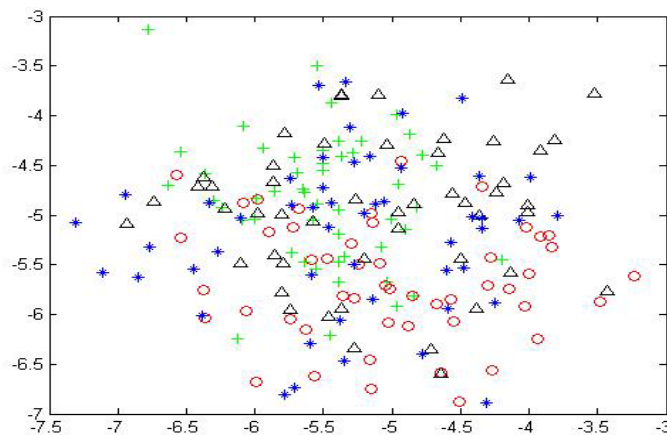
5.1 Two Class Analysis

Dataset IVa, provided by BCI competition III is used for the evaluation of two classes. This dataset contains EEG data from five healthy subjects. The data was recorded without feedback. Visual cue instructing a specific movement was presented for a period of three and a half seconds. After this the movement imagination signal was recorded for three seconds. Data was recorded for the imagination of three motor actions, but only two were released on internet. These two were the imagination of right hand movement and the foot movement. The data was recorded from 118 EEG channels [11]. Total two hundred and eighty samples were recorded and they were divided into training and the testing data.

Table 1: Division of dataset in different subjects

Subjects	aa	al	av	aw	ay
Number of training samples	28	56	84	224	168
Number of test samples	252	224	196	56	112

As mentioned earlier the best results were obtained when the KNN-Classifier was used along with features extracted using log transformation from the selected spatial patterns with maximum mutual information. The extracted features for the two class analysis are shown in Figure 3, where *circles* represents training class 1, '+' represents training class2, *triangles* represents testing class 2 and '*' represents testing class 1.

**Figure 3: Feature separation**

Proposed method is applied instead of original CSP algorithm which results in improvement of accuracies. This improvement in feature separation is illustrated in Figure 4. The dignity of proposed method allows us to use KNN classifier, which is simplest to implement as compare to other two classifiers.

5.2 Parameter Optimization for Classification Accuracies Improvement

For the analysis all the data was divided into E-matrices such that $E_{N \times T1} \in R_{N \times T}$ where "T1" is a subset of "T". In this way all the EEG data taken from the dataset was divided into equal matrices. Moreover, it was found that if the starting points of the window (Km) then the overlaps in time between the windows and window size (total sample points in window) was varied, and its effect on results is illustrated in Figure 4.

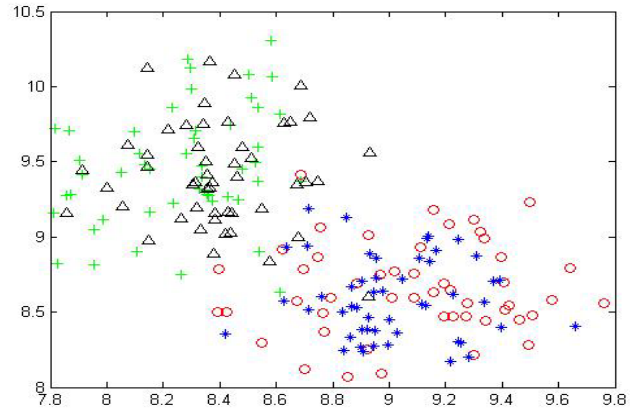


Figure 4: Feature separation

By sliding the window for a trail we select only that portion of the overall imagination trail, which gives the best classification for a given class. Furthermore, the classification accuracies are also affected by varying the dimension N of the feature space which actually corresponds to the number of columns of transformation matrix with highest mutual information. The factor “ k ” number of nearest neighbor used in classification is also optimized and best classification is obtain when $k = 2$, as illustrated in Figure 5.

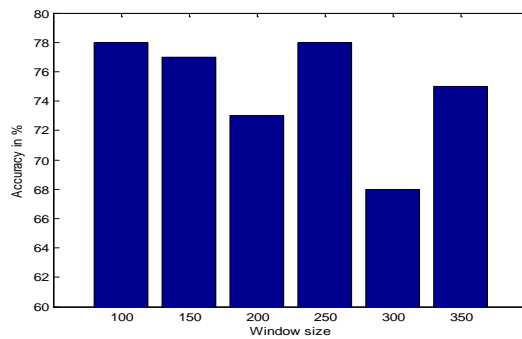


Figure 5: Effect of varying window size on accuracy.

