Multiclass MI-Task Classification Using Logistic Regression and Filter Bank Common Spatial Patterns

Md. Mahmudul Haque Joy, Mohammad Hasan, Abu Saleh Musa Miah, Abir Ahmed, Sadia Anwar Tohfa, Md. Farukul Islam Bhuaiyan, Ashrafun Zannat, and Md Mamunur Rashid

Department of CSE, Bangladesh Army University of Science and Technology (BAUST), Saidpur, Bangladesh

hasancse.cuet13@gmail.com, abusalehcse.ru@gmail.com

Abstract. We proposed a classification technique of EEG motor imagery signals using Logistic regression and feature extraction algorithm using filter bank common spatial pattern (FBCSP). Main theme of FBCSP is that the signals decomposed into 5 subband then calculated CSP for each subband, this algorithm also allows automated frequency band selection. We combined each subband CSP feature vector, feed this feature vector into machine learning algorithm. In the paper Logistic regression is used to classify among multiple classes. To evaluate this method, we used here publicly available dataset namely Brain-Computer Interface competition IV-2a. Because of high accuracy and kappa that shown in accuracy table that proposed method is promising.

Keywords: Motor Imagery (MI) · Brain Computer Interface (BCI) · Electroencephalogram (EEG) · EEG signal classification · Logistic Regression

1 Introduction

BCI uses EEG signal to establish a direct communication path away between an enhanced or wired brain and external device based on signal processing system [1].

For the people suffering from neural disorders, motor imagery brain computer interface provides a non-muscular solution [2].

EEG based invasive BCI system is immense temporal solution, relatively high portability, low-cost and non-invasive. The raw EEG signals are noisy or corrupted with various artifacts or noises like cardiac activity, muscular activity, eye blinking etc. [3].

So, some initial steps is necessary to minimize background noise and artifacts named preprocessing. Preprocessing of high dimensional data results in noise-free and artifact less data which can be converted into commands by classifying the preprocessed data using a machine learning algorithm.

The EEG signal from motor cortex is used in this paper. Which is correlate to the environment when a subject thinking a movement like left or right-hand movement. It is called sensory-motor activity. Movement imagined or done in real-life cases increases...
or decreases in $\mu$ activity in the device of motor-cortex-sensor. The band frequencies $\mu$ (8–15 Hz) and beta $\beta$ (15–25 Hz) rhythms are activated when the persons imagined motor movement and the amplitude decreases before the actual movement [4].

2 Related Works

The proposed system is dealing with the BCI competition where many numbers of the methodology used to solve the BCI competition problem since the start of the BCI competition. The success of brain signal classification in BCI competition significantly depends on feature extraction from the observed data, firstly used auto-regressive (AR) model with adaptive auto-regressive models (AAR) [5], fast Fourier analysis and windowed and cross-correlation. Then Principle Component Analysis (PCA), Common spatial patterns (CSP) are broadly used [6]. In BCI research common spatial pattern is a very popular method to obtain features from the EEG signal. LDA (Linear Discriminate Analysis), ANN (Artificial Neural Network), KNN (K-Nearest Neighbor) classification algorithm, SVM (Support Vector Machine) algorithm are used for classify into right intention [5, 7]. First developed the support vector machine (SVM) by Vapni based on statistical learning theory in 1995, which is used for nonlinear regression and classification. The main concept of SVMs is to make samples linearly separable by calculating optimal planes from high dimensional space. Based on the high dimension, local minima, non-linear relationships, and small sample size SVMs can solve practical problems.

For the random and non-stationary nature of electro encephalography (EEG) signals is the main difficulty of SVM to choose kernel function in practical application [8]. The concept of the Naive Bayes (NB) algorithm is to find the probability of a dependent event based on occurring given the probability of another event that has already occurred. The problem of Naïve-Baise algorithm are, produced estimated probability can be inaccurate, sometimes maximum probability assigns to the correct class. In our proposed methodology for feature extraction. The method we proposed here namely filter bank common spatial pattern is the modified version of common spatial pattern and for classification we used the Logistic Regression model.

3 Methodology

For feature extraction we used in the paper Filter Bank Common Spatial Pattern (FBCSP) and Logistic regression is used for identifying right intention of human thinking. Feature selection is performed in FBCSP where as we are classifying the data directly by Logistic regression [9].

3.1 Preprocessing the Data

Cross Validation
Cross validation tests the model in training phase for the dataset in order to restrict problems like overfitting and under fitting. It’s important that the validation and training set is drown from the same distribution.
Validation helps us to find out the model which will perform best on unseen data. It also evaluates the quality of the model. It mainly depends on the number of splits and fold in the datasets to produce random train validation split. With stratification when splitting data, we achieve similar target distribution over different folds. Cross validation is very useful to tackle overfitting and underfitting in addition it also determines which parameters will result in lowest test error.

**Noise Removal**

Obtained signal from signal acquisition steps are usually contains a lot of artifacts because of high frequency noise such as EOG’s and EMG’s because of electrical interference. Besides the distance between the scalp and the neurons makes it difficult to pinpoint the exact location of where an activation took place can also be an issue. As a result, preprocessing these signals is an import step for such experiments.

**Band Pass Filtering**

A band pass filter passes signals between two specific frequencies.

Most human brain activity produced within the frequency band of (2–40 Hz). For while, a band pass filter with range of frequency 2–40 Hz used here. Most of the high frequency noises can be filtered using band pass filter. The filter can have as much as sub bands as one wants. Sub band is used because alpha beta brain rhythms resides within such frequency [8].

**Butterworth Filtering**

It has a maximum flat frequency response with slow cut off and no gain ripple in pass and stop band. In the butterworth filter zero frequency is used as stop band and maximally in pass band.

**Standard Scaling**

Data have collected from multiple sources so lack of standardization can be create problem during data preprocessing. Large scale collection development is the process to test the neurological phenomenon across experiment and subject with robustness of approaches. If we use only raw data without standard preprocessing then it might create conflicts with respect to collection development. So, standardization format is very much important to progress in large scale in EEG. This kind comparison needs to start from datasets that are well documented and analysis ready. The important step for preprocessing of large scale is noise removal and inefficient channel detection. Several criteria use standard deviation of Z-score to replace the mean by the median and the standard deviation. It also detects nan-data from the channels. So, the key step of mining EEG across large collections is to develop a standardize preprocessing pipeline that will allow us to preform various analysis with reference to the raw data [9].

**Normalization**

Normalization mainly is removal of mean and division by standard deviation which can be performed across the band over time. There are two types of normalization:

i) **Temporal:** Subtraction of mean of each window and division by standard deviation.
ii) **Ensemble:** Pointwise subtraction of an ensemble means and division by ensemble standard deviation.

Temporal normalization is usually a good idea when possible and reasonable

### 3.2 Logistic Regression

We got a feature vector $f_i$ from FBCSP from both training and testing EEG dataset. Now we used Logistic Regression model used to feed the training feature then evaluate this model using testing feature. Classification algorithm Logistic Regression is a statistic technique borrowed by machine learning. It is used for estimating values from parameters coefficient. It predicts the outcome based on the given parameters. Logistic Regression can be divided in to two kinds. They are:

(i) Binary (example: cancer yes or no)
(ii) Multi-linear functions (example: Book, pencil, pen).

Based on the number of parameters, there are three kinds of logistic regression:

1. Binominal: Categorical output with two values ‘0’ or ‘1’.
2. Multinomial: Categorical output with more than two values: good, better, best
3. Ordinal: If multiple categories are in orders. Like 0, 1, 2, 3, 4 or A, B, C etc.

In this paper, we are dealing with ordinal or multi-linear functions. Our feature vector can be represented as matrix $M$.

\[
M = \begin{bmatrix}
    m_{11} & m_{12} & \cdots & m_{1k} \\
    m_{21} & \vdots & \vdots \\
    \vdots & \ddots & \vdots \\
    m_{n1} & \cdots & m
\end{bmatrix}
\]

Where ‘$k$’ is the feature variables. And $m_{i,j}$ represents values of features and observation. A single observation can be represented as below $M$.

\[
X = \begin{bmatrix}
    1 \\
    m_{i1} \\
    m_{i2} \\
    m_{i3} \\
    \vdots \\
    \vdots \\
    m_{in}
\end{bmatrix}
\]

The hypothesis of Logistic Regression $h(x_i)$ presents the predicted response for $i^{th}$ observation.
The hypothesis is described in Eq. (i):

\[ 0 \leq h_\theta(x) \leq 1 \]  

(i)

Where \( h(x_i) = g(z) = \frac{1}{1+e^{-z}} \)

Logistic function can be defined using Eq. (i). This function can take input any range’s number and produced a output value between 0 and 1, making the function useful in the prediction in probabilities (Fig. 1).

![Sigmoid function outcome.](image)

The hypothesis also can be written as:

\[ \sigma(z) = \sigma(\beta_0 + \beta_1 x) \]

Where, \( z = \beta_0 + \beta_1 x \)

The cost function of linear regression is,

If \( y = 1 \log(h_\theta(x)) \)

If \( y = 0 \log(1 - h_\theta(x)) \).

Cost function measures the machine learning performance for the given data. This calculates the difference between the expected value and predicted value called error value. Presents it in the form of a single real number which should not be negative. To minimize the cost value gradient descent is used. A gradient descent function should be run on every parameter to minimize the cost. The equation is:

\[ \theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} j(\theta) \]

Hence \( \theta_j \) is the gradient descent and \( j(\theta) \) is the cost function.

### 3.3 Filter Bank Common Spatial Pattern

In a task like motor imagery, FBCSP is one of the most popular features obtaining method based on a common spatial pattern. It consists of four steps which use the Common Spatial Pattern algorithm for feature selection and classification on selected feature [10].
The first multiple bandpass filter namely the Chebyshev type II filter is used to remove the artifact from EEG.

To compute the most effective feature of the EEG signal, we used spatial filtering with CSP on each frequency subband. For finding ERD and ERS from EEG, CSP is one of the powerful algorithms which is used [6, 11].

The method named common spatial pattern is used to designs spatial filters that have an ideally fixed variance between the filtered time-series data.

We have given input data denoting EEG/ECoG data from trial \( i \) for class \( c \) \{1, 2, 3, 4\} (e.g., left hand, right hand, tongue versus foot motor imagery). Each of the sets is an \( X \times Y \) matrix, where \( X \) is the number of channel used during recording time, and \( Y \) denotes the number of sample points in time per channel [12].

Here used a Butterworth filter to scale those data. Low-pass Butterworth filter design can be defined as:

\[
\frac{v_0(s)}{v_i(s)} = \frac{R}{s^3(L_1C_2L_3) + s^2(L_1C_2R) + s(L_1 + L_2) + R}
\]

After scaling the data, we will perform a Common spatial pattern on it.

Our spatial filter we obtained from the CSP method can be used to linearly transform the EEG measurement in Eq. (ii) [13, 18].

\[
Z = WT E
\]

\( W \in \mathbb{R}^{ch\times ch} \) represents CSP projected matrix, \( E \in \mathbb{R}^{ch\times s} \) is the trial of EEG signal measurement, and \( Z \in \mathbb{R}^{ch\times s} \) which are the filtered signals where \( ch \) represents the number of electrode and \( s \) represents the number of sample point per channel [14]. Projected matrix \( W \) also called a Transformation matrix that sums up the features whose variance is used to determine the EEG between two classes [15, 19]. Solving the equivalent decomposition problem can be useful to calculate \( W \) using Eq. (iii).

\[
\Sigma_1 W = (\Sigma_1 + \Sigma_2) WD
\]

\( \Sigma_1 \) and \( \Sigma_2 \) are calculated from the respective motor imagery task by estimating the band pass filter EE measurement from the covariance matrix. The eigenvalue of \( \Sigma_1 \) is contained in a diagonal matrix \( D \). For \( i \)th trial the CSP features for the EEG measurements are calculated by Eq. (iv) by:

\[
f_i = \log \left( \frac{\text{diag}(\tilde{W}^T E_i \tilde{E}_i^T \tilde{W})}{\text{trace}(\tilde{W}^T E_i \tilde{E}_i^T \tilde{W})} \right)
\]

Here \( f_i \) is the features which we get from the common spatial pattern. We are calculating the \( \text{diag}() \) using transposed projection matrix and with the Eigen value. The trace is similarly calculated with the same attribute. The divisor result is then passed as parameter into a logarithmic function to calculate the features.

### 3.4 Random Forest

To classify our CSP feature vector \( f_i \) we also used the Random Forest machine learning algorithm as a classifier [16, 17]. The concept of this algorithm is to calculate Gini index
and information gain from the feature vector \( f_i \) to build up the decision tree. Multiple trees are created to classify EEG label in accordance with their attributes, each tree individually gives a classification result and saves them as for appropriate class, which has the most overall the trees is chosen for classification. For regression, the average output by different trees is taken. Pseudocode of random forest is given below:

- A number of training features set is N. then, randomly take samples for these N cases with replacement.
- For M input features or variables, m variables are selected as \( m < m \). m is constant although growing the forest. We used the best split of \( m \) to split the given node.
- Each tree is expanded as large as possible without pruning.
- By summing up the prediction odd n trees new data is predicted.
- Maximum numbers of votes are accountable for classification.
- Average votes are used for classification.

Random Forest with two decision trees is given below:

![Random Forest Example](image)

In Fig. 2, \( V[i] \) is the \( i \)-th value of the feature vector. The random forest depicted in figure predicts one of 3 class labels: A, B, or C. Many kinds of decision trees are used for classification in a random forest. It uses bagging and feature randomness to build individual trees to build an uncorrelated forest of the tree, to make predictions more accurate than individual trees. It is an ensemble machine learning algorithm, uses a divide and concurs approach.

### 4 Experimental Analysis and Discussion

#### 4.1 Dataset Description

For dataset 2a of Brain Computer Interface (BCI) competition IV that is contained Electroencephalogram (EEG) data from 9 subjects. It has 4 types of motor imagery tasks, these are the thinking imagination of right-hand movement, left-hand movement, both feet movement, and tongue.
Data has recorded from individual subject over two sessions. 6 runs combined into one session and separated each run by a short gap. In front of a computer screen, we make the subjects to sit comfortably. When starting a trail \((t = 0)\), on the black screen taxation cross was appeared.

Besides, these as a short aural warning accent was presented. Subsequently 2 s \((t = 2)\) a pointer in the form of an indicator is goes may be to the up, down, left, or right (Fig. 3).

One session has 288 trails in total, 48 trails per run, 12 for each of the four classes. To record the EEG signal 22 Ag/AgCl electrode distance of 3.5 mm was used; the image is shown in Fig. 4 most signals were recorded as mono-polar with left and right side serving as reference and ground respectively. 250 Hz sampling rate was used on the signal during signal transformation and passes those signal whose have frequency between 0.5 Hz to 100 Hz using band pass filter. The amplifier had a sensitivity of 100 \(\mu\)V. To discard line noise, we used an additional notch filter of 50 Hz.

4.2 Results

To obtain 9 band-pass filters here band pass filter is applied to cover 4–40 Hz. Later we have used CSP for extracting features for each band.

The best results appeared for subject A01T, A02T, A07T, A08T, A09T is giving the best possible results using our proposed method and the overall result is better than the results we found for Naïve Base Classifier and Random Forest classifier for the same dataset. While we are getting a comparatively less performing model for other subjects like A03T (Table 1).
The best results appeared for subjects 1, 3, 7, 8, 9 using our proposed method and the overall result is better than the results we found for Random Forest classifier for the same dataset. As we can see our model is giving better results than Random Forest Classifier, there are some factors working behind it: Sometimes Logistic regression gives better results than Random Forest Classifier because of when the dataset has a higher impurity or higher Gini index. In this case, this factor is playing a major role.

Random forest is better when it predicts actual result with a lower accuracy but if the class label of a dataset is labelled correctly then logistic regression can play the role
as well. In this case our model is giving accuracy because of the accurate labeling in the preprocessing step (Fig. 5).

The result can be seen from the bar chart where the black bar represents our proposed method which is significantly rising higher than the others.

5 Conclusion and Future Work

For four class EEG motor imagery classification we have used logistic regression and FBCSP is used for feature extraction. Our experiment gives a remarkable result but the accuracy doesn’t remain constant. Due to subject variability we were unable to find a method that can give equally good results for every subject. FBCSP is giving a better classification rate. As future work, we are trying to create a method which is equally effective for every subject and every dataset.

References

1. Shih, J.J., Krusienski, D.J., Wolpaw, J.R.: Brain-computer interfaces in medicine. Mayo Clin. Proc. 87, 268–279 (2012). https://doi.org/10.1016/j.mayocp.2011.12.008
2. Lazarou, I., Nikolopoulos, S., Petrantonis, P.C., Kompatsiaris, I., Tsolaki, M.: EEG-based brain–computer interfaces for communication and rehabilitation of people with motor impairment: a novel approach of the 21st century. Front. Hum. Neurosci. 12, 14 (2018). https://doi.org/10.3389/fnhum.2018.00014
3. Nicolas-Alonso, L.F., Gomez-Gil, J.: Brain computer interfaces, a review. Sensors 12, 1211–1279 (2012). https://doi.org/10.3390/s120201211
4. Padfield, N., Zabalza, J., Zhao, H., Masero, V., Ren, J.: EEG-based brain-computer interfaces using motor-imagery: techniques and challenges. Sensors 19(6), 1423 (2019). https://doi.org/10.3390/s19061423
5. Wang, T., Deng, J., He, B.: Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns. Clin. Neurophysiol. 115, 2744–2753 (2004)
6. Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Müller, K.R.: Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Process. Mag. 25, 41–56 (2008)
7. Allison, B.Z., Pineda, J.A.: ERPs evoked by different matrix sizes: implications for a brain computer interface (BCI) system. IEEE Trans. Neural Syst. Rehabil. Eng. 11, 110–113 (2003)
8. Bashashati, A., Fatourechi, M., Ward, R.K., Birch, G.E.: A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. J. Neural Eng. 4, R32–R57 (2007)
9. Ang, K.K., Chin, Z.Y., Wang, C., Guan, C., Zhang, H.: Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. Front. Neurosci. (2012). https://doi.org/10.3389/fnins.2012.00039
10. Ha, K. W. & Jeong, J. W.: Motor imagery EEG classification using capsule networks. Sensors (Switzerland) 19, (2019). https://doi.org/10.3390/s19132854
11. Djemal, R., Bazey, A.G., Belwafi, K., Gannouni, S., Kaaniche, W.: Three-class EEG-based motor imagery classification using phase-space reconstruction technique. Brain Sci. 6, 36 (2016). https://doi.org/10.3390/brainsci6030036
12. Wang, D., Miao, D., Blohm, G.: Multi-class motor imagery EEG decoding for brain-computer interfaces. Front. Neurosci. 6, 151 (2012). https://doi.org/10.3389/fnins.2012.00151
13. Pfurtscheller, G., da Lopes Silva, F.: Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin. Neurophysiol. 110, 1842–1857 (1999)
14. Cheng, D., Liu, Y., Zhang, L.: Exploring motor imagery EEG patterns for stroke patients with deep neural networks. In: IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings, ICASSP, April 2018, pp. 2561–2565. Institute of Electrical and Electronics Engineers Inc. (2018)

15. Oikonomou, V.P., Georgiadis, K., Liaros, G., Nikolopoulos, S., Kompatsiaris, I.: A comparison study on EEG signal processing techniques using motor imagery EEG data. In: Proceedings - IEEE Symposium on Computer-Based Medical Systems, June 2017, pp. 781–786. Institute of Electrical and Electronics Engineers Inc. (2017)

16. Bashashati, H., Ward, R.K., Birch, G.E., Bashashati, A.: Comparing different classifiers in sensory motor brain computer interfaces. PLoS ONE 10, e0129435 (2015). https://doi.org/10.1371/journal.pone.0129435

17. Garrett, D., Peterson, D.A., Anderson, C.W., Thaut, M.H.: Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Trans. Neural Syst. Rehabil.Eng. 11, 141–144 (2003)

18. Bentlemsan, M., Zemouri, E.T., Bouchaffra, D., Yahya-Zoubir, B., Ferroudji, K.: Random forest and filter bank common spatial patterns for EEG-based motor imagery classification. In: Proceedings - International Conference on Intelligent Systems, Modelling and Simulation, ISMS (2014). https://doi.org/10.1109/ISMS.2014.46

19. Padfield, N., Zabalza, J., Zhao, H., Masero, V., Ren, J.: EEG-based brain-computer interfaces using motor-imagery: techniques and challenges. Sensors (Switzerland) (2019). https://doi.org/10.3390/s19061423