CHAPTER 5

5 Motor imagery EEG spectral-spatial feature optimization using Dual-Tree Complex Wavelet and Neighbourhood Component Analysis.

5.1 Introduction

Brain-computer interface (BCI) is a communication technique that builds a direct pathway between the human brain and an external device [1]**.** BCI systems recognize electroencephalogram (EEG) patterns, originated from the neural responses to different stimulus, using pattern recognition method [67]. In the past few years, BCI finds application in rehabilitation of stroke patients by facilitating them to communicate with surrounding environment [225]. The popular EEG patterns in BCI are the mu and beta rhythms, steady-state visual evoked potentials (SSVEP), and event-related potentials (ERP) [9,226–228].

In all the CSP based algorithms discussed in section 2.5, filter -- used for filtering the EEG into multiple sub bands -- was one from finite impulse response (FIR) or Infinite impulse response (IIR) filters. In EEG signal processing, filter type and properties (ripple, cut-off frequency, roll off, and attenuation) affect the temporal structure of the filtered EEG [231]. Although, there cannot be a certain recommendation on choice of a given filter type, properties or parameters, but to avoid filter artifacts and effects, it is recommended to analyse the effect of different filters on the EEG data and then cautiously select the filter [232].

In many studies [233–237]**,** wavelets decomposed the EEG signals into details of different frequency range and then reconstructed the EEG using details of a particular frequency band of interest. This provides filtering on the EEG. However, wavelets suffer with problem of power loss and aliasing at transition states [238]. Dual tree complex wavelet transform (DTCWT) overcomes these shortcomings of wavelet and provides nearly perfect reconstruction of the signal [209]. Moreover, with properties such as multidimensional direction selectivity and shift invariance, DTCWT is suitable for biomedical signal processing. In this study, we designed a DTCWT based filter and compared its ERD/ERS detection effectiveness on MI EEG data with three IIR filters: Butterworth, elliptical, and Chebyshev Type II.

Further, fisher score, and mutual information-based feature selection approaches are categorized under filter feature selection methods. A comparison, between filter and wrapper based algorithms, presented in [121], suggested the superiority of wrapper methods over filter methods in rejecting irrelevant features. A popular fast supervised learning based feature optimization algorithm namely, neighbourhood component analysis (NCA) is proposed in [239]. More recently, many MI feature optimization algorithms [28,240] are presented based on supervised learning and promised better classification performance than that by filter methods.

This work proposes a robust feature selection approach based on NCA for frequency band optimization to enhance the classification performance of MI signals. Further, a filter bank using DTCWT is designed to filter the EEG, instead of traditional filtering techniques. The DTCWT filter bank filters the EEG at three frequency bands 8-16 Hz, 16-24Hz and 24-32Hz. Then, CSP extracts the spatial features from EEG at multiple frequency bands. Afterwards, NCA with a regularization parameter evaluated the most relevant spatial features. Subsequently, selected features are fed into a support vector machine (SVM) to perform the MI classification task. Public BCI datasets (BCI Competition IV (Dataset 2b), and BCI completion III Dataset IIIa) validated the proposed method. A comparative study, with standard MI EEG feature extraction algorithms such as CSP with a filter of 8-13 Hz, CSP with a filter of 8-30Hz, filter bank CSP (FBCSP), discriminative FBCSP (DFBCSP), shows that the proposed algorithm achieved a higher classification performance.

Following sections in this chapter are organized as: first, elaborating the methods used for MI classification, followed by a brief description of the experimental study conducted on MI dataset; second, results and discussions are presented in the subsequent sections. Finally, last section presents the conclusion.

5.2 Methods and Materials

5.2.1 Notations

In this work, the following notations are used to represent scalers, vectors and matrices.

1. Scalers are denoted as italic letters.
2. Vectors are presented by uppercase letters, The *i*th element of a vector X is denoted as $X^{(i)}$
3. Matrices are indicated by boldface uppercase letters. The *i*th row and *j*th column of a matrix $\mathbf{X} = x_{i,j}$ are denoted as $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ respectively.