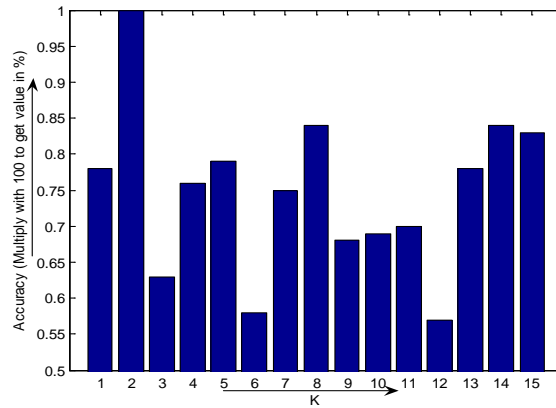


Figure 6: Effect of varying k (number of nearest neighbors) on accuracy

Table 2 shows the best accuracies obtained by varying window size of different subjects.

Table 2 : Best window size accuracies for different subjects

Subject	Window Size	N	Number of Test Samples	Accuracy
aa	100	2	112	77.7%
al	200	2	56	100%
av	250	2	196	75.5
aw	150	2	224	74%
ay	200	2	252	88.9%

Table 3 : Effect of Varying N for subject aa. Best results are obtained for N=2.

Number of training samples	Number of testing Samples	Window Size	Dimension N of feature space	Accuracy
140	140	350	2	92.85%
168	112	350	2	90.476%
168	112	350	6	82.14%
168	112	350	4	80.95%
168	112	350	8	79.76%
140	140	350	118	79.76%

During the analysis it is also observed that if clustering of the data is good then the values of “k” (selected nearest neighbors) did not much affect the classification accuracies. For the subject ‘aa’ the value of “k” was varied from 3 to 21 and the accuracies were same keeping all the other parameters unchanged. Analysis was also performed to check whether the classification accuracy depends on subjects or this algorithm is subject independent. Results

obtained for two class analyses are summarized in the Table 4, which illustrates that the algorithm is subject dependent.

Table 4 : Subject independent case

Training Subject	Testing Subject	Samples		Accuracy
		Training	Testing	
aa,al,av,aw	ay	532	252	65.6%
aa,al,av, ay	aw	504	224	52.2%
aa,al,aw,ay	av	476	196	59.6%
aa,av,aw,ay	al	336	56	50%
al,av,aw,ay	aa	329	112	44.6%
aa,al,av,aw,ay	aa,al, av,aw,ay	560	840	70%

5.3 Multiclass Analysis.

Similar analysis was done for the increased number of classes. KNN-Classifer was the best classifier and again parameter optimization improved the accuracies. Three classes of dataset “2a BCI competition IV” are selected for evaluation and parameter optimization. The optimized value of window size was taken as 500. Similarly the value of “N” (number of spatial patterns) was selected to be two and four. The value of K for nearest neighbor classifier was five. The results of three class analysis using the mentioned parameters are summarized in Table 5.

Table 5 : Results for three class analysis with equal training and testing samples

Subject	Correctly Detected Classes	Total Training Samples	Total Testing samples	Accuracy
A01T	90	108	108	83%
A03T	95	108	108	87.96%
A04T	65	108	108	60.18%
A07T	88	108	108	81.48%
A08T	95	108	108	87.96%

The nobility of this work is that analysis with greater amount of testing samples than the training samples can be done, which is illustrated in Table 6.

Table 6 : Results for three class using unequal training and testing samples

Subject	Correctly Detected Classes	Total Training Samples	Total Testing Samples	Accuracy
A01T	107	90	126	85.50%
A03T	102	105	111	91.89%
A08T	96	96	120	80.46%

It is observed that different accuracies are obtained for different subjects. Hence it signifies that this algorithm is subject dependent and accuracies greatly depend on the subject's concentration. It is also observed that in this dataset when only three movements imagination (Right hand, left hand, both feet) were considered results were good but including the fourth class i.e. the imagination of tongue movement greatly degraded the accuracies. This can be observed in the Table 7.

Table 7 : Accuracies for four classes

Subject	Training Samples	Testing Samples	Correctly Classified	Accuracy
A01T	144	144	121	84.02%
A03T	144	144	99	68.75%
A04T	144	144	72	50%
A08T	144	144	112	77.77%

5.4 Effective Electrode Reduction.

Using Multiclass CSP algorithm along with information theoretic feature extraction as well as log transformation and KNN classifier resulted in effective reduction of electrodes used to get the brain signals. It is observed that if we optimize the values for the parameters: features space dimension (N), number of nearest neighbors and sliding window, the number of electrodes can be reduced. The offline datasets used for two class analysis used 59 and 118 electrodes (channels) to record the brain signals. In feature space optimization we reduced the dimension space and selected the optimum value of N that gave us the best results (effect of varying N is demonstrated in

Table 3). Actually here we were selecting the minimum number of electrodes required to give best accuracies. It is observed that if we implement the whole procedure on data collected from lesser number of electrodes the results are comparable to those obtained from 118 or 59 numbers of electrodes. Accuracies can be further improved by selecting optimum value of "k"

(number of nearest neighbors), greater number of training samples as compared to testing samples and by utilizing the concept of window sliding. Table 8 shows the results when only three electrodes placed at C3, C4 and reference CZ were used to predict the right hand and left hand movement.

Table 8: Classification accuracies using only 3 electrodes

Training Samples	Testing Samples	Correctly detected	Confused	Accuracy
50	6	4	2	66.66%
120	50	33	17	66%
120	6	5	1	83%
120	6(all left)	6	0	100%
120	6(all right)	6	0	100%

This is a subject dependent case and training along with parameter optimization is done every time there is a new subject. So here we get improved accuracies with lesser number of electrodes making it cost effective, but with a disadvantage of greater training requirements. It is suggested that if two movements can be predicted using 3 electrodes only, then 4 movements can be predicted using at least 6 electrodes. Accuracies can be improved by recursive parameter optimization.

6. CONCLUSION

Proper filtration of the EEG data gives significant improvement in the results. The filtration is done in between 8-30Hz, to filter out any kind of artifacts. We are selecting this bandwidth as we are interested in only the alpha α and the beta β region. In this way we train our algorithm to detect the event related de-synchronization, i.e. the change from Mu-Rhythm (α) to active frequency region when the mind is alert/working (β), or in simple words imagining any movement. Filtration is also improvement as it helps the algorithm to get trained on the movement's imagination rather than any artifacts caused by physical movements specifically eye movement. In this way correctly trained algorithm gives good accuracies for testing.

If information theoretic feature extraction is used for the selection of spatial patterns with maximum mutual information along with proper filter implementation and log transformation for features, it gives better separation in feature plotted on a feature space for two classes as compared to conventional CSP algorithms. The features separation can also be improved if we optimize the value for dimensions of the feature space. By dimension N, we actually select only N rows of the spatial patterns that are extracted from a sample of EEG data on the basis of maximum mutual information. It was found that instead of going towards higher dimensions

lower values of N also give comparable results with decreased complexity in computations as well. Decreased values of N also assist easier classification algorithm selections. This helps to overall reduce the complexity of design. If N is varied in turn we are selecting lesser number of channels. So it also gives decreased hardware complexity. It was found that rather than using higher number of channels, if only two channels C3 and C4 along with Cz as reference is used it also gives the comparable accuracies. Thus selection of lower value of N not only improves the features representation but also provides an overall simplicity to the design. In this way KNN-classifier also becomes the most effective classifier with CSP in terms of simplicity and accuracies.

Concept of window sliding helps us to select the most optimum window from the training data which gives the best predication of the class label. This window selection is subject dependent as it is selecting a window in which a particular subject EEG gives best results. So window sliding gives improved accuracies but with a constraint of being subject dependent. However if the parameters are first optimized before testing and are then keep fixed for all the subjects then this can help us to choose most appropriate window that can be applied for every EEG sample and get good results.

In summary comparable results can be obtained by a simpler implementation if we first filter the EEG samples, then apply information theoretic feature extraction with lower dimension N instead of conventional CSP and then apply KNN-classifier along with parameters optimization.

REFERENCES

- [1] G. Pfurtscheller, "Graphical display and statistical evaluation of event related desynchronization (ERD)," *Electroenc. Clin. Neurophys.*, vol. 43, pp. 757–760, 1977.
- [2] S. C. Gandevia and J. C. Rothwell, "Knowledge of motor commands and the recruitment of human motor neurons," *Brain*, vol. 110, no. 5, pp. 1117–1130, 1987.
- [3] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG based discrimination between imagination of right and left hand movement," *Electroenc. Clin. Neurophys.*, vol. 103, no. 5, pp. 1–10, 1997.
- [4] H. Ramoser, J. Mueller-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.
- [5] Cedric Gouy-Pailler, Jeremie Mattout, BCI Competition IV, Dataset 1: Motor Imagery, uncued classifier Application, August 2008, France.
- [6] Benjamin Blankertz, Klaus-Robert Muller, The BCI Competition III: Validating alternative Approaches to Actual BCI Problems, *Neural and Rehabilitation Engineering*, 2006

- [7] Hyohyeng Kang, Yunjun Nam, Seungjin Choi, Composite Common Spatial Pattern for subject to subject transfer, IEEE Signal Processing letters, Vol. 16, No-8, pages. 683-686, August 2009
- [8] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG based discrimination between imagination of right and left hand movement," *Electroenc. Clin. Neurophys.*, vol. 103, no. 5, pp. 1–10, 1997.
- [9] Moritz Grosse-Wentrup, Martin Buss, Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction, IEEE transactions on Biomedical Engineering, Vol. 55, page. 1991-1999, August 2008
- [10] Artificial Neural networks and digital signal processing, danikos, 2007-2010: <http://www.learnartificialneuralnetworks.com/>
- [11] C.Brunner, R. Leeb, G.R. Muller-Putz, A. Schlogl, and G. Pfurtscheller, BCI competition 2008, Graz dataset A, Institute of Knowledge Discovery, Graz University of Technology, Austria, and Institute for Human-Computer Interfaces, Graz University of Technology, Austria.
- [12] Benajmin Blankertz, Klaus Robert Muller, Berlin BCI Group, Datasets-1V(a) BCI IV Description, (motor Imagery, uncued classifier Application), 2008