DUJASE Vol. 4(2) 83-89, 2018 (January & July)

A Comparative Analysis on Feature Extraction and Classification of EEG Signal for Brain-Computer Interface Applications

Mohammed Nowaz Rabbani Chowdhury and Subrata Kumar Aditya

Department of Electrical and Electronic Engineering, University of Dhaka, Dhaka-1000, Bangladesh

E-mail: nowazrabbani@rocketmail.com, aditya@du.ac.bd

Received on 03.05.18. Accepted for publication on 08.08.18

ABSTRACT

Classification of EEG signal for Brain-Computer Interface (BCI) applications consists of three stages: *Pre-processing; Feature extraction* and *Classification*. There are different methods implemented in these stages found in existing literature. However, the performance of the methods has been measured on different datasets which made the results incomparable to each other. To address this problem, in this paper, different combination of feature extraction and classification methods has been implemented to classify a well known dataset (dataset 2A, BCI Competition IV) so that a comparative analysis can be made based on identical platform to find out the best combination of methods. In the pre-processing step, the EEG data was band-pass filtered to remove the artifacts and Common Spatial Pattern (CSP) was applied to increase the discriminativity of the data. Two types of features: *Time Domain Parameters* (TDP) and *Adaptive Auto-Regressive* (AAR) parameters were extracted from the pre-processed EEG signal. The features were classified using two types of classifiers: *Linear Discriminant Analysis* (LDA) and *Support Vector Machine* (SVM). A comparative analysis has been conducted to identify the best combination of feature and classifier. The analysis reveals that, TDP features classified using LDA classifier provides best performance and hence demands application in real time BCI system.

Keywords: Brain-Computer Interface (BCI), Time Domain Parameters (TDP), Adaptive Auto-Regressive Parameters (AAR), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM)

1. Introduction

Brain-Computer Interface (BCI) provides the only way of communication for the people affected by motor disabilities. More formally, BCI is a communication system that doesn't require any peripheral muscular activity. Indeed, BCI systems enable a subject to send commands to an electronic device only by means of brain activity [1]. Not only as the communication tool for the people with motor disabilities, BCIs also have a wide range of other applications. Some of these are videogames and virtual reality, creative expression, access to internet etc [2].

In order to control through BCI, the user must produce different brain activity patterns that will be identified by the system and translated into commands. There is a variety of invasive and non-invasive techniques to measure the brain activity. The non-invasive techniques can be EEG, MEG, PET or fMRI. The invasive technique can be ECoG [3]. However, among all of the invasive and non-invasive techniques, due to its easy implementation facility, high time resolution, real time availability and price constraints, most of the BCI systems use EEG as the measurement technique of brain activity [4].

To identify the measured brain activity patterns, a BCI system relies on an identification algorithm. The goal of this identification algorithm is to automatically estimate the class of data. Usually, the identification algorithm consists of three stages as shown in Fig 1: *Pre-processing* (enhance the separability of classes), *Feature extraction* (extract features of the patterns) and *classification* (classify the features to identify the classes). The performance of the BCI system readily depends on the methods implemented in these stages. Hence, selection of best algorithms in each of these stages is very important in BCI design.

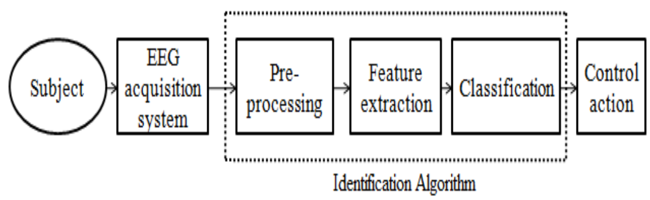

Fig. 1: Structure of a BCI system based on EEG

There are different pre-processing methods for signal enhancement in BCI designs found in literature such as common average referencing (CAR), surface laplacian (SL), independent component analysis (ICA), common spatial patterns (CSP) and principle component analysis (PCA) [5]. CSP enhance the spatial resolution of EEG signals and hence the separability between classes can be increased. Also, selection of optimal spatial filter can reduced the dimensionality of EEG data [6].

One of the important stages in BCI design is the selection of features used to classify the EEG signal. There are different types of features used in BCI systems based on neurological mechanisms used in those systems. However, BCI systems based on sensorimotor activities (ERD/ERS) uses spectral parameters (e.g. band power), auto-regressive parameters (AR), adaptive-auto-regressive (AAR) parameters, multivariate AAR, time-frequency (TF) features, wavelet packets, time domain parameters etc. as features [5].

However, the most important stage of a BCI design is the selection of classification algorithm. There is variety of classifiers found in literature to classify the EEG data for BCI systems. Some of them are different variations of neural-networks (NN), LDA, Quadratic discriminant analysis (QDA), support vector machine (SVM), k-nearest neighborhood (kNN), Bayesian classifiers etc. It was found in the literature that different classifiers performed well with different features [5].

However, one problem with previous research works concerning BCI systems is that very often the performance of the methods are measured using different procedures or different EEG data sets. This makes the obtained results incomparable to each other. To address this problem, in this paper, different methods of pre-processing, feature extraction and classification was implemented to automatically classify four class motor imagery (imagination of left hand, right hand, both feet and tongue movement) EEG signal. Then, the best combination of methods of the three stages was identified as the best algorithm according to a comparative analysis between them.

2. Materials and Methods

2.1 Data Set

The BCI data used in this research work is the Dataset 2a of BCI Competition IV. The data set consists of EEG data from 9 subjects while they were performing four different motor imagery tasks [7]:

1. Movement imaginary of left hand (Class 1)
2. Movement imaginary of right hand (Class 2)
3. Movement imaginary of both feet (Class 3)
4. Movement imaginary of tongue (Class 4)

There were two sessions for recording data from each subject. These sessions were taken place at different days. Each session consists of 6 runs and they are separated by short breaks. Each run consists of 48 trials (12 for each class. Hence, $4 \times 12 = 48$ trials). Hence, in each session there are 72 trials for each class and 288 trials in total.

There are 22 EEG channels in the data set. Also, 3 monopolar EOG channel are included to deal with eye movement artifacts. The duration of each trial is 7.5s with 250 Hz sampling rate. Hence, the dimension of the raw dataset for each trial is 25 channels with 1875 data points. The electrode montage for EEG and EOG channels has been given in Fig 2.

According to the rules of BCI Competition IV, data from one session (for each subject) should be used as training set and other session has to be used as test set. In this research work, the test set was kept completely hidden until the final evaluation experiment. All the analysis was done using the training set.

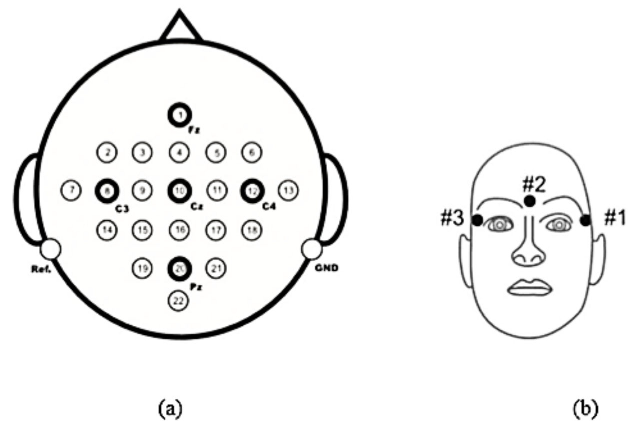

Fig. 2: (a) Electrode montage for EEG channels (b) Electrode montage for EOG channels

2.2 Pre-processing

At first, the EEG signal was pre-processed to remove the EOG artifacts. To increase the separability between classes and to reduce the dimensionality of the data, spatial filtering was implemented in this step.