5.2.2 Band Pass Filter design using dual-tree complex wavelet transform

In this work, we designed a bandpass filter bank using a dual-tree complex wavelet transform (DTCWT) to avoid the artifacts and effects of filter. The DTCWT utilizes two discrete wavelet transforms (DWT) working in parallel to provide the real and imaginary coefficients of the transform. In DTCWT, a signal is decomposed using high pass and low pass filters simultaneously. The output gives coefficients of details (from high pass filter) and approximation (from low pass filter). The approximation is repeatedly decomposed into high and low pass filters after each level, whereas the details coefficients are the output coefficients.

Since ERD/ ERS patterns are prominently observed in mu (8-13Hz) and beta (13-30Hz) frequency ranges [14], we design a DTCWT filter bank to filter the EEG into three sub bands that has frequency ranges 8-16 Hz, 16-24 Hz, and 24- 32 Hz. For the filter design, we first decomposed the EEG using five-level DTCWT into four details (D4, D3, D2, and D1), and one approximation (A1). For first sub band filtering, we selected the DTCWT coefficients with sub band frequency between 8 and 16 Hz and omitted the other coefficients by setting zero in their position. Then we performed inverse DTCWT to reconstruct the filtered EEG signal. Similarly, we filtered the EEG into two more sub bands: 16-24Hz and 24-32 Hz.

The effective strength of ocular artifacts is up to 10 Hz [206]. In this study, we used an adaptive threshold method for removing ocular artifacts presented in [234].

5.2.3 Common Spatial Patterns

The commonly used technique for feature extraction in motor imagery based BCI research is common spatial patterns (CSP). Spatial filters of CSP extract spatial features for binary MI classification. To compute spatial features for the ith trial EEG $\mathbf{X}_i \in \mathbb{R}^{C \times T}$, we consider input parameter $\mathbf{X}_L$ and $\mathbf{X}_R$ representing the EEG matrices for the left-hand and right-hand MI task, respectively. Where C is the total number of channels and T is the total time points. First, normalized covariance EEG matrices $\mathbf{R}_L$ and $\mathbf{R}_R$ are evaluated as

$$\mathbf{R}_L = \frac{\mathbf{X}_L \mathbf{X}_L^T}{trace(\mathbf{X}_L \mathbf{X}_L^T)} \qquad \mathbf{R}_R = \frac{\mathbf{X}_R \mathbf{X}_R^T}{trace(\mathbf{X}_R \mathbf{X}_R^T)} \tag{5.1}$$

Where $\mathbf{X}^T$ gives the transpose of $\mathbf{X}$ and $trace(.)$ returns the sum of the diagonal values. Second, the sum of the covariance matrices $\mathbf{R}_L, \mathbf{R}_R$ is decomposed to obtain eigenvector matrix U_0, which is defined as

$$\mathbf{R} = \mathbf{R}_L + \mathbf{R}_R = \mathbf{U}_0\mathbf{E}\mathbf{U}_0^T \tag{5.2}$$

Where $\mathbf{E}$ is the diagonal eigenvalue matrix. Third, the whitening transformation matrix $\mathbf{P}$ is generated as $\mathbf{P} = \mathbf{E}^{-1/2}\mathbf{U}_0$. Fourth, the covariance matrix $\mathbf{S}_L$ and $\mathbf{S}_R$ are calculated as

$$\mathbf{S}_L = \mathbf{P}\mathbf{R}_L\mathbf{P}^T \qquad \mathbf{S}_R = \mathbf{P}\mathbf{R}_R\mathbf{P}^T \tag{5.3}$$

