**IJCTA,** 9(36), 2016, pp. 443-450


# A Comparative Performance Analysis for Classification of Multiclass Motor Imagery Movements

**Deepa Beeta Thiyam* and Rajkumar E. R.***

***Abstract :*** The effectiveness of Brain-computer interface system (BCI) system is determined by its classification performance. The performance of BCI highly depends on the proper selection of signal processing algorithm and the classifier. It also decreases with the increasing number of discriminating classes. In this paper, we present the performance analysis of Linear discriminant analysis (LDA) and support vector machine (SVM) using two pre-processing algorithms- multiclass common spatial pattern (mCSP) and thinICA-CSP (thin independent component analysis-common spatial pattern) algorithms using the multiclass motor imagery movements EEG signals. From the experiment; we observed that the combination of thinICA-CSP with LDA performs better than the other combination of pre-processing and classifiers methods. In overall, the LDA performs better than SVM for discrimination of multiclass movement for BCI competition IV dataset 2*a*.

***Keywords :*** *m*CSP, ThinICA-CSP, SVM, LDA, BCI

## 1. INTRODUCTION

Electroencephalogram (EEG) measures the electrical activity of the brain by placing electrodes on the scalp. It is one of the most convenient methods for non-invasive measurement of brain activity. The first human EEG was measured by Hans Berger in 1924 (1) . Later, it was used for discrimination of normal and abnormal components. The main disadvantages of EEG are a low spatial resolution because of the volume conduction and nonstationarities characteristics. From the past few decades, it has been commonly used in the field of BCI because of its ability to measure the brain signal from the scalp directly. A person suffering from amyotrophic lateral sclerosis (ALS), multiple sclerosis or totally locked in cannot perform any voluntary movements of body limbs due to damage in the peripheral nervous system but their cognitive functions are not affected. Here comes the role of BCI system. For these group of people, a BCI(2) serves as a communication with the external world without using the muscular activities. The BCI uses different types of control signals such as evoked potential, slow cortical potentials, and sensorimotor rhythms. The evoked potential measures the EEG signals in response to a stimulus. Slow cortical potentials are generated by the slow variation of the cortical activity which lasts for only some milliseconds. Among these, a sensory motor rhythm which denotes the event-related synchronization and desynchronization activity of the brain rhythm is used for motor imagery based BCI (MI-based BCI). The main aim of MI-based BCI is to translate the different motor imagery movements into control commands. It has been known that the imagination of motor movements leads to the changes in brain rhythm in the corresponding sensorimotor cortex regions. This changes in rhythm during the motor imagery movement can be observed in the  and  rhythms of the EEG signals. By considering only this range of frequencies, we can observe the changes in signal band power for the corresponding motor movements which can be decoded as

* Department of Sensors and Biomedical Technology School of Electronics Engineering VIT University, Vellore-632014 Tamil Nadu, India

features for classification of different classes. One of the challenges in for analysis and classification of EEG-based BCI is the frequent changes in the signal properties from time to time and variation of EEG signals from subject to subject. Moreover, the classification performance of the BCI decreases with the increasing number of motor imagery movements. Therefore, it is necessary to overcome these challenges for developing an effective BCI system for real-time applications. One of the approaches is by improving the signal processing unit. A signal processing unit of a BCI system consists of the pre-processing, feature extraction and classification steps. Various signal processing techniques have been proposed and presented for discriminating the motor imagery movements.

Common spatial pattern (CSP) algorithm is one of the most popular algorithms for discrimination of two class motor imagery movements. It maximizes the variance of one class and simultaneously minimizes the variance of the other. However, the CSP is not robust to outliers since its computation depends directly on the covariance matrices. Other variants of CSP algorithms were also proposed to improve the classification performance and to increase the robustness. Some of them incorporated additional EOG signals, prior information, and other subject data to extract the stationary features. Other groups of researcher aim at developing a robust algorithm for multiclass movements. The two class CSP was extended for multiclass discrimination by defining the multiclass problem as binary problems (3). The author of (3) uses the idea of joint approximate diagonalization for discriminating multiclass EEG. This idea was further extended by proposing information theory based feature extraction methods in (4) the question of optimality of CSP in terms of the minimal achievable classification error remains unsolved. Second, CSP has been initially proposed for two-class paradigms. Extensions to multiclass paradigms have been suggested, but are based on heuristics. We address these shortcomings in the framework of information theoretic feature extraction (ITFE. Other author proposed the multiclass filter bank common spatial pattern by using a bank of band-pass filter (5). A different approach for obtaining a multiclass spatial filter using Riemannian geometry and tangent space was proposed in (6). Independent component analysis (ICA) is also one of the blind source separation (BSS) techniques commonly used for separating the multiclass signals into independent sources. It is also used in the field of BCI for as a feature extraction technique. The author of (7) compared the performance of different ICA techniques with multiclass CSP. The classification accuracy of BCI system also depends on the proper selection of the classifier. The classifiers mainly used for discrimination of multiclass MI movements are LDA and SVM. Both LDA and SVM are a linear classifier. However, SVM can produce a nonlinear decision boundary for discriminating nonlinearly separable data using the kernel function. Since we don't have the prior information of the distribution of data, sometimes it is useful to use a non-linear decision boundary for separation of classes.

In this paper, we have presented performance comparison analysis of LDA and SVM classifier using two feature extraction algorithms- *m*CSP and thinICA-CSP. Both LDA and SVM is the most effective classifier common used in machine learning and pattern recognition. The LDA uses the linear decision boundary to discriminate the class whereas SVM with kernels uses non-linear decision boundary for classification of classes. Here, we have used SVM with Gaussian radial basis function kernel (RBF).

## 2. METHODS

The EEG signal processing of MI-based BCI consists of the following steps:1) acquisition of the EEG signals during the motor imagery movements, 2) bandpass filtering of the EEG signal in the particular frequency range and extraction of features for each class, 3) classification of the extracted features into different classes. In this section, we have explained the detailed steps used for the experimental studies in this paper.

**Data set :** In this paper, the experiment was done using BCI competition IV data set 2*a*. The signals were acquired using 22 channels during the motor imagery movements of left hand, right hand, tongue and foot from nine healthy subjects. The dataset consists of two sessions where the signals were acquired on a different day for each session. Each session consists of 72 trials for each class which gives a total of

288 trials for all the class. The signals were sampled at 250 Hz and applied band-pass filtering between the cut-off frequencies of 0.5 Hz and 100 Hz. A notch filter of 50 Hz was added to reduce the noise.

**Pre-processing :** The MI-based BCI decodes the rhythm change of brain signals during the motor imagery process as a control commands. These changes in brain rhythm are observed only in the μ and β rhythm. Therefore, to get only the useful information the acquired signals were band-pass filtered between 8 to 30 Hz which includes only μ and β rhythms. The time segment of 2*s* per trial was selected after the que from the filtered signals for each subject. All the four classes of movements were used for further processing. The extracted signals were used for computation of spatial filters using multiclass CSP and thin ICA-CSP methods.

**Multiclass CSP (*m*CSP) :** The CSP algorithm which is one of the most effective in discriminating the two class motor imagery movements by maximizing the variance ratio of the two classes. Let us consider the acquired EEG signal as $X \in \mathbb{R}^{nXt}$ where, $n$ is the no. of channels and $t$ is the no. of sample points. The covariance matrices of each class can be determined by

$$\Sigma_C = \frac{1}{t}\sum_{i=1}^{t}\frac{X_i X_i^T}{\mathrm{trace}(X_i X_i^T)}, \; c \in (1, 2) \tag{1}$$

The problem can be presented as a Rayleigh quotient and the solution *i.e*. the spatial filters can be computed by solving the generalized eigenvalue problem

$$R(w_i) = \frac{w_i^T \Sigma_1 w_i}{w_i^T (\Sigma_1 + \Sigma_2) w_i} \tag{2}$$

The extracted spatial filters, $W = [w_1 \ldots, w_d]$ can discriminate the two classes. $\Sigma_1$ and $\Sigma_2$ are the average covariance matrices of the two class. T denotes the transpose function. The solution can be obtained by eigenvalue decomposition

$$\Sigma_1 W = (\Sigma_1 + \Sigma_2) WD \tag{3}$$

Where D is the eigenvalues sorted in the descending order. For the multiclass problem, the highest and the lowest eigenvalues are selected and the corresponding eigenvectors represent spatial filters for the discrimination of two classes. This approach was extended for solving a multiclass problem by computing the CSP solution for one class versus the other remaining class (3). In the multiclass approach, the spatial filters for each class are obtained by computing CSP filter between the class and the other remaining class in a similar way. The spatial filters for a respective class are obtained by selecting the corresponding eigenvectors for the lowest eigenvalues. The number of filters is kept equal for all the class. In this paper, we have initialized the total number of filters, $d = 8$ with 2 filters for each class.

**Thin ICA-CSP (8) :** The thin ICA-CSP algorithm extracts the independent components by performing simultaneous joint diagonalization of the second order and fourth order statistic of the observed signals. Let us consider the acquired signal as $\in \mathbb{R}^{nXt}$, which can be considered as a linear model of the mixing matrix, source signals and noise. It can be represented as:

$$x(t) = As(t) + n(t) \tag{4}$$

Where A is the mixing matrix, $s(t)$ is the source and $n(t)$ is the noise term. The whitening of a signal is common in ICA techniques. The whitening transformation enforces the data to have unit variance and uncorrelated. The observed signals are prewhitened using the whitening matrix B. The pre-whitened matrix can be defined as

$$z(t) = Bi(t) \tag{5}$$

Using (4), we can defined as

$$\begin{aligned} z(t) &= BAs(t) + Bn(t) \\ &= Us(t) + Bn(t) \end{aligned} \tag{6}$$

Where U = BA denotes the residual mixing matrix. The estimated sources signals can be obtained by multiplying the whitened signals by the transpose of reduced semi-orthogonal matrix $\widehat{\mathrm{U}} \in \mathbb{R}^{nXd}$. The estimated output signals can be denoted as

$$y(t) = \widehat{\mathrm{U}}^{\mathrm{T}}_{\mathrm{Z}}(t) \tag{7}$$

The main aim of the Thin ICA (9) is to maximize the square sum of the fourth order cumulant of the estimated output. However, for non-stationaries data, it is better to analyze the data by splitting it into smaller blocks and evaluating the contrast function for each block. The variance increases with the statistic order for short data, therefore the estimates of mixing matrix using higher order are not very accurate. In order to overcome this problem, the thin ICA-CSP combined the second order and fourth order statistics of the output signals to obtain the contrast function for estimating the demixing matrix. Since there is no correlation between the symmetrically distributed data, the third order statistic was not included in the computation of the contrast function. Moreover, we computed the second order cross-cumulant of the output signals at different time delay $\tau \in$ T. To make the optimization easier, a contrast function that uses four different estimates of the sources signals is obtained:

$$\begin{aligned}\Phi(\widehat{\mathrm{U}}^{[1]}, \ldots, \widehat{\mathrm{U}}^{[4]}) = \; & \gamma_4 \sum_{c=1}^{c}\sum_{i=1}^{d}\sum_{\tau} |\,\mathrm{Cum}(y_i^{[1]}(t_k), \ldots, y_i^{[4]}))\,|^2 \\ & + \frac{\gamma_2}{3}\sum_{c=1}^{c}\sum_{i=1}^{d}\sum_{\tau} |\,\mathrm{Cum}(y_i^{[1]}(t_k+\tau), y_i^{[2]})(t_k))\,|^2 \\ & + \frac{\gamma_2}{3}\sum_{c=1}^{c}\sum_{i=1}^{d}\sum_{\tau} |\,\mathrm{Cum}(y_i^{[1]}(t_k+\tau), y_i^{[3]}(t_k))\,|^2 \\ & + \frac{\gamma_2}{3}\sum_{c=1}^{c}\sum_{i=1}^{d}\sum_{\tau} |\,\mathrm{Cum}(y_i^{[2]}(t_k+\tau), y_i^{[4]}(t_k))\,|^2\end{aligned}$$

Where $\gamma_4$ and $\gamma_2$ are the weighting proportion of the estimates of fourth order and second order statistics. The maximization of the contrast function is done sequentially by evaluating the gradient function and singular value decomposition of the gradient is performed to update the extraction matrix. Another problem with the ICA method for this application is the selection of MI related components. In order to overcome this problem, the unmixing matrix is initialized with CSP solutions. To have the proper comparison for both the pre-processing methods, an equal number of filters have been used for both the methods.

**Feature extraction :** The extracted spatial filter was used for filtering the EEG signals. The training and testing features were obtained by taking log of the variance of filtered data *i.e.* Feature = $\log(var w_i^{\mathrm{T}} x)$). The logarithmic operation enforced the features to have a normal distribution.

**MI classification :** The extracted training features are used to train the classifier and the classification performance is obtained by using the testing features. In this paper, we have analyzed the classification performance of the two feature extraction methods using two different classifiers. The LDA is commonly used for discriminating the linear data as well as in the field of BCI. Additionally, SVM is also one of the most effective classifiers in the field of pattern recognition. The basic principle of the both the classifiers are described in the below section:

**Linear discriminant analysis: Linear discriminant analysis (LDA) :** LDA (10) is the most commonly used techniques for data reduction and pattern classification because of its closed form solution and easy computation. It is mainly used when the number of observations is larger than the number of features. But this classifier is not effective for the samples with more number of features and less number of samples. The main aim of LDA is to project the multidimensional data into a reduced dimensional subspace with higher class separability. LDA approach mainly considers the data for each class as a model of probability density functions. The class of the input data is determined by the larger value of probability density function from the others.

LDA assumes that all the class has a normal distribution and have the same covariance matrix. Let us consider there are c classes and X = [$x_1$, …, $x_N$] be the samples to be classified, where N represents the no. of samples. The mean, $\mu_c$ for each and the global covariance matrix $\sum$, can be represented as:

$$\mu_c = \sum xcP(x_c) \tag{9}$$

$$\sum = \frac{\sum_{i=1}^{c}(x_k - \mu_k)\sum_{i=1}^{c}(x_k - \mu_k)^T}{N - c} \tag{10}$$

Then, the classification of data point x is done by

$$\underset{c}{\text{argmax}}\; x^T \Sigma^{-1} \mu_c - \frac{1}{2}\mu_c^T \Sigma^{-1} \mu_c \tag{11}$$

Which decision boundaries is a linear function. The LDA is mainly used for binary classification but it can be used for classifying the multiple classes using one versus others methods where one class is classified against the other remaining classes.

**Support vector machine (SVM) (11) :** The SVM approach has a wide application in the fields of machine learning and pattern recognition. It has the ability to deal with high dimensional and non-linear data. The SVM with kernel can generate non-linear decision boundaries which make it suitable for discriminating non-linearly separable data. In this method that data were mapped into a high-dimensional space where the data were spread in such a way that a linear hyper-plane can be fitted. The decision function for kernel-based SVM can be defined by:

$$g(x) = \left(\sum_{i=1}^{N} \alpha_i y_i k(x, x_i) + b\right) \tag{12}$$

Where $x$ = [1,…, N] is the set of training samples, $y$ represents the class labels, $\alpha_i \geq 0$ is a Lagrangian multiplier which is a solution of the quadratic optimization problem, $k(x, x_i)$ represents the kernel and b is the bias. The selection of kernel and setting of the hyperparameters value are important steps in designing SVM classifier. For the experiment in this paper, we have selected Gaussian radial basis function (RBF) which can be defined as

$$k(x, x_i) = e^{-\gamma \|x - xi\|2} \tag{13}$$

Where $\gamma = \frac{1}{2\sigma^2} > 0$, controls the width of the Gaussian function, $\|x - x_i\|$ is the norm of $x$. In this paper, we have used the M-SVM$^2$ method from MSVM package (12) for the discrimination of multiclass MI movements. The M-SVM2 is the extended version of 2-norm SVM for the multiclass approach. The detail explanation can be found in (13).

## 3. RESULTS

**Table 1**

**Accuracy obtained for different combination of pre-processing and classifier methods**

| *Pre-Processing techniques* | *Classifier* | *Accuracy (%)* |
|---|---|---|
| *m*CSP | LDA | 62 |
| | SVM | 60 |
| Thin ICA-CSP | LDA | 64 |
| | SVM | 63 |

The comparison of performance was done for both *m*CSP and thin ICA-CSP algorithms using two classifiers *i.e*. LDA and SVM with Gaussian RBF kernel in the BCI competition IV dataset 2a. In the experiment, all the four classes were used for determining the performance. Both the feature extraction methods are tested using LDA and SVM and the observations are summarized in Table 1. For more analysis, the bar graph comparison of all the combined methods is shown in Fig 1. From the figure, we can observe that for this dataset the performance of thinICA-CSP with LDA gives the maximum classification performance of. In overall, it is also observed that the thin ICA-CSP performs better than the *m*CSP.

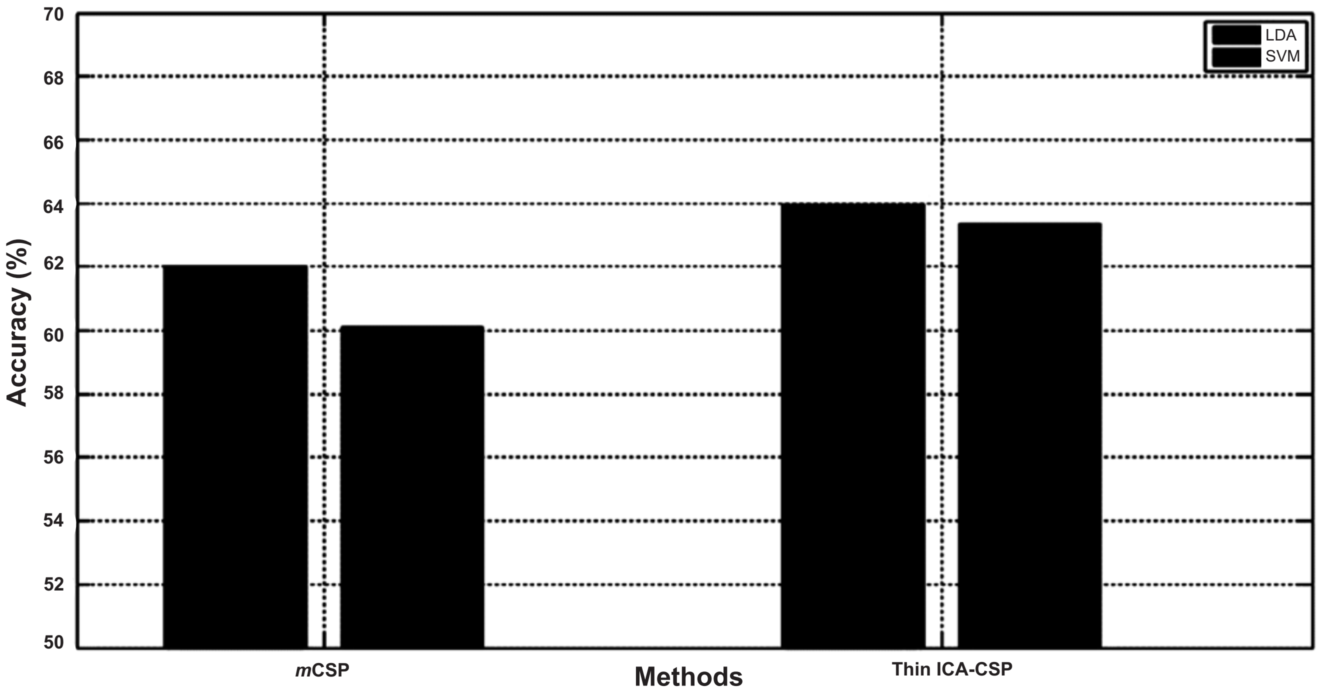


**Figure 1: Performance comparison of mCSP and thinICA-CSP using LDA and SVM classifier**

To support our results we have also performed the boxplot analysis. The boxplot analysis is mainly useful for observing the distribution of performance. The boxplot comparison of all the combinations is shown in fig 2. From the boxplot analysis, we can predict that the percentile will be less if there are more number of a subject with low performance for the corresponding methods. Similarly, the percentile will be high if many subjects are performing better for that method. From the figure, we can observe that the percentile is almost equal for mCSP-SVM, thinICA-CSP-LDA and thinICA-CSP-SVM but the percentile is higher for thinICA-CSP-LDA.

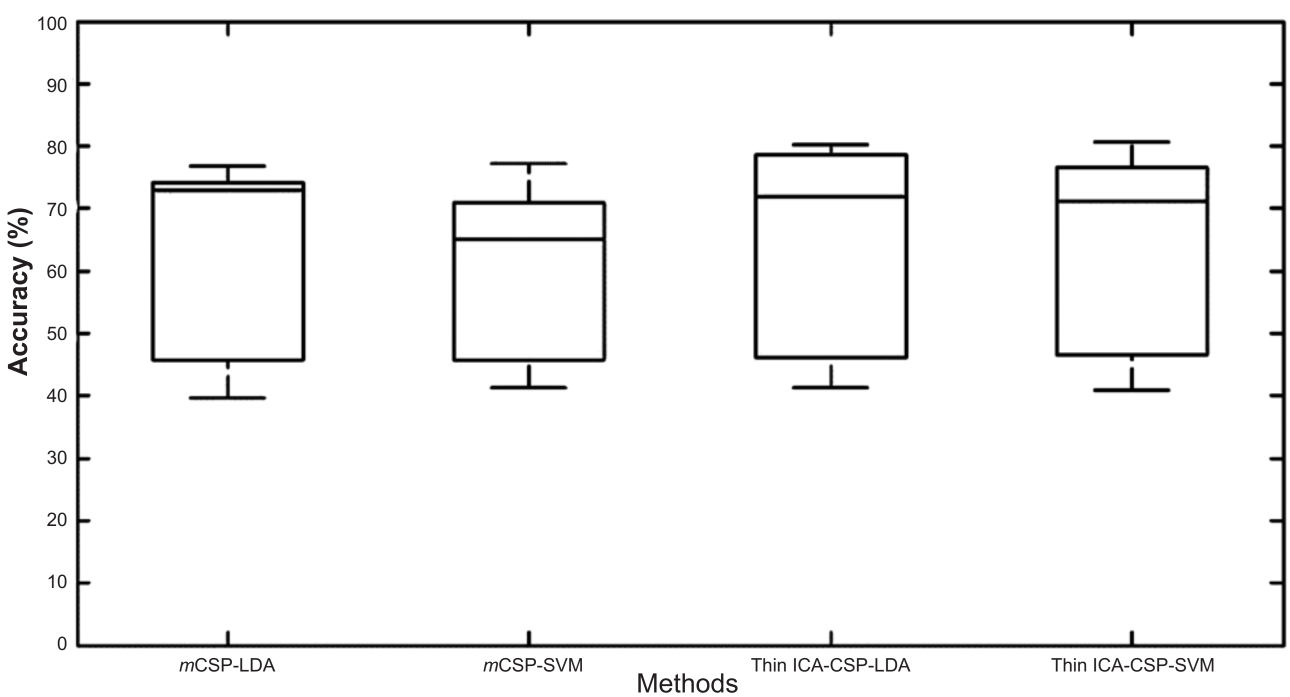


**Figure 2: Boxplot comparison of mCSP and thinICA-CSP using LDA and SVM classifier**

From the obtained results, we observe that among the two feature extraction algorithms-mCSP and thinICA-CSP, thinICA-CSP outperforms mCSP algorithm. We also observe from the boxplot analysis that the thinICA-CSP performs well for a subject which gives low performance with mCSP. In order to examine the performance of different classifier, we have tested both the feature extraction algorithms with LDA and SVM classifier. SVM with Gaussian RBF kernel was used for this experiment. From the comparative study, we observe that the LDA performs better than the SVM for this dataset.

## 4. DISCUSSIONS

The experiment in this paper was done using two feature extraction methods for obtaining the different class motor imagery features. As mentioned before, the presence of outliers reduces the classification performance of the system. Another factor that reduces the classification performance is using more number of classes for classification. The CSP method is considered as a very effective method for feature extraction of two classes, but it is not robust to the outliers. The thinICA-CSP based algorithm is based on joint diagonalization optimization which is robust to the presence of outliers. Moreover, the mixing matrix is initially initialized with the CSP solution for have a closed solution and selects the components that are related only to MI movements. From the results, we observed that the performance of thin ICA-CSP outperforms the *m*CSP. The results indicate that thinICA-CSP method is more robust than CSP for this dataset. In order to understand the effect of a classifier in obtaining the classification performance, we have obtained the performance using two classifiers: LDA and SVM. The performance is quantified by computing the classification accuracy. Both LVM and SVM are the popular methods for classification. LDA is famous for its easy computation but its decision boundary is not flexible and it is not robust to the outliers. Additionally, it is not accurate for the sample with more number of features than the observation. On the other hand, SVM can define a non-linear boundary which makes it useful for discrimination of data with unknown distribution by selecting proper kernels and parameters. For our studies, we have tested using LDA and SVM with Gaussian RBF. The obtained classification accuracy shows better results using LDA for this data. The reason may be because this set of data is linearly separable into multiple classes. From the overall observation, we can suggest that thinICA-CSP with LDA classifier could be the best choice for discrimination of multiclass MI movements in BCI applications.

## 5. CONCLUSIONS

In this paper, we have studied the performance analysis of two feature extraction methods-*m*CSP and thinICA-CSP using LDA and SVM. The experiment was done on BCI competition IV dataset 2a using four different MI EEG signals. For this dataset, we observed that the thin ICA-CSP with LDA classifier performs better than other combination of methods. We can conclude that thin ICA is more robust than *m*CSP and LDA are more effective than SVM for the classification of multiclass movement for the given dataset.

## 6. REFERENCES

1. Berger H. Über das elektrenkephalogramm des menschen II. J für Psychol und Neurol. 1930;40:160–79.
2. Vidal J. Towards direct brain-computer communication. Annu Rev Biophys ans Bioeng. 1973;2(1):157–80.
3. Dornhege G, Blankertz B, Curio G, Muller KR. Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. IEEE Trans Biomed Eng. 2004;51(6):993–1002.
4. Grosse-Wentrup M, Buss M. Multiclass common spatial patterns and information theoretic feature extraction. IEEE Trans Biomed Eng. 2008;55(8):1991–2000.
5. Chin ZY, Ang KK, Wang C, Guan C, Zhang H. Multi-class Filter Bank common spatial pattern for four-class motor imagery BCI. Proc 31st Annu Int Conf IEEE Eng Med Biol Soc Eng Futur Biomed EMBC 2009. 2009;138632:571–4.

6. Barachant A, Bonnet S, Congedo M, Jutten C. Multiclass brain-computer interface classification by Riemannian geometry. IEEE Trans Biomed Eng. 2012;59(4): 920–8.
7. Naeem M, Brunner C, Leeb R, Graimann B, Pfurtscheller G. Seperability of four - class motor imagery data using independent components analysis. J Neural Eng. 2006;3(3):208–16.
8. Thiyam DB, Cruces S, ElagiriRamalingam R. ThinICA-CSP algorithm for discrimination of multiclass motor imagery movements. IEEE TENCON conference (accepted). 2016.
9. Cruces S, Cichocki A, De Lathauwer L. Thin QR and SVD factorizations for simultaneous blind signal extraction. European Signal Processing Conference. 2004. p. 217–20.
10. Balakrishnama S, Ganapathiraju A. Linear Discriminant Aalysis -A Brief Tutorial. 1998 p. 1–8.
11. Cortes C, Vapnik V. Support-Vector Networks. Mach Learn. 1995;20(3273-297).
12. Cluster based Key Management Authentication in Wireless Bio Sensor Network “, , International Journal of pharma and bio sciences.