



## Feature selection using regularized neighbourhood component analysis to enhance the classification performance of motor imagery signals



Nitesh Singh Malan\*, Shiru Sharma

School of Biomedical Engineering, Indian Institute of Technology, Banaras Hindu University, Varanasi, India

### ARTICLE INFO

#### Keywords:

Brain-computer interface  
Genetic algorithm  
Motor imagery  
Neighbourhood component analysis  
Principal component analysis  
Support vector machine

### ABSTRACT

In motor imagery (MI) based brain–computer interface (BCI) signal analysis, mu and beta rhythms of electroencephalograms (EEGs) are widely investigated due to their high temporal resolution and capability to define the different movement-related mental tasks separately. However, due to the high dimensions and subject-specific behaviour of EEG features, there is a need for a suitable feature selection algorithm that can select the optimal features to give the best classification performance along with increased computational efficiency. The present study proposes a feature selection algorithm based on neighbourhood component analysis (NCA) with modification of the regularization parameter. In the experiment, time, frequency, and phase features of the EEG are extracted using a dual-tree complex wavelet transform (DTCWT). Afterwards, the proposed algorithm selects the most significant EEG features, and using these selected features, a support vector machine (SVM) classifier performs the classification of MI signals. The proposed algorithm has been validated experimentally on two public BCI datasets (BCI Competition II Dataset III and BCI Competition IV Dataset 2b). The classification performance of the algorithm is quantified by the average accuracy and kappa coefficient, whose values are 80.7% and 0.615 respectively. The performance of the proposed algorithm is compared with standard feature selection methods based on Genetic Algorithm (GA), Principal Component Analysis (PCA), and ReliefF and performs better than these methods. Further, the proposed algorithm selects the lowest number of features and results in increased computational efficiency, which makes it a promising feature selection tool for an MI-based BCI system.

### 1. Introduction

Brain–computer interfaces (BCIs) find applications in assistive technology to assist people with neuromuscular disorders. The BCI system establishes a direct communication pathway between the electrophysiological signals originating from the brain and the external devices, such as a robotic arm, prosthetic device, or wheelchair [1]. Its performance is independent of nerves and muscles which are damaged or not fully functional to perform movements. For the design and development of the BCI system, information from electroencephalogram (EEG) signals of a subject is used as the main feature and is acquired while performing a specific task for which the BCI system is intended to be developed. These signals originate from the different brain locations depending on the mental task performed by the subject, and these signals are then used to generate command signals to control the external devices. In motor imagery (MI) we pick up the biopotential signals originating from the sensorimotor cortex area of the brain while a person imagines the motor movements (without actually performing

the real movements) [2]. Event-related desynchronization (ERD) and event-related synchronization (ERS) patterns are observed in the mu rhythms (8–13 Hz) and beta rhythms (13–25 Hz) of the brain activity during the motor imagery task [3].

Motor imagery-based brain–computer interface (MI-BCI) systems have been finding wide application in medical fields, for instance, in neuro-rehabilitation of stroke patients, real time control of a robotic arm, and wheelchair control [4,5]. In the work of Cantillo-Negrete et al. [6], a scheme has been proposed to couple MI-BCI with a robotic orthosis device to rehabilitate human upper extremities post-stroke. In general, the raw EEG data of the patients are analysed using computational models to classify various mental tasks. The EEG pattern classification can be performed using pattern recognition methods such as feature extraction, feature selection, and classification or regression [7,8]. To date many algorithms have been proposed to classify the mental tasks for different BCI applications [9–15]. However, obtaining a higher classification performance is still a challenging task, and further, the high dimensions of the features extracted from EEG data make

\* Corresponding author.

E-mail addresses: [niteshsm.rs.bme16@itbhu.ac.in](mailto:niteshsm.rs.bme16@itbhu.ac.in) (N.S. Malan), [shiru.bme@itbhu.ac.in](mailto:shiru.bme@itbhu.ac.in) (S. Sharma).

the analysis task a complex and computationally expensive exercise [16]. To address these issues, feature selection techniques have been implemented to reject the redundant features. Some of the feature selection approaches applied in the recent studies for MI classification include Principal Component Analysis (PCA) [17], Independent Component Analysis (ICA) [18], and evolutionary algorithms based on the Firefly Algorithm (FA) [19], Differential Evolution (DE) [20], and Artificial Bee Colony (ABC) optimization [21]. Broadly, the feature selection techniques are categorized as filter methods and wrapper methods [22]. In the filter approach, features are selected based on the statistical properties of the input data, whereas wrapper methods use a classifier to select the best subset of features. In the work of Ramos et al. [23], a comparison of one wrapper method [Genetic Algorithm (GA)] and five filter methods, (i) Correlation-based Feature Selection (CFS), (ii) ReliefF, (iii) Consistency, (iv) mRmR, and (v) C4.5, was done. In this comparison, feature selection techniques were combined with different classifiers and the performance was compared using the classification accuracy and kappa value. Ramos et al. concluded that GA with a linear discriminant analysis (LDA) classifier outperforms all the other combinations. Thus, wrapper procedures provide better performance than the filter methods but are computationally expensive as multiple training sessions of the classification algorithm are involved. Moreover, the performance of the wrapper methods is highly dependent on the initial configurable parameters, which need to be manually defined by the user. Goldberger et al. [24] proposed Neighbourhood Component Analysis (NCA), a feature weighting approach which optimizes the nearest neighbour classifier performance to address the issue of high dimensionality of the training data. However, the cost function of NCA is prone to overfitting, and an improved version of NCA with a regularization term was proposed by Zhirong Yang et al. [25], where the regularization term was selected empirically. To alleviate the problem of optimal selection of the regularization term, we introduced a learning-based method for the tuning of the regularization parameter to be added to the cost function of the conventional NCA. This algorithm mainly targets the problem of optimizing the subject-specific EEG features to further improve the MI classification performance. The improved method, namely regularized NCA (RNCA) has been validated as a feature selection technique on two public MI datasets. The effectiveness of the proposed feature selection RNCA algorithm is compared with the following standard feature selection algorithms: (i) PCA, (ii) GA, and (iii) ReliefF. The RNCA is a relatively new approach that, to the best of our knowledge, has not been applied to an MI dataset. In our approach, differentiation of EEG corresponding to mental tasks is performed using steps that include statistical, phase, and frequency features extraction followed by selection of the best subset of features using the RNCA and finally classification utilizing a support vector machine (SVM) classifier.

The subsequent sections of this paper are presented as follows. Section II describes the methods and materials used in this work such as dataset descriptions, feature extraction techniques, and competing feature selection algorithms including the proposed RNCA feature selection method. The results and discussion are presented in Sections III and IV, respectively, and finally, the conclusions drawn are given in Section V.

## 2. Methods and materials

### 2.1. EEG dataset and paradigm

In this work, two EEG datasets provided by BCI Competitions II and IV are used. Brief descriptions of the datasets used are as follows.

**Dataset 1:** The first dataset used in this work was provided by BCI Competition II (Dataset III) [26]. The Department of Medical Informatics, Institute for Biomedical Engineering, Graz University of Technology (Gert Pfurtscheller, Alois Schlögl) performed experiments with the aim of providing continuous control over a BCI-feedback system.

This dataset was recorded from a 25-year-old female while she performed a motor imagery task of two classes (left-hand or right-hand movement) with a feedback session. The dataset contains a three-channel EEG from channel locations C3, CZ, and C4 sampled at 128 Hz and bandpass filtered between 0.5 and 30 Hz. The subject was sitting comfortably in an armchair and was instructed to control a feedback bar by imagining right-hand or left-hand movement according to the displayed cue. A total of 140 labelled training trials and 140 unlabelled test trials were recorded. Every trial consists of a 9-s EEG recording with the initial 2 s as a rest period; at the time  $t = 2$  s an acoustic sound indicating the start of the trial occurred for 1 s. At  $t = 3$  s, the subject started to see an arrow as a cue pointing either right or left and the subject performed the motor imagery task by moving the feedback bar in the direction of the cue by imagining the right-hand or left-hand movement.

**Dataset 2:** BCI Competition IV (Dataset 2b) [27] provided the second dataset used in this work. Dataset 2b was recorded by the Institute for Knowledge Discovery (Laboratory of Brain–Computer Interfaces), Graz University of Technology (Robert Leeb, Clemens Brunner, Gernot-Müller-Putz, Alois Schlögl, Gert Pfurtscheller). This dataset consists of three bipolar EEG channel signals acquired from nine subjects for two classes (left- and right-hand motor imagery). Bandpass filtering is applied between 0.5 and 100 Hz with a notch filter at 50 Hz to remove the power line noise.

### 2.2. Preprocessing

Due to the randomness and non-stationary nature of the EEG time series, there is a need to preprocess the data, which helps in noise reduction and identification of the main components present in the EEG signal. Further temporal characteristics of the EEG signal suggest that it varies significantly between different subjects and also between different training sessions with the same subject. To avoid such issues, time-frequency analysis of the EEG signal is performed to extract features in the time and frequency domains at the same time. To accomplish this, we selected a dual-tree complex wavelet transform (DTCWT) to decompose the EEG signal. The DTCWT is an enhancement of a discrete wavelet transform (DWT). DWTs are widely used to deal with non-stationary signals like EEGs, and problems with DWT such as aliasing and power losses at the transaction bands can be resolved using DTCWT [28].

In the EEG signal, ocular artefacts are more prominent up to 10 Hz. To remove ocular artefacts, an adaptive threshold presented by Zikov et al. [29] is applied to the coefficients of the sub-bands with a decomposition level below 10 Hz. In DTCWT, two DWTs work in parallel to compute the real and imaginary parts of the transform. Fig. 1 shows the analysis and synthesis filter banks of DTCWT used to analyse the EEG signal in the wavelet domain. In the present work, the level of decomposition selected was 4, and the signal was decomposed into four details, D1–D4, and one approximation A5.

### 2.3. Feature extraction

The decomposition of the EEG signal into frequency sub-bands after preprocessing results in increased dimensionality. The feature extraction is applied to reduce the high dimensionality and improve the differentiable capacity of the dataset between the different motor imagery classes. In other words, feature extraction provides the primary information carried by the raw EEG signal that can easily distinguish between different mental states. The feature space that contains the highlighted information is generated.

In this method, instead of using a wider frequency band, we have used three frequency sub-bands which are most associated with the mu and beta rhythms to analysis the ERS/ERD patterns. From the preprocessing step, wavelet coefficients of the three decompositions, D2–D4, are selected, considering the frequency range of interest

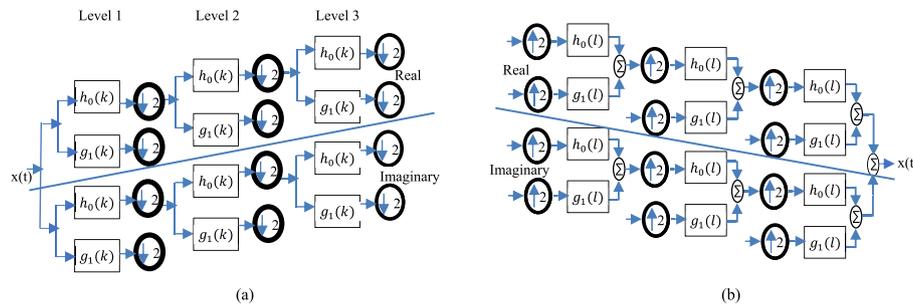


Fig. 1. DTCWT (a) Analysis filter bank of DTCWT and (b) synthesis filter bank of DTCWT.

4–30 Hz. The frequency ranges of D2 and D3 are 4–8 Hz and 8–15 Hz, respectively. Thus, they give only the mu rhythms of the brain activity and filter the other frequency components. However, for beta activity, we have used the sub-band D4 ranging between 15 and 30 Hz.

Parameters used to create feature space are a combination of various statistical, frequency, and phase information features of the EEG signal for each frequency sub-band. The parameters used in this work are as follows:

2.3.1. Statistical features

Discriminatory information from the wavelet coefficients of the raw EEG signal can be defined and calculated using statistical analysis [30,31]. In our approach, five statistical features are evaluated, namely the mean absolute values (MAV), standard deviation, variance, sample entropy, and root mean square (RMS) values of the coefficients of the details D2–D4 for channel locations C3 and C4.

2.3.2. Frequency features

The power spectral density (PSD) of the EEG signal describes the power carried by the signal as a function of frequency [32]. The EEG signal of two channels, C3 and C4, is filtered using DTCWT in three separate frequency bands ranging between 4 and 30 Hz. The PSD of each sub-band is evaluated.

2.3.3. Phase features

Phase locking value is a method of measuring the instantaneous phase relationship between two signals [33]. We use the EEG signal from the channel location CZ as a reference signal, and the phase relationship of C3 and C4 is calculated with respect to the reference CZ. The PLV is calculated for all three sub-bands.

Considering three sub-bands and the signal from two channel locations with seven features extracted, the outcome dimension of the feature space is  $3 \times 2 \times 7 = 42$  features. Hence, the complete training set is a  $42 \times N$  matrix, where  $N$  represents the total number of trials. Table 1 gives a brief description of all the features.

2.4. Feature selection

Feature selection is performed to convert the m-dimensional feature vector to a lower p-dimensional feature vector by rejecting the

redundant features. Also, the feature selection procedure plays a vital role in reducing the amount of data used for training the classifier. As a result, the execution speed of the classifier increases. The feature selection is said to be adequately performed if it enhances the generalization performance. In this section, four feature selection methods and the proposed algorithm are explained in detail.

2.4.1. Genetic Algorithm

Charles Darwin’s theory of natural evolution inspired the Genetic Algorithms (GAs) [34]. GAs are computational models which replicate the natural process of selecting the fittest individuals to reproduce the best offspring of the next generation. As a feature selection technique, it works to preserve the critical information carried by the features. GAs solve the feature selection problem by considering a chromosome-like data structure where a population of chromosomes is chosen and each chromosome is encoded as an array of binary bits. The length of the array is taken to be equivalent to the size of the feature set in the problem. Thus, each bit in the array represents one particular feature. Features with the bit value ‘high’ are selected for classification while those with the bit value ‘low’ are rejected. Hence, each chromosome represents one feature subset. Then, the fitness value of every chromosome from the population is measured as the kappa coefficient and classification accuracy using a learning algorithm. In our approach, the population size was equal to the length of the features set, and the tournament selection method was used with the elitist size set to two [35]. Also, the arithmetic crossover function was applied to every generation for the creation of next-generation offspring. The generation size was varied during the experiment.

2.4.2. Principal Component Analysis

Principal component analysis (PCA) is an unsupervised feature selection technique that can linearly transform a higher dimensional feature space into a lower dimensional feature space using the statistical approach [36]. Some of the variables in the original feature space are correlated with one another, and there exists some redundancy. These correlated variables are linearly combined into a smaller number of principal components that preserve the maximum amount of variance in the variables. Principal components are orthogonal to one another to avoid redundancy.

For feature selection, we evaluated the contributions of individual

Table 1 Description of different features extracted using statistical, frequency and phase analysis.

Feature Index	Features description
1–6	MAV of the Details D2-D4 for two EEG channels ( $3 \times 2 = 6$ )
7–12	The standard deviation of the Details D2-D4 for two EEG channels ( $3 \times 2 = 6$ )
13–18	The variance of the Details D2-D4 for two EEG channels ( $3 \times 2 = 6$ )
19–24	Sample entropy of the Details D2-D4 for two EEG channels ( $3 \times 2 = 6$ )
25–30	Root mean square (RMS) of the Details D2-D4 for two EEG channels ( $3 \times 2 = 6$ )
31–36	PSD of the Details D2-D4 for two EEG channels ( $3 \times 2 = 6$ )
37–42	Phase locking value of the Details D2-D4 for two EEG channels( $3 \times 2 = 6$ )

features to the principal components. We let the size of the training set  $S$  be  $n - by - p$ , where rows of  $S$  correspond to trial observations and columns correspond to features. In PCA,  $S$  is transformed into a coefficient matrix of size  $p - by - p$ , in which each column represents the coefficients of one principal component and columns are arranged in descending order of features variance [37]. To evaluate each feature's contribution, the original data set  $S$  is multiplied by the PCA coefficient matrix, which will project the original data onto the principal component vector space. In our method, we calculated the mean and variance of the columns of the projected data matrix to read the contribution of each feature to the principal components. Thus, features with a larger absolute mean compared with the variance are selected for classification.

### 2.4.3. ReliefF

As a filter approach, ReliefF [38] is an effective algorithm to solve multiclass data problems. The algorithm randomly selects a sample  $x$  from the training set and searches for  $k$  nearest neighbour samples of the same class and  $k$  nearest neighbour samples of the non-similar classes. Using the Euclidean distance, the closest nearest neighbour samples from each class are selected. A relevant weight is assigned to each feature by comparing the interclass distance and intraclass distance from the neighbour samples. This procedure is repeatedly performed on each feature sample, and each feature is assigned a weight. The features selected as the best subset have weights larger than a predefined threshold.

### 2.5. Neighbourhood component analysis

Neighbourhood component analysis (NCA) is a learning algorithm to measure the Mahalanobis distance used in the KNN classification algorithm [24]. NCA as a feature selection technique is a feature weighting scheme to select the best subset of features by maximizing an objective function that evaluates the average leave-one-out classification accuracy over the training data [39]. The algorithm works to assess a weighting vector  $w$  that corresponds to the feature vector  $x_i$  by optimizing the nearest neighbour learning classifier. In the NCA framework, like 1-nearest neighbour classifier, a reference sample point  $x_j$  is selected for the sample  $x_i$  from all the samples. The probability  $P_{ij}$  of  $x_j$  being chosen as a reference point for  $x_i$  from all the samples is higher depending on the closeness of the distance between the two samples. This distance can be measured by a weighted distance  $D_w$  defined as

$$D_w(x_i, x_j) = \sum_{m=1}^r w_m^2 |x_{im} - x_{jm}| \tag{1}$$

where  $w_m$  is the assigned weight of the  $m$ th feature. The relation between the probability  $P_{ij}$  and the weighted distance  $D_w$  can be established by introducing a kernel function  $k$  which returns large values for small  $D_w$ . The  $P_{ij}$  can be defined as

$$P_{ij} = \frac{k(D_w(x_i, x_j))}{\sum_{j=1, j \neq i}^n k(D_w(x_i, x_j))} \tag{2}$$

Also, if  $i = j$ , then  $P_{ii} = 0$ . The kernel function  $k$  is defined as  $k(z) = \exp\left(-\frac{z}{\sigma}\right)$  and the parameter  $\sigma$  is the kernel width, which affects the probability that a sample  $x_j$  will be selected as a reference point. Now, the probability of  $x_i$  being correctly classified can be written as

$$P_i = \sum_{j=1, j \neq i}^n P_{ij} Y_{ij} \tag{3}$$

where  $Y_{ij}$  indicates one only if  $y_i = y_j$ . Hence, the average leave-one-out classification accuracy is the summation of  $P_i$  over all the trials divided by the total number of trials, which can be seen as an objective function which needs to be maximized. However, this objective function is prone to overfitting. A term-regularization parameter  $\lambda$  is introduced in the

final objective function to avoid overfitting of the NCA model [39]. Thus, the objective function can be expressed as

$$A = \sum_{i=1}^n P_i - \lambda \sum_{m=1}^r w_m^2 \tag{4}$$

The objective defined in Eq. (4) is known as the regularized NCA (RNCA). The target of RNCA is to maximize the objective function  $A$ . To perform this,  $A$  can be solved using the conjugate gradient approach. If  $A$  is limited to be a diagonal matrix, then its diagonal values provide the weight of each feature. Based on the weights outcome, the best subset of features is selected.

### 2.6. Proposed method: RNCA as feature selection

This work assesses the performance of feature selection on the BCI dataset using RNCA. The idea of this method is to select the subset of features which gives the maximum classification accuracy or minimum generalization error. Further, this study presents a technique to regularize the NCA model of Eq. (4) for the feature selection of the MI data. Our approach contains the following steps:

Step 1: To apply feature selection on the BCI dataset, we begin by considering the training set  $S = \{(x_i, y_i), i = 1, 2, \dots, N\}$ , where  $x_i$  are the feature vectors,  $N$  is the number of trials,  $y_i \in \{1, 2, \dots, C\}$  defines its class label, and  $C$  is the number of classes. In this work,  $x_i$  comprise 42 features explained in Section 3 for two motor imagery classes.

Step 2: We perform the fivefold cross-validation on the training set  $S$  and evaluate the generalization error,  $err$ , defined as

$$err = \frac{1}{n} \sum_{i=1}^N I(k_i \neq t_i) \tag{5}$$

where  $k_i$  represents the predicted class,  $t_i$  is the true class, and  $I(x)$  returns 1 when  $k_i$  is not equal to  $t_i$  and otherwise returns 0.

Step 2 is executed to check whether feature selection is required. Now we fit the NCA model defined by Eq. (4) while keeping regularization parameter  $\lambda$  equal to zero and again calculate the generalization error. If the value of the generalization error after fitting the NCA model is less than that obtained before fitting the model, then there is a need for feature selection.

Step 3: Now, we tune the regularization parameter  $\lambda$  to obtain the minimum classification loss. To do this, we generate a uniformly distributed array  $\lambda_{val}$  of length  $L$ . Then, we fit the NCA model for each  $\lambda$  and store the estimated generalization error in an array.

Step 4: Step 2 is executed repeatedly for all folds and all values of  $\lambda$ . Simultaneously, the average classification loss is calculated from all the folds for each  $\lambda$  value. Subsequently, the value of  $\lambda$  corresponding to the minimum average classification loss is selected as the best lambda,  $\lambda_{Best}$ .

Step 5: Using the  $\lambda_{Best}$  value, the NCA model is run on the complete data and the feature weights of each feature are evaluated. Features with weights higher than 5% of the maximum feature weight are selected to classify the data.

Step 6: Next, the SVM classifier is trained using the updated training set selected in the previous step. We calculate the evaluation parameters such as confusion metrics, kappa value, and classification accuracy.

Note: The classification accuracy (CA) is equal to  $1 - err$ . Therefore, this algorithm works well if we search for a subset with the maximum CA instead of the minimum classification error.

The complete scheme of the proposed method is summarized in Fig. 2.

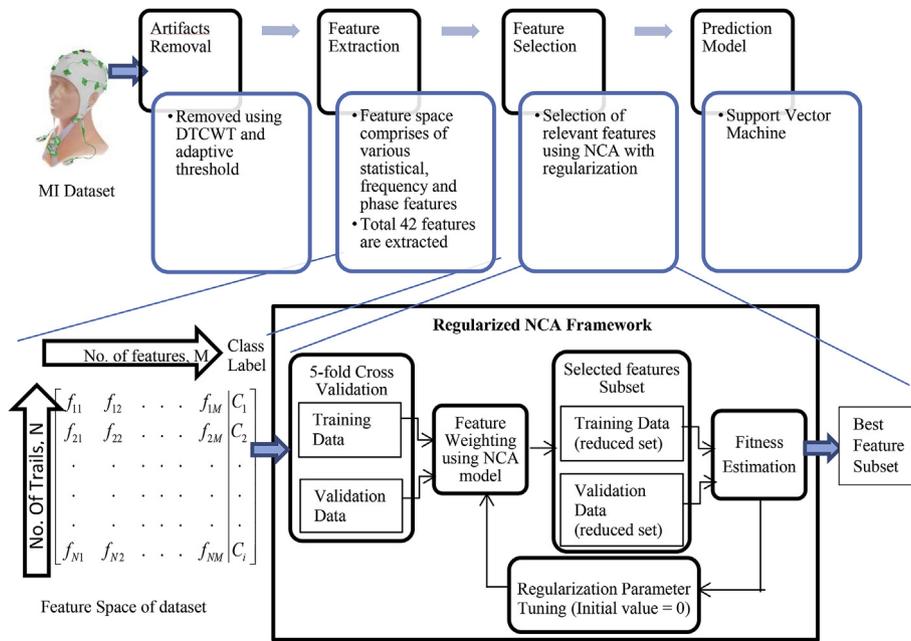


Fig. 2. Workflow elucidates the proposed method used to reduce the high dimensional MI Dataset.

2.7. Classification

This work implemented a support vector machine (SVM) as a classifier to perform the classification task on the two MI datasets. The SVM classifier creates a discriminant hyperplane to improve the generalization capabilities by maximizing the margin between the classes. The SVM is a relatively fast algorithm and is capable of dealing with a large dataset [40]. Since the MI datasets used in this work have two classes (right-hand and left-hand motor imagery), the SVM model creates a hyperplane to differentiate the two classes in such a way that the gap between them is globally maximized.

3. Results

This section elucidates the feature selection and the classification performances of the proposed method in comparison with the baseline methods explained in this paper. Algorithms used in this work were developed and implemented on a computer having 12 GB of RAM and an Intel Core i7 (@ 3.4 GHz) processor using the 64-bit version of Matlab R2018a software and were applied to the two different BCI datasets explained in Section II. Dataset 1 has EEG data from one subject for one training session, while Dataset 2 consists of EEG data from nine different subjects obtained during two training sessions for each subject. Hence, the EEG recordings of 19 training sessions were used in this work.

3.1. Feature selection results

The selection of a particular feature is based on the weight of that feature calculated by the feature selection algorithm. The data of subject ID B0101T of BCI Competition IV Dataset 2b are chosen to represent the results of the feature selection methods. Specifically, preceded by preprocessing and feature extraction using the DTCWT, a feature space of dimension  $42 \times N$  features is generated.

Fig. 3(a) illustrates the weights evaluated for each feature using ReliefF. It shows the feature weights in descending order, and accordingly, a rank is provided to each feature; that is, the feature with the most significant weight has been assigned rank 1 and so on. For classification, the features with the positive weights are selected.

In PCA analysis, the contribution of each feature to the principal

components is evaluated by projecting the dataset on another coordinate, that is, the principal axis. Then, the absolute mean and variance of each feature in the principal coordinate are calculated and plotted as shown in Fig. 3(b). The features whose absolute mean is greater than the variance are selected for classification.

The performance of RNCA for feature selection is shown in Fig. 4. First, the best regularization parameter was determined as 0.0077 with an average classification loss of 0.105. Using this value of the regularization parameter, the weights of all the features were evaluated. Subsequently, a threshold of 5% of the maximum weight is set to select the features. It is notable that only six features have weights significantly more than 5% of the maximum weight, and hence these features are selected for classification.

Further, Figs. 3 and 4 show the feature selection procedure applied to EEG data of only one training session; the same methodology was applied to all the EEG data from the 19 training sessions. Fig. 5 presents the comparison of the average number of features selected using RNCA, ReliefF, PCA, and GA. Although all the algorithms considerably reduce the irrelevant features, it is notable that RNCA outperforms the other algorithms on the data from most of the training sessions with an average of  $6.6 \pm 1.88$  features selected. The number of features selected (FS) using various algorithms for each training session is listed in Table 2. However, the reduced number of features improves the execution speed of the classifier; it is also crucial that the selected subset of features must enhance the classification performance. The following section presents the classification results.

3.2. Classification performance

In the present study, two evolution criteria [41], namely the classification accuracy and the kappa coefficient, have been chosen to compare the effectiveness of the different feature selection approaches for the classification performance. The classification accuracy is defined as

$$classification\ accuracy = p_0 = \frac{\sum_{i=1}^C n_{ii}}{\sum_{i=1}^C \sum_{j=1}^C n_{ij}} \tag{6}$$

where  $n_{ii}$  and  $n_{ij}$  represent the elements of the confusion matrix and indicate how many times class  $i$  has been predicted as class  $j$ . If  $i = j$

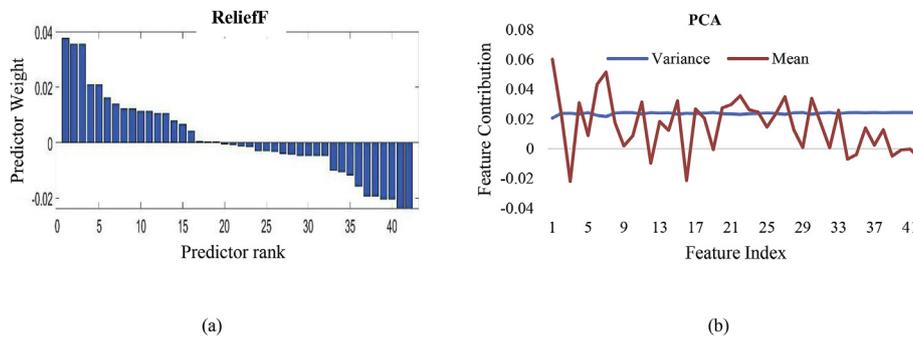


Fig. 3. Weights assigned to different features using (a) ReliefF algorithm and (b) Principal Component Analysis (PCA).

then a true class is predicted by the classifier.  $C$  is the number of classes, which is two in our case. The kappa coefficient is defined as

$$\text{kappa coefficient} = \frac{p_0 - p_e}{1 - p_e} \tag{7}$$

where  $p_e$  is the expected accuracy, which turns out to be 0.5 for a two-class problem. SVM is used as a classifier, and fivefold cross-validation has been applied to divide the data into training and test sets. The comparison of the classification accuracy, kappa coefficient, and the number of features selected using feature selection techniques such as RNCA, PCA, ReliefF, and GA is shown in Table 2. It is observable that the learning-based feature selection methods such as GA and RNCA perform better than rank-based feature selection methods such as ReliefF and PCA in terms of classification performance. Also, in learning-based feature selection approaches, RNCA achieved better classification performance than the GA. The average classification accuracy and kappa coefficient achieved by RNCA were 80.7% and 0.615, which are better values than those produced by GA, which were 78.9% and 0.579. The best classification accuracy of 99.2% was achieved for subject ID B0401T using RNCA. As can be seen from Table 2, on comparing the classification accuracy and kappa coefficient, RNCA achieved the highest values for all subjects, except for subject IDs BC11, B0302T, B0702T, B0801T, B0902T.

Besides the classification accuracy and kappa coefficient, the confusion matrix is further analysed in detail and parameters such as precision, recall or sensitivity, specificity, and F1-score are evaluated for each subject and averaged over the data from 18 training sessions of Dataset 2. These parameters range between 0 and 1, with 1 indicating the highest classification performance. Table 3 lists the average values of the confusion metrics obtained using different feature selection approaches. The obtained results suggest that the RNCA improved the classification performance in comparison to the other feature selection algorithms.

Further, the Friedman test and a Wilcoxon signed ranks test as the post hoc statistical test are performed to statistically validate the results

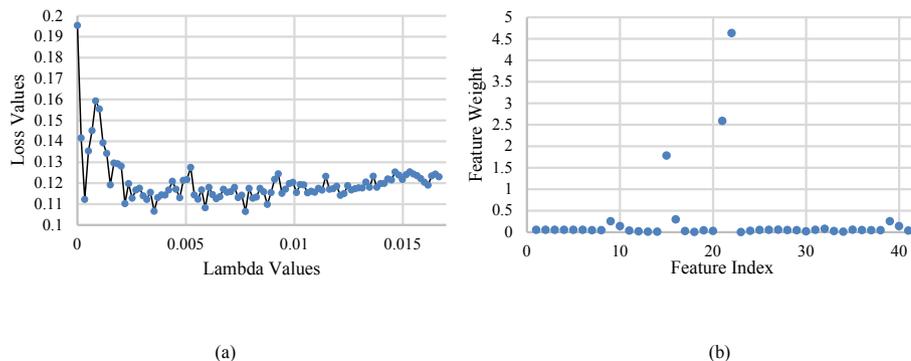


Fig. 4. Regularized Neighbourhood component analysis as feature selection (a) Estimation of the regularization parameter  $\lambda_{Best}$  at minimum loss value. (b) Feature weights calculated using  $\lambda_{Best} = 0.0077$ .

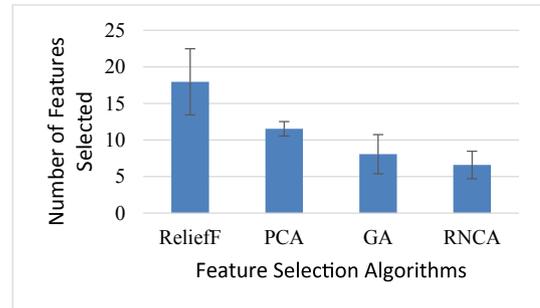


Fig. 5. Comparison of the average number of features selected between four feature selection algorithms: ReliefF, PCA, GA, and RNCA for 19 different training sessions of 2 motor imagery datasets from BCI competition II (dataset III) and IV (dataset 2b). It also indicates the standard deviation from the mean value.

given in Table 2. The null hypothesis assumes that the performance of all the algorithms is identical. After ranking all the algorithms for each subject ID separately, the Friedman test averages the ranks over all the subject IDs and calculates the p-value. A low p-value indicates that the null hypothesis is rejected and that there is a statistical difference. To find out which algorithm has the source of statistical difference, the Wilcoxon signed ranks test is conducted as the post hoc statistical test. The algorithm with the best rank is selected as the control algorithm, and then a pairwise comparison of all the other algorithms and the control algorithm is performed. The Friedman test was applied to the classification accuracies obtained in Table 2 of Dataset 2 and it was found that the null hypothesis was rejected ( $p = 2.55e-9$ ). The RNCA achieved the best rank of 1.33 and was selected as the control algorithm to perform the Wilcoxon signed ranks as the post hoc test. Table 4 shows the mean rank obtained across the data from 18 training sessions of Dataset 2, and Table 5 presents the p values obtained by pairwise comparison of the algorithms. The results of the statistical tests indicate

**Table 2**

Comparison of classification accuracy (CA), kappa coefficient and the number of features selected (FS) between ReliefF, PCA, GA and RNCA for motor imagery data of 19 different training sessions from BCI competition II (dataset III) and IV (dataset 2b). Values in boldness indicate the largest value compared with all others.

	Subject ID	All Features		ReliefF			PCA			GA			RNCA			
		CA	Kappa	FS	CA	Kappa	FS	CA	Kappa	FS	CA	Kappa	FS	CA	Kappa	FS
Dataset 1	BCI1	77.1	0.542	42	80.7	0.614	26	69.3	0.586	11	<b>81.4</b>	<b>0.628</b>	7	80.7	0.614	11
Dataset 2	B0101T	84.2	0.684	42	88.3	0.766	16	72.2	0.444	15	91.7	0.834	19	<b>93.3</b>	<b>0.866</b>	6
	B0102T	82.5	0.650	42	91.7	0.834	10	65	0.300	12	91.7	0.834	8	<b>92.5</b>	<b>0.850</b>	7
	B0201T	62.5	0.250	42	64.2	0.284	22	57.5	0.150	12	72.5	0.450	8	<b>73.3</b>	<b>0.466</b>	16
	B0202T	60.0	0.200	42	62.5	0.250	12	56.7	0.134	15	70.8	0.416	8	<b>73.3</b>	<b>0.466</b>	7
	B0301T	58.3	0.166	42	64.2	0.284	13	54.2	0.084	11	66.7	0.334	4	<b>67.5</b>	<b>0.350</b>	7
	B0302T	49.2	-0.016	42	51.7	0.234	33	50.8	0.016	10	<b>54.2</b>	<b>0.084</b>	4	49.2	-0.016	3
	B0401T	87.5	0.75	42	80.0	0.600	10	97.5	0.950	12	95.8	0.916	25	<b>99.2</b>	<b>0.984</b>	4
	B0402T	92.1	0.842	42	91.4	0.828	38	77.1	0.542	10	<b>94.3</b>	<b>0.886</b>	3	<b>94.3</b>	<b>0.886</b>	5
	B0501T	75.8	0.516	42	90.0	0.800	11	64.2	0.284	9	<b>91.7</b>	<b>0.834</b>	6	<b>91.7</b>	<b>0.834</b>	6
	B0502T	<b>73.6</b>	<b>0.472</b>	42	72.1	0.442	13	70.0	0.600	15	70.7	0.414	7	<b>73.6</b>	<b>0.472</b>	8
	B0601T	79.2	0.584	42	80.8	0.616	30	60.0	0.200	11	86.7	0.734	8	<b>94.2</b>	<b>0.884</b>	3
	B0602T	76.7	0.534	42	85.0	0.700	12	60.8	0.216	9	87.5	0.750	5	<b>88.3</b>	<b>0.766</b>	2
	B0701T	67.5	0.350	42	65.8	0.316	19	59.2	0.184	11	65.8	0.316	5	<b>70.8</b>	<b>0.416</b>	11
	B0702T	61.7	0.234	42	<b>65.8</b>	<b>0.316</b>	23	50.8	0.016	9	63.3	0.266	11	64.2	0.284	9
	B0801T	67.5	0.350	42	73.1	0.462	13	62.5	0.250	13	<b>78.8</b>	<b>0.576</b>	5	78.1	0.562	2
	B0802T	75.8	0.516	42	65.8	0.316	7	70.0	0.400	10	78.3	0.566	6	<b>80.8</b>	<b>0.616</b>	6
	B0901T	67.5	0.35	42	85.0	0.700	8	66.7	0.334	11	75.0	0.500	6	<b>88.3</b>	<b>0.766</b>	2
B0902T	73.3	0.466	42	81.7	0.634	25	57.5	0.150	13	<b>83.3</b>	<b>0.666</b>	8	81.7	0.634	11	
Mean Values		72.2	0.444	42	75.7	0.526	18	64.3	0.307	11.52	78.9	0.579	8	<b>80.7</b>	<b>0.615</b>	6.6

**Table 3**

Averaged classification performance metrics of SVM classifier on BCI competition IV dataset 2b present a comparison between compared feature selection methods. Values in boldness indicate the largest value compared with all others.

	All Features	ReliefF	PCA	GA	RNCA
Precision	0.7103	0.7518	0.6454	0.8277	<b>0.8541</b>
Recall or Sensitivity	0.6836	0.7408	0.6287	0.7408	<b>0.7895</b>
Specificity	0.7305	0.7483	0.6785	0.8224	<b>0.8341</b>
F1 Score	0.6951	0.7419	0.6355	0.7625	<b>0.8089</b>

**Table 4**

Mean rank of feature selection methods for the Friedman test.

Feature Selection Method	All Features	ReliefF	PCA	GA	RNCA
Mean Rank	3.83	2.97	4.52	2.33	1.33

**Table 5**

Statistical analysis of classification accuracy differences between the compared methods on BCI competition IV Dataset Iib: Results of Wilcoxon signed ranks test as post hoc with alpha = 0.05.

Hypothesis	RNCA vs. All Features	RNCA vs. PCA	RNCA vs. ReliefF	RNCA vs. GA
P-value (alpha = 0.05)	0.000438	0.001169	0.000232	0.017135

that the classification performance of SVM with RNCA used for feature selection is better than that obtained without feature selection ( $p < 0.05$ ), feature selection with ReliefF ( $p < 0.05$ ), PCA ( $p < 0.05$ ), and GA ( $p < 0.05$ ).

**4. Discussion**

The present study analysed the effectiveness of the proposed feature selection approach RNCA in MI EEG classification. Although comparison with some baseline feature selection methods has proven the superiority of RNCA, further discussion of the mechanism of the algorithm is provided in this section to present the potential advantages and limitations of the proposed method.

**4.1. Selection of optimal features by RNCA**

For MI EEG pattern recognition, sensory motor rhythms (SMRs) are analysed to discriminate between two classes of motor imagery [3]. In various studies, it is noted that the selection of an inappropriate frequency band leads to suboptimal classification performance; hence care has to be taken while selecting a frequency band [9,42]. In our approach, filtering of EEG into three sub-bands with frequency ranges of 4–8 Hz, 8–15 Hz, and 15–30 Hz, respectively, is performed by DTCWT, and accordingly, time, frequency, and phase features are extracted. Afterwards, features with greater importance have been selected using RNCA. The results presented in Tables 2–5 show that RNCA selects a smaller number of features with increased classification performance compared to GA, ReliefF, and PCA. Another finding of the results is that the frequency domain feature, that is, the power spectral density (PSD) of the channels C3 and C4 in the frequency range 8–15 Hz, has high importance and has been selected for eight out of ten subjects (see Fig. 6). These results are backed by numerous studies [3,31,43] that have suggested that the PSD of the mu rhythm (8–13 Hz) carries vital information about motor imagery tasks. Furthermore, as we have used multiple feature extraction methods, our proposed algorithm RNCA has selected significant features which are different for different subjects, which shows that RNCA is a robust algorithm for use in the design of a subject-specific BCI system.

However, although RNCA has shown potential in selecting optimal features from appropriate frequency bands, a few studies [9,44] have demonstrated the importance of choosing a potential time window, since the neural response time to different MI tasks is subject specific. Therefore, use of a fixed time window may degrade the classification performance of the MI BCI system, and this is worth our consideration. Keeping this in view, our work can be further extended by using shifted and varying time windows, but this will also increase the computational cost.

**4.2. Configurable parameter**

So far, numerous wrapper structure-based feature selection approaches have been proposed to improve the classification performance of MI EEG [9,15,45,46]. The experimental results have shown that the behaviour of most of these algorithms is highly variant because of the diversity in the selection of the configurable parameters. The parameter

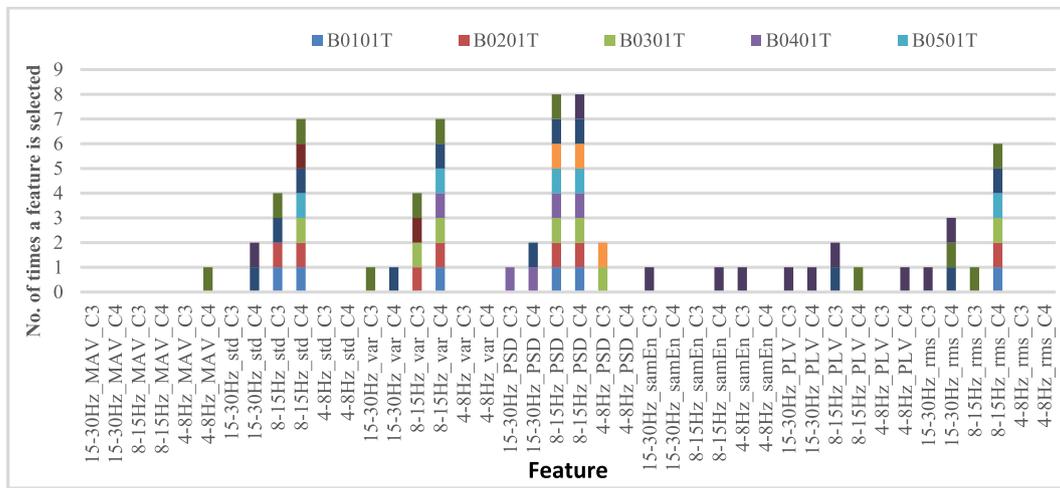


Fig. 6. The number of times a feature is selected by RNCA for ten subjects. Results indicate that the RNCA feature selection method is capable of selecting optimal features occur at different frequency bands for different subjects.

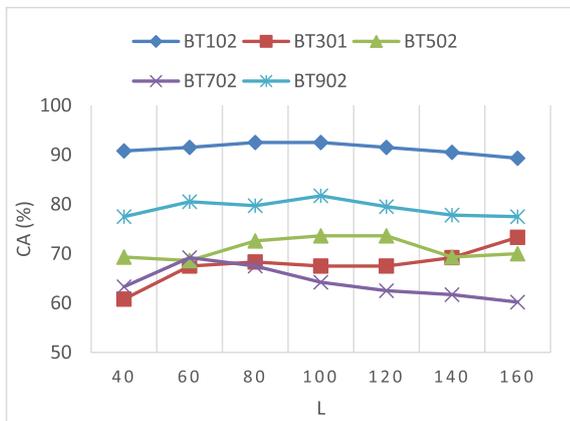


Fig. 7. Change in classification Accuracy (CA) with respect to parameter L of RNCA for five subjects BT102, BT301, BT502, BT702, and BT902, respectively.

controlling the nature of our approach is the length  $L$  of the uniformly distributed array  $\lambda_{val}$ . We have investigated the classification accuracy against varying values of  $L$ . In particular, we varied the value of  $L$  from 40 to 160 in steps of 20 and evaluated the classification accuracy for five subjects (see Fig. 7). It should be noted that the stability of our proposed algorithm is affected by varying values of  $L$  to some extent. For most of the subjects, CA is very stable over varying values of  $L$ , especially in the range from 80 to 120. Although this range of the parameter  $L$  has given a potentially improved classification accuracy, this method of selecting the optimal range of parameter  $L$  decreases the processing speed of the algorithm and also requires additional datasets. This limits the practicality of the proposed feature selection method to some extent. Hyperparameter optimization techniques in BCI have been introduced in some recent studies [47,48] and have shown potential in automatic selection of the optimal hyperparameter. As an extension of the proposed feature selection approach, some parameter optimization methods can be used to improve the BCI performance, which is worth our future consideration.

4.3. Computational complexity

The complexity of the feature selection algorithm increases the total processing time. Table 6 lists the processing time, training time, and testing time of the different algorithms used in our work. From the results, it can be observed that feature selection methods such as RNCA and GA that use a learning algorithm to optimize the classification

Table 6

Comparison of the average processing time of four feature selection methods: ReliefF, PCA, GA, and RNCA. SVM. Values in boldness indicate the largest value compared with all others.

	ReliefF	PCA	GA	RNCA
Algorithm Processing Time (s)	1.257415	<b>0.3984</b>	38.9404	22.1616
Training Time (s)	0.3478	0.3598	0.3758	<b>0.3137</b>
Testing Time (s)	0.0392	0.0250	0.0183	<b>0.0124</b>

performance have higher computational times than PCA and ReliefF. On the whole, the processing time of PCA is the fastest, with an average of 0.3984 s, while the processing times of RNCA and GA are 22.16 s and 38.94 s, respectively. Although the RNCA feature selection time is higher due to the inner loop for estimation of the classification error, it is a one-time procedure and does not affect the testing time. Also, due to the exclusion of a higher number of irrelevant features, RNCA with the SVM classifier built a prediction model that attained a faster classification speed compared with the prediction models made using GA, PCA, and ReliefF as feature selection methods and the SVM classifier.

For the design of a practical BCI system, it is essential that the system responds to different subjects with the same efficiency. Since the neural response to the MI task is subject and frequency band specific, high system performance cannot be achieved with a fixed set of features. Hence, there is a need for a practical feature selection approach capable of selecting an optimal subset of features for different subjects. This experimental study has demonstrated that the proposed RNCA feature selection method can be used efficiently for different subjects. Although the present study is conducted in an offline scenario, this work will be further extended to implement a real-time MI BCI system to control a prosthetic arm.

5. Conclusion

This study presents the effectiveness of RNCA as a feature selection method to enhance the classification performance of motor imagery tasks on datasets provided by BCI Competitions II and IV. For comparative study, we used two rank-based feature selection methods, PCA and ReliefF, and one learning-based method, GA. The SVM classifier has been used to evaluate the classification performance of each algorithm. We have investigated whether the subset of features estimated by the proposed RNCA algorithm successfully eliminated the irrelevant features and improved the overall classification performance of the SVM classifier for a two-class motor imagery problem. The results reveal that

the RNCA performed better on the feature selection task when compared with PCA, ReliefF, and GA. An important issue is the processing time of RNCA. Due to the diversity in the selection of configurable parameters that control the nature of RNCA and GA, the computational cost of these methods is higher than that of the rank-based feature selection methods. However, the execution time of the RNCA is faster than that of GA. It is also concluded that due to the reduction in the feature space, the training and testing times of the prediction model were faster for RNCA than for ReliefF, PCA, and GA. Further, this work can be extended by varying the feature space dimensions and optimizing the configurable parameters that control the nature of RNCA to attain a reasonable processing speed. This study concludes that the feature selection using RNCA is the best choice to discriminate between different motor-related mental tasks where high classification accuracy with a reasonable processing speed is required.

## References

- U. Chaudhary, N. Birbaumer, A. Ramos-Murguialday, Brain-computer interfaces for communication and rehabilitation, *Nat. Rev. Neurol.* 12 (2016) 513–525, <https://doi.org/10.1038/nrneuro.2016.113>.
- M. Hamed, S.-H. Salleh, A.M. Noor, Electroencephalographic motor imagery brain connectivity analysis for BCI: a review, *Neural Comput.* 28 (2016) 999–1041, [https://doi.org/10.1162/NECO\\_a\\_00838](https://doi.org/10.1162/NECO_a_00838).
- G. Pfurtscheller, F.L. Da Silva, Event-related EEG/MEG synchronization and de-synchronization: basic principles, *Clin. Neurophysiol.* 110 (1999) 1842–1857.
- S.N. Abdulkader, A. Atia, M.-S.M. Mostafa, Brain computer interfacing: applications and challenges, *Egyptian Informatics Journal* 16 (2015) 213–230, <https://doi.org/10.1016/j.eij.2015.06.002>.
- J.N. Mak, J.R. Wolpaw, Clinical applications of brain-computer interfaces: current state and future prospects, *IEEE Rev. Biomed. Eng.* 2 (2009) 187–199, <https://doi.org/10.1109/RBME.2009.2035356>.
- J. Cantillo-Negrete, R.I. Carino-Escobar, P. Carrillo-Mora, D. Elias-Vinas, J. Gutierrez-Martinez, Motor imagery-based brain-computer interface coupled to a robotic hand orthosis aimed for neurorehabilitation of stroke patients, *J. Healthc. Eng.* (2018), <https://doi.org/10.1155/2018/1624637>.
- M. Grosse-Wentrup, C. Liefhold, K. Gramann, M. Buss, Beamforming in noninvasive brain-computer interfaces, *IEEE Trans. Biomed. Eng.* 56 (2009) 1209–1219, <https://doi.org/10.1109/TBME.2008.2009768>.
- G. Pfurtscheller, C. Neuper, Motor imagery and direct brain-computer communication, *Proc. IEEE* 89 (2001) 1123–1134, <https://doi.org/10.1109/5.939829>.
- Y. Zhang, C.S. Nam, G. Zhou, J. Jin, X. Wang, A. Cichocki, Temporally constrained sparse group spatial patterns for motor imagery BCI, *IEEE Trans. Cybern.* (2018) 1–11, <https://doi.org/10.1109/TCYB.2018.2841847>.
- Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, A. Cichocki, Sparse bayesian classification of EEG for brain-computer interface, *IEEE Trans. Neural Netw. Learn. Syst.* 27 (2016) 2256–2267, <https://doi.org/10.1109/TNNLS.2015.2476656>.
- H. Wang, Y. Zhang, N.R. Waytowich, D.J. Krusienski, G. Zhou, J. Jin, X. Wang, A. Cichocki, Discriminative feature extraction via multivariate linear regression for SVEP-based BCI, *IEEE Trans. Neural Syst. Rehabil. Eng.* 24 (2016) 532–541, <https://doi.org/10.1109/TNSRE.2016.2519350>.
- Y. Zhang, G. Zhou, Q. Zhao, J. Jin, X. Wang, A. Cichocki, Spatial-temporal discriminant analysis for ERP-based brain-computer interface, *IEEE Trans. Neural Syst. Rehabil. Eng.* 21 (2013) 233–243, <https://doi.org/10.1109/TNSRE.2013.2243471>.
- S.-Y. Dong, B.-K. Kim, S.-Y. Lee, EEG-based classification of implicit intention during self-relevant sentence reading, *IEEE Trans. Cybern.* 46 (2016) 2535–2542, <https://doi.org/10.1109/TCYB.2015.2479240>.
- Z. Jin, G. Zhou, D. Gao, Y. Zhang, EEG Classification Using Sparse Bayesian Extreme Learning Machine for Brain-Computer Interface, *Neural Computing and Applications*, (2018), pp. 1–9, <https://doi.org/10.1007/s00521-018-3735-3>.
- A. Liu, K. Chen, Q. Liu, Q. Ai, Y. Xie, A. Chen, Feature selection for motor imagery EEG classification based on firefly algorithm and learning automata, *Sensors* 17 (2017), <https://doi.org/10.3390/s17112576>.
- A. Rakotomamonjy, V. Guigue, G. Mallet, V. Alvarado, Ensemble of SVMs for improving brain computer interface P300 speller performances, *Proceedings of the 15th International Conference on Artificial Neural Networks: Biological Inspirations - Volume Part I*, Springer-Verlag, Berlin, Heidelberg, 2005, pp. 45–50, [https://doi.org/10.1007/11550822\\_8](https://doi.org/10.1007/11550822_8).
- X. Yu, P. Chum, K.-B. Sim, Analysis the effect of PCA for feature reduction in non-stationary EEG based motor imagery of BCI system, *Optik* 125 (2014) 1498–1502, <https://doi.org/10.1016/j.jleo.2013.09.013>.
- X. Guo, X. Wu, X. Gong, L. Zhang, Envelope detection based on online ICA algorithm and its application to motor imagery classification, 2013 6th International IEEE/EMBS Conference on Neural Engineering, NER, 2013, pp. 1058–1061, <https://doi.org/10.1109/NER.2013.6696119>.
- S. Bhattacharyya, P. Rakshit, A. Konar, D.N. Tibarewala, R. Janarthanan, Feature selection of motor imagery EEG signals using firefly temporal difference Q-learning and support vector machine, in: B.K. Panigrahi, P.N. Suganthan, S. Das, S.S. Dash (Eds.), *Swarm, Evolutionary, and Memetic Computing*, Springer International Publishing, 2013, pp. 534–545.
- Y. Wang, V. Gonuguntla, G. Shafiq, K.C. Veluvolu, BMFLC with neural network and DE for better event classification, 2013 International Winter Workshop on Brain-Computer Interface, BCI, 2013, pp. 34–35, <https://doi.org/10.1109/IWW-BCI.2013.6506621>.
- P. Rakshit, S. Bhattacharyya, A. Konar, A. Khasnobish, D.N. Tibarewala, R. Janarthanan, Artificial bee Colony based feature selection for motor imagery EEG data, in: J.C. Bansal, P. Singh, K. Deep, M. Pant, A. Nagar (Eds.), *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)*, Springer India, 2013, pp. 127–138.
- Feature selection on movement imagery discrimination and attention detection, (n.d.), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2946110/> (accessed February 5, 2019).
- A.C. Ramos, R.G. Hernández, M. Vellasco, Feature Selection methods applied to Motor Imagery task classification, 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCL), 2016, pp. 1–6, <https://doi.org/10.1109/LA-CCL.2016.7885731>.
- J. Goldberger, G.E. Hinton, S.T. Roweis, R.R. Salakhutdinov, Neighbourhood components analysis, *Advances in Neural Information Processing Systems*, 2005, pp. 513–520.
- Z. Yang, J. Laaksonen, Regularized neighborhood component analysis, in: B.K. Ersbøll, K.S. Pedersen (Eds.), *Image Analysis*, Springer Berlin Heidelberg, 2007, pp. 253–262.
- B. Blankertz, K. Müller, G. Curio, T.M. Vaughan, G. Schalk, J.R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schroder, N. Birbaumer, The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials, *IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng.* 51 (2004) 1044–1051, <https://doi.org/10.1109/TBME.2004.826692>.
- M. Tangermann, K.-R. Müller, A. Aertens, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K.J. Miller, G. Mueller-Putz, Review of the BCI competition IV, *Front. Neurosci.* 6 (2012) 55.
- K.N. Chaudhury, M. Unser, On the shiftability of dual-tree complex wavelet transforms, *IEEE Trans. Signal Process.* 58 (2010) 221–232.
- T. Zikov, S. Bibian, G.A. Dumont, M. Huzmezan, C.R. Ries, A wavelet based de-noising technique for ocular artifact correction of the electroencephalogram, *Engineering in Medicine and Biology*, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint, IEEE, 2002, pp. 98–105.
- R.K. Chaurasiya, N.D. Londhe, S. Ghosh, Statistical wavelet features, PCA, and SVM based approach for EEG signals classification, *Int. J. Electr. Comput. Electron. Commun. Eng.* 9 (2015) 182–186.
- S.N. Resalat, V. Saba, A study of various feature extraction methods on a motor imagery based brain computer interface system, *Basic Clin. Neurosci.* 7 (2016) 13–19.
- C. Kim, J. Sun, D. Liu, Q. Wang, S. Paek, An effective feature extraction method by power spectral density of EEG signal for 2-class motor imagery-based BCI, *Med. Biol. Eng. Comput.* (2018) 1–14.
- W.-Y. Hsu, Wavelet-coherence features for motor imagery EEG analysis posterior to EOG noise elimination, *Int. J. Innov. Comput. Inf. Contr.* 9 (2013) 465–475.
- D. Whitley, A genetic algorithm tutorial, *Stat. Comput.* 4 (1994) 65–85.
- O.H. Babatunde, L. Armstrong, J. Leng, D. Diepeveen, A Genetic Algorithm-Based Feature Selection, (2014).
- C. Boutsidis, M.W. Mahoney, P. Drineas, Unsupervised feature selection for principal components analysis, *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2008, pp. 61–69.
- F. Song, Z. Guo, D. Mei, Feature selection using principal component analysis, *System Science, Engineering Design and Manufacturing Informatization (ICSEM)*, 2010 International Conference on, IEEE, 2010, pp. 27–30.
- I. Kononenko, Estimating attributes: analysis and extensions of RELIEF, *European Conference on Machine Learning*, Springer, 1994, pp. 171–182.
- W. Yang, K. Wang, W. Zuo, Neighborhood component feature selection for high-dimensional data, *J. Clin. Pharm.* 7 (2012) 161–168.
- J. Shawe-Taylor, S. Sun, A review of optimization methodologies in support vector machines, *Neurocomputing* 74 (2011) 3609–3618.
- A. Schlögl, J. Kronegg, J. Huggins, S. Mason, 19 Evaluation Criteria for Bci Research, toward Brain-Computer Interfacing, (2007).
- K.K. Ang, Z.Y. Chin, H. Zhang, C. Guan, Filter bank common spatial pattern (FBCSP) in brain-computer interface, *Neural Networks*, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on, IEEE, 2008, pp. 2390–2397.
- Y. Wang, Y.-T. Wang, T.-P. Jung, Translation of EEG spatial filters from resting to motor imagery using independent component analysis, *PLoS One* 7 (2012), <https://doi.org/10.1371/journal.pone.0037665>.
- J. Feng, E. Yin, J. Jin, R. Saab, I. Daly, X. Wang, D. Hu, A. Cichocki, Towards correlation-based time window selection method for motor imagery BCIs, *Neural Network*. 102 (2018) 87–95, <https://doi.org/10.1016/j.neunet.2018.02.011>.
- M.S. Uzer, N. Yilmaz, O. Inan, Feature selection method based on artificial bee Colony algorithm and support vector machines for medical datasets classification, *J. Sci. World* (2013), <https://doi.org/10.1155/2013/419187> 2013.
- R. Corralajo, R. Hornero, D. Álvarez, Feature selection using a genetic algorithm in a motor imagery-based Brain Computer Interface, 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 7703–7706, <https://doi.org/10.1109/IEMBS.2011.6091898>.
- B. Nakisa, M.N. Rastgoo, A. Rakotonirainy, F. Maire, V. Chandran, Long short term memory hyperparameter optimization for a neural network based emotion recognition framework, *IEEE Access* 6 (2018) 49325–49338, <https://doi.org/10.1109/ACCESS.2018.2868361>.
- H. Bashashati, R.K. Ward, A. Bashashati, Bayesian optimization of BCI parameters, 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 2016, pp. 1–5, <https://doi.org/10.1109/CCECE.2016.7726801>.