Also, S_L and S_R always have the same eigenvector matrix $\mathbf{U}$ and the sum of their eigenvalues is unity. In last, the spatial filter is designed as $\mathbf{A} = \mathbf{U}^T\mathbf{P}$. The ith trial EEG data matrix $\mathbf{X_i}$ is projected using spatial patterns as

$$\mathbf{Z} = \mathbf{A}\mathbf{X} \tag{5.4}$$

In general, the designed spatial filter consists of 2m columns selected from the first and last m columns of $\mathbf{A}$, which maximizes the variance of one class whereas minimizes the variance of the other class. Thus, the final feature vector has spatial features defined as logarithmic values of normalized variance of p row of $\mathbf{z}_p$, where $p = 1, \ldots.2m$. The generated CSP feature for ith trial EEG is written as

$$f_p = \log\left(\frac{var(\mathbf{z}_p)}{\sum_{p=1}^{2m}(var(\mathbf{z}_p))}\right) \tag{5.5}$$

Where $var(.)$ computes the variance. By applying filter bank before extracting CSP features, we created a feature set consisting of CSP features for all the trials and is defined as

$$\mathbf{FV} = \begin{pmatrix} f_{1,1} & \cdots & f_{1,N} \\ \vdots & \ddots & \vdots \\ f_{R,1} & \cdots & f_{R,N} \end{pmatrix} \tag{5.6}$$

Where R is the total number of trials, and N is the product $2m \times J$, where J is the total number of filters in the filter bank.

5.2.4 Neighbourhood Component Analysis

Neighbourhood Component Analysis learns the Mahalanobis distance metric to linearly transform the training dataset to a subspace such that the average leave-one-out classification accuracy gets maximized. Let W be the feature vector needs to be learned, like 1-nearest neighbour classifier, the probability p_{ij} of x_j being chosen as a reference point for x_i from all the samples is given as

$$p_{i,j} = \begin{cases} -\dfrac{\exp\left(-\left\|\mathrm{W}x_i - \mathrm{W}x_j\right\|^2\right)}{\sum_k \exp(-\|\mathrm{W}x_i - \mathrm{W}x_k\|^2)}, & j \neq i \\ 0, & j = i \end{cases} \tag{5.7}$$

Now, the probability of x_i accurately classified as its true class is expressed as

$$p_i = \frac{1}{R} \sum_{j=1, j \neq i}^{R} p_{ij} y_{ij} \tag{5.8}$$

where y_{ij} returns one for $y_i = y_j$ else its value is zero. Mathematically, Equation (5.8) defines the average leave-one-out classification accuracy evaluated over the train set. Since the target is to attain a higher classification accuracy, solve for transformation matrix W which maximizes Equation (5.8). When maximized, this objective function attains a higher classification accuracy. As can be seen from Equation (5.8), it can subject to overfitting for higher dimension data. In the work of [128] a regularization parameter,

λ is induced to stabilize the NCA objective function. With regularization parameter, λ and to solve for transformation matrix W the Equation (5.8) can be rewritten as

$$\widehat{\mathrm{W}} = \underset{w}{\operatorname{argmax}} \frac{1}{R} \sum_{i=1}^{R} p_i \ -\lambda \sum_{m=1}^{N} {w_m}^2 \tag{5.9}$$

Where w_m are the feature weights and R represents the total number of trials. The objective defined in Equation (5.9) is termed as the regularized NCA (RNCA). To solve the objective function defined by Equation (5.9) for $\widehat{\mathrm{W}}$, the conjugate gradient approach can be utilized. The argument of the maximum $\widehat{\mathrm{W}}$, is the weight vector that provides maximum classification accuracy. Based on the outcome of the weights, the best subset of features is selected by comparing the weights of each feature with a threshold value.