2.2.1 Band-pass filtering to remove EOG artifacts

During the recording of dataset used in this research, the non-physiological artifact such as line frequency noise was suppressed using a notch filter of 50 Hz. For the purpose of removing the EOG artifacts (due to eye blinking or rolling) EOG data of three channels is provided also. However, there may be EMG artifacts present due to the movement of head, body or facial muscles. There is no information provided about EMG artifacts in the dataset. EOG artifacts have wide frequency range, but they are maximal in 0-4Hz range. EMG artifacts are maximal in the range above 30 Hz [8]. However, the neurological phenomenon utilized in this research is Event Related De-synchronization/Synchronization (ERD/ERS) which is prominent in μ (8-13 Hz) and β (13-30 Hz) band. Hence, there is no overlapping of artifacts with the neurological phenomenon. Therefore, band-pass filtering can be an efficient tool to remove the artifacts. Hence, the data was filtered using a fifth order digital Butterworth filter with the pass-band of 7-30 Hz.

After, band-pass filtering, as there is no EOG artifacts presents in the data, the 3 EOG channels was removed and hence the dimensionality has been reduced to 22 channels with 1875 data points for each trial.

2.2.2 Common Spatial Pattern (CSP)

To increase the discriminativity between classes and to reduce dimensionality of the data Common Spatial Pattern (CSP) was implemented on the band-pass filtered EEG data. Also, as the EOG artifacts were removed, the EOG channels were also removed from the data.

CSP is a technique to analyze multi-channel EEG data based on recordings from two or more than two classes. It maximizes the variance of the spatially filtered signal under one class and minimizes it for other class. Since variance of band pass filtered signals is equal to the band power, CSP increases the discrimination of mental states that are characterized by ERD/ERS [9].

Let, $\boldsymbol{x}(t) \in \mathbb{R}^{N\times T}$ is the multichannel EEG data of N channels and T sample points with $c \in \{c_1, c_2\}$ is the set of classes. CSP finds a linear transformation $\mathbf{W} \in \mathbb{R}^{\mathbf{N}\times\mathbf{L}}$ with $L \leq N$ such that,

$$\boldsymbol{x}_{CSP}(t) = \boldsymbol{W}^T \boldsymbol{x}(t) \quad (1)$$

When $L = N$, there is no reduction of dimensionality. Only the discriminativity will be increased. On the other hand, for $L < N$, dimensionality reduction is also provided along with increased discrimination between classes. Columns of $\boldsymbol{W}$ is represented by $\boldsymbol{w}_j \in \mathbb{R}^N$ where, $j = 1,2,3, \dots . L \leq N$. Each columns of $\boldsymbol{W}$ is called a spatial filter.

However, to calculate the transformation matrix $\boldsymbol{W}$, two covariance matrices are estimated for two classes. Then optimized value of $\boldsymbol{W}$ is computed by simultaneous diagonalization of the two covariance matrices.

However, the simultaneous diagonalization problem can be solved using generalized eigenvalue problem and the dimensionality reduction can be implemented by selecting the spatial filters with smallest/largest eigenvalues.

As the dataset used in this research indicates the implementation of multiclass CSP, Joint Approximate Diagonalization (JAD) was used to solve the multiclass problem [9].

However, to implement CSP along with dimensionality reduction scheme, in this paper, L was taken to 8 and hence, the dimensionality of the processed dataset reduced to 8 spatial filtered channels with 1875 data points.

2.3 Feature extraction

In this research, two different types of features were extracted from the pre-processed EEG data. One of them is time domain feature; another is based on time series analysis. These are: *Time Domain Parameters* (TDP) and *Adaptive Auto-Regressive parameters* (AAR).

These features for different classes show different discriminative properties which is useful for a classifier (at the next step) to classify the data.

2.3.1 Time Domain Parameters (TDP)

Time Domain Parameters (TDP) was first introduced in [10], and considered as more generalized representation of Hjorth Parameters. TDPs are obtained by calculating the variances of derivatives of the signal with different order. For each order of derivatives a TDP can be obtained as,

$$p_i(t) = var\left(\frac{d^i x(t)}{dt^i}\right);\ i = 0, \dots, k \quad (2)$$

However, for better performance, the parameters are smoothed using an IIR filter by implementing the following expression.

$$y(n) = up_i(n) - (1-u)y(n-1) \quad (3)$$

Where, u is called update co-efficient and has an important effect in the performance of TDP as features in BCI.

However, it has been shown in result section that, the optimal value of k is 5. Hence, the size of the feature vector is 40 for TDP features.

2.3.2 Adaptive Auto-Regressive (AAR) parameters

The AAR method is appropriate for on-line and single trial analysis of the time varying EEG spectrum. When there is no averaging of an ensemble of recordings, AAR method is very useful to extract features from EEG signal for BCI classifiers [11].

An Auto-Regressive model is useful for describing the stochastic behavior of an EEG time series. This can be described as,

$$y_k = a_1\, y_{k-1} + \cdots + a_p\, y_{k-p} + x_k \quad (4)$$

$$\text{with, } x_k = N\{0, \sigma_x^2\} \quad (5)$$

Where, x_k is a zero-mean-Gaussian-noise process with variance σ_x^2. The index k is an integer number and describes discrete, equidistance time points. Here, p is the model order and AR parameters $a_1, a_2, \dots, a_p$ of an AR model can be used as features.

However, to consider non-stationarity of the EEG signal, the AR parameters are allowed to vary in time and hence these time-varying parameters are known as AAR parameters. Hence the model is changed to represent by following equation.

$$y_k = a_{1,k}\, y_k + \cdots + a_{p,k}\, y_{k-p} + x_k \quad (6)$$

The AAR parameters were estimated using scalar Kalman filtering.

However, it has been shown in result section that, the optimal value of AAR model order is 6 and hence the size of the feature vector is 48 for AAR features.

2.4 Classification

To classify each trial of the multichannel EEG data or in other words predicting class label for each trial, the extracted features were fed to a classifier. The performance of a BCI system significantly depends on which type of classifier is used. In this research, two different types of classifiers were used to classify the EEG features extracted. These are: *Linear Discriminant Analysis* (LDA) and *Support Vector Machine* (SVM).

2.4.1 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a generalization of Fisher's linear discriminant, and used in statistics, pattern recognition and machine learning to find a linear combination of features that characterizes or separates two or more classes of objects or events [12].

Let, $\boldsymbol{x} = [x_1, x_2, \dots, x_n]^T$ is the feature vector of EEG data of M different classes represented by the set $c \in$

$\{c_1, c_2, \dots, c_M\}$. Then, the discriminant functions are represented as,

$$g_i(\boldsymbol{x}) = \boldsymbol{w}_i^T\boldsymbol{x} + w_{i0} \tag{7}$$

Where, $i = 1,2, \dots, M$

The classification is done as:

If, $g_i(\boldsymbol{x}) > 0$; $i = 1,2, \dots, M$; **Then,** $\boldsymbol{x} \in c_i$; **Otherwise,** $\boldsymbol{x} \notin c_i$

Then training of the classifier is done using one-vs-rest scheme by calculating the optimal value of the weight vector $\boldsymbol{w}$ using Equation 8 for the discriminant function expressed in Equation 7.

$$\widehat{\boldsymbol{w}} = (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)^T\boldsymbol{S}_{pooled}^{-1} \tag{8}$$

Where, $\overline{\boldsymbol{x}}_1$= Mean value of the data of the class for which $\boldsymbol{w}$ to be calculated; $\overline{\boldsymbol{x}}_2$= Mean value of the data for rest of the classes; $\boldsymbol{S}_{pooled}$ = pooled covariance matrix of two classes;

2.4.2 Support Vector Machine (SVM)

SVM classifies linearly separable two class data by constructing an optimal hyperplane so that the margin of separation between the two classes is maximized [13].

Let, $\boldsymbol{x} = [x_1, x_2, \dots x_n]^T$ is the feature vector of EEG data for two different classes c_1 and c_2 which are linearly separable. Then, the mathematical representation of the hyperplane separating the two classes can be represented as,

$$\boldsymbol{w}^T\boldsymbol{x} + b = 0 \tag{9}$$

Where, $\boldsymbol{w}$ is the weight vector and b is the bias. Then, for a given weight vector $\boldsymbol{w}$ bias b, the separation between the closest data point (known as support vector) and the hyperplane representing by Equation 9 is called the *margin of separation*,ρ. The training of SVM is nothing but finding an optimal value of $\boldsymbol{w}$ and bias b so that ρ is maximized. This is done by solving an optimization problem by the *method of Lagrange multipliers*. However, for the data which are not linearly separable, a non-linear mapping to higher dimensional space is necessary. To execute the non-linear mapping, inner product kernels are used in SVM. These kernels are of different types such as polynomial kernel, radial-basis function (RBF) kernel etc. In this research, RBF kernel was used which can be expressed as,

$$K(\boldsymbol{x}, \boldsymbol{x}_i) = \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{x} - \boldsymbol{x}_i\|^2\right) \tag{10}$$

The parameter σ is called kernel width and it has a significant effect on the performance of the classifier.

However, as the EEG data classified in this research consists of more than two classes, one-vs-one scheme of SVM was implemented [13].

2.5 Evaluation

To analyze the performance of BCI systems, some evaluation criteria must be applied. The most popular is accuracy. However, because of some strict prerequisites, accuracy is not always a suitable criterion, and other evaluation criteria have been proposed [14]. In this research work, the Cohen's kappa co-efficient was applied to evaluate the performance of the classifiers with different features.

At first, a confusion matrix as shown in Table 1 was calculated from the output of the classifier, where the rows represent the true classes and the columns represent the predicted classes by the classifier.

Table 1. Example of a Confusion Matrix for M=4 classes

Class	1	2	3	4	Total
1	63	4	2	3	72
2	2	67	2	1	72
3	5	3	58	6	72
4	2	3	3	64	72
Total	72	77	65	74	288

Then, the classification accuracy can be calculated as follows.

$$ACC = p_0 = \frac{\sum_{i=1}^{M} n_{ii}}{N} \tag{11}$$

Where, $N = \sum_{i=1}^{M}\sum_{j=1}^{M} n_{ij}$

Finally, the Cohen's kappa co-efficient, κ was calculated as,

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{12}$$

Where, $p_0 = ACC$ (Overall agreement);

$p_e = \frac{\sum_{i=1}^{M} n_{:i} n_{i:}}{N^2}$ (Chance agreement);

2.6 *Optimization of performance parameters*

The performance parameters (e.g. update co-efficient u and no. of derivatives k for TDP, update co-efficient UC and model order p for AAR, kernel-width σ for SVM) were optimized using cross-validation. However, it is important to notice that during cross-validation only the training set was used.

In this process, for each of the different values of the parameter, the 8-fold cross-validation was implemented by dividing the training set of each subject into 8 subsets, 7 of them were used for training and the rest of the subset was used for test. The process is repeated for 8 times, every time with different subset as test set. Finally, average value of kappa was calculated. In this way, the value of the parameter with highest kappa was chosen as optimal value.

3. Results and Discussion

3.1 Optimized Performance parameters

As stated earlier, the performance parameters were optimized using crossvalidation. These results are presented below.

Figure 3 represents the variation of crossvalidation kappa, k with the change of update coefficient, u for TDP feature extraction. As seen in the figure, crossvalidation kappa decreases with the increase of u and the optimal value of u was found 0.0045.

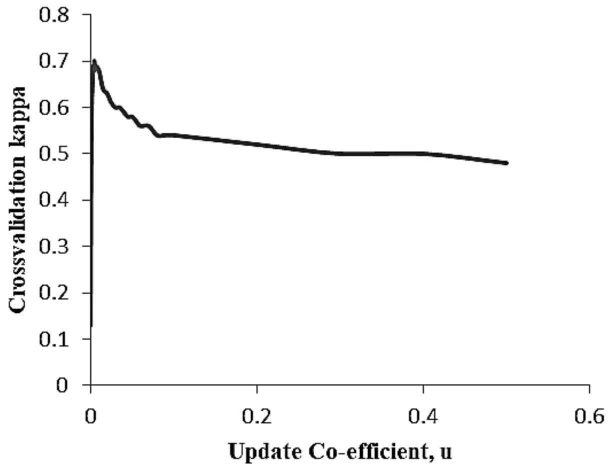

Fig. 3: Change of Crossvalidation kappa with update coefficient, u for TDP features.

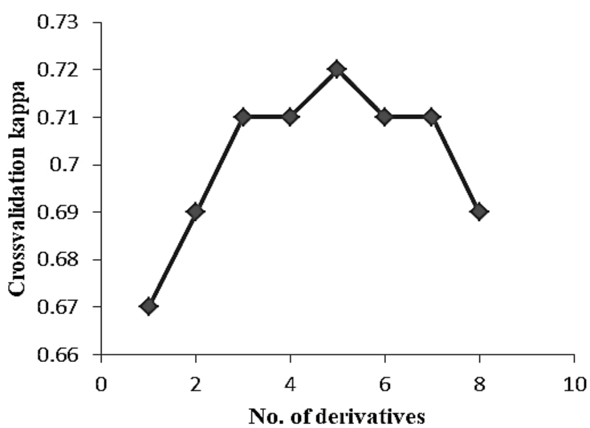

Fig. 4: Variation of Crossvalidation kappa with no. of derivatives, d as TDP features.

Figure 4 represents the variation of crossvalidation kappa with no. of derivatives, d used for TDP feature extraction and the optimal value is 5. It's important to note that, inclusion of higher order of derivatives causes an increase of processing time for TDP feature extraction.

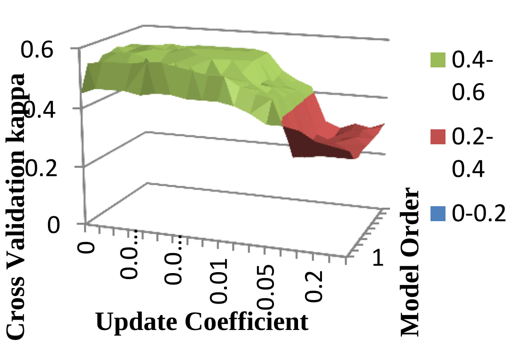

Fig. 5: Surface plot of cross validation kappa with the change of UC and p for AAR features.

Figure 5 represents the variation of crossvalidation kappa with the change of AAR update coefficient, UC and model order, p. The figure suggests that, selection of UC above 0.01 decreases the kappa drastically and also a model order of 4 reduces the performance. However, selection of higher order model doesn't increase the kappa; rather it may increase the computation burden. However, combination of UC and p of 0.004 & 6 respectively gives the highest peak and hence these values were taken as optimal value.

Figure 6 shows the variation of crossvalidation kappa with the change of the kernel width, σ for SVM. It can be seen that, kappa increases with the increase of σ beyond a certain level and then it becomes saturated. The optimal value of σ is 4.

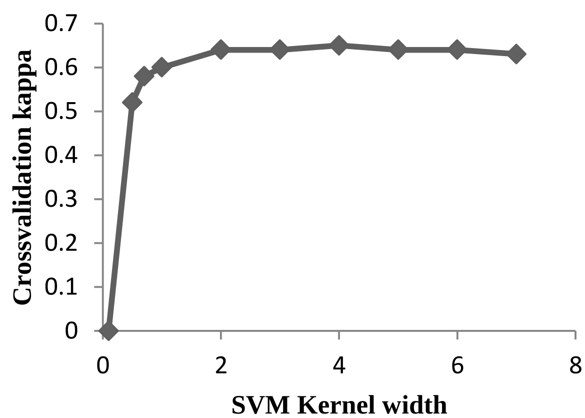

Fig. 6: A plot of Crossvalidation kappa vs Kernel width, σ.

A summary of all the optimal values of performance parameters has been given in Table 2.

Table 2. The optimal value of performance parameters.

Parameter Name	Optimal Value
AAR Update Coefficient, UC & Model Order, p	0.004 & 6
Update Coefficient to extract TDP features, u	0.0045
No. of derivatives extracted as TDP features, d	5
SVM Kernel Width, σ	4

3.2 Final Experiment Results on Test Set

Finally, after selection of optimized parameters, these parameters were used in corresponding methods to classify EEG trials of test set. In these final experiments, the corresponding classifiers were trained using the corresponding features extracted from the training set and then were tested using the corresponding features extracted from test set for each subject. Then, the kappa or accuracy values of test set for all the 9 subjects were averaged to get the mean kappa/ accuracy.

There are three different combinations of features and classifiers were tested in this paper:

1) Features: AAR; Classifier: LDA
2) Features: TDP; Classifier: SVM
3) Features: TDP; Classifier: LDA

A comparison of performance for each of these combinations has been presented on Figure 7.

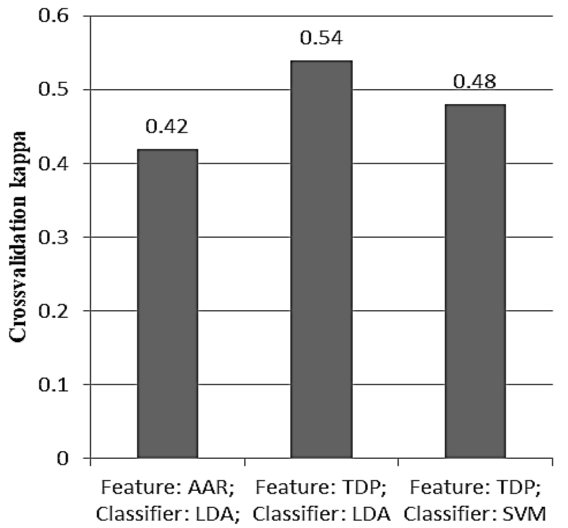

Fig. 7: Comparison of classification performance between the methods implemented.

From the figure, it is seen that, best kappa (0.54) is achieved by LDA classifier with TDP features. The kappa value for AAR features is significantly less than kappa value for TDP features with any combination of classifiers. Also, it was found during the experiment that, it takes much longer time extract AAR features.

Table 3 shows the subject-wise comparison of kappa/ accuracy for each combination of features and classifiers implemented. It can be noticed from the table that, the mean kappa value decreased by the fact that, all of the methods show poor performance to classify the EEG data of subject 2, 5 & 6. This is due to the fact that, BCI illiteracy (inability to use the BCI system) presents in the EEG data of these subjects which can be confirmed by comparing the results with BCI Competition IV for the dataset [15].

Table 3. Subject-wise comparison of implemented methods.

Method	**Classifier:** LDA **Feature:** AAR		**Classifier:** LDA **Feature:** TDP		**Classifier:** SVM **Feature:** TDP	
Subject	Acc. (%)	Kap (k)	Acc. (%)	Kap (k)	Acc. (%)	Kap (k)
Subject 1	76.87	0.69	83.99	0.79	82.21	0.76
Subject 2	49.12	0.32	54.77	0.40	49.82	0.33
Subject 3	63.00	0.51	78.02	0.71	75.09	0.67
Subject 4	47.37	0.30	59.65	0.46	53.95	0.39
Subject 5	39.49	0.19	46.74	0.29	35.14	0.13
Subject 6	43.26	0.24	52.56	0.37	46.98	0.29
Subject 7	67.15	0.56	72.56	0.64	69.31	0.59
Subject 8	60.89	0.48	69.00	0.59	69.74	0.60
Subject 9	62.12	0.49	70.45	0.61	68.56	0.58
Mean Acc./kap	56.58	0.42	**65.30**	**0.54**	61.20	0.48

4. Conclusion

In this research, to classify four class motor imagery EEG signal, two types of features: TDP and AAR were extracted. The extracted features were classified using two types classifiers: LDA and SVM. Before extracting the features, the EEG data were pre-processed using band-pass filtering and CSP. The performance parameters of all of these methods were optimized using crossvalidation. Among the two types of features, best kappa was achieved by TDP features classified using LDA classifier. This suggests the usefulness of these methods (Features: TDP; Classifier: LDA) with improved algorithm in real time BCI system.

References

1. Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J. R., & Birbaumer, N. (2001). Brain–computer communication: Unlocking the locked in. *Psychological bulletin*, *127*(3), 358.
2. L´ecuyer, A., Lotte, F., Reilly, R.B., Leeb, R., Hirose, M. & Slater, M. (2008). Brain-computer interfaces, virtual reality, and videogames. *Computer*, 41(10), 66–72.
3. Hwang, H.-J., Kim, S., Choi, S. & Im, C.-H. (2013). EEG-Based Brain-Computer Interfaces: A Thorough Literature Survey *Internatinal Journal of Human–Computer Interaction*, 29, 814-826.
4. Al-ani, T., & Trad, D. (2010). *Signal processing and classification approaches for brain-computer interface*. INTECH Open Access Publisher.
5. Bashashati, A., Fatourechi, M., Ward, R. K., & Birch, G. E. (2007). A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. *Journal of Neural engineering*, *4*(2), R32.
6. Blanchard, G., & Blankertz, B. (2004). BCI competition 2003-data set IIa: spatial patterns of self-controlled brain rhythm modulations. *IEEE Transactions on Biomedical Engineering*, *51*(6), 1062-1066.
7. Brunner, C., Leeb, R., Muller-Putz, G.R., Schlogl, A. & Pfurtscheller, G. “BCI Competition 2008 – Graz data set A” http://www.bbci.de/competition/iv/desc_2a.pdf
8. Fatourechi, M., Bashashati, A., Ward, R. K., & Birch, G. E. (2007). EMG and EOG artifacts in brain computer interface systems: A survey. *Clinical neurophysiology*, *118*(3), 480-494.
9. Grosse-Wentrup, M., & Buss, M. (2008). Multiclass common spatial patterns and information theoretic feature extraction. *IEEE transactions on Biomedical Engineering*, *55*(8), 1991-2000.
10. Vidaurre, C., Krämer, N., Blankertz, B., & Schlögl, A. (2009). Time domain parameters as a feature for EEG-based brain–computer interfaces. *Neural Networks*, *22*(9), 1313-1319.
11. Schlögl, A. (2000). *The electroencephalogram and the adaptive autoregressive model: theory and applications*. Aachen: Shaker.
12. Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis.: Pearson Prentice Hall. *Upper Saddle River, NJ*.

13. Haykin, S. S., Haykin, S. S., Haykin, S. S., & Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. 3). Upper Saddle River, NJ, USA:: Pearson.

14. Schlogl, A., Kronegg, J., Huggins, J., & Mason, S. (2007). 19 evaluation criteria for bci research. *Toward brain-computer interfacing*.

15. *BCI Competition IV: Results*. [online] Available at: http://www.bbci.de/competition/iv/results/index.html [Accessed 22 Aug. 2018].