5.2.5 Proposed spectral-spatial feature optimization approach

Apart from frequency band optimization, application of proper filtering on the MI EEG is also very important to attain optimal classification performance. In this study, we devise an improved filtering method using DTCWT and features optimization approach using NCA with a regularization parameter to improve the classification performance of the MI based BCIs. Figure 5.1 shows an illustration of the proposed algorithm and Algorithm 1 elucidates the pseudocode of the proposed approach. First, we filtered the EEG dataset into three sub bands using the DTCWT and applied k-fold cross validation to generate train and test sets. The frequency ranges of the selected sub bands are 8-16 Hz, 16-24 Hz, and 24-32 Hz. Second, CSP extracted the spatial features from each sub band in train set. Considering three sub bands and CSP parameter, m, the dimension of the feature set is $3 \times 2 \times m = 6m$ features. Finally, we applied a feature optimization algorithm based on

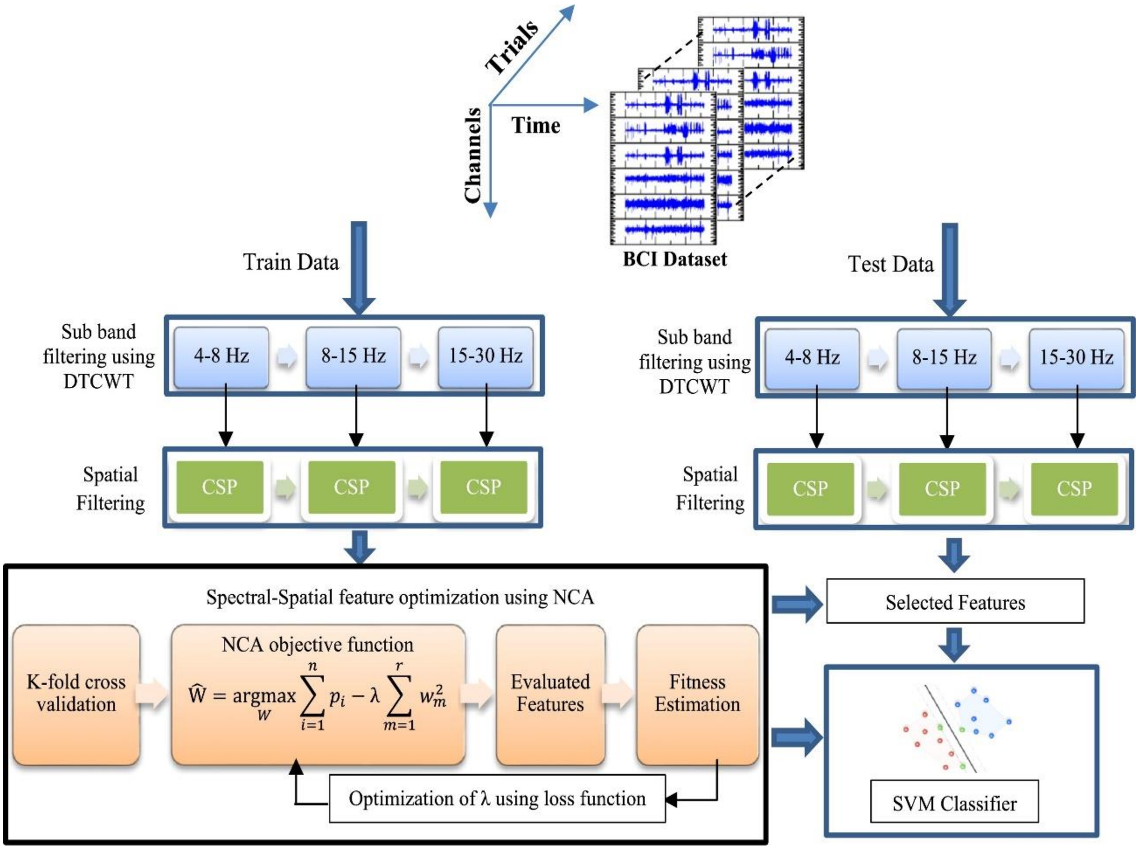

Figure 5.1 Workflow presents the working of the proposed MI spectral-spatial feature optimization algorithm.

NCA with a regularization parameter on the feature set to select the best subset of features. The target of the proposed approach is to select a subset of features which will maximize the classification accuracy or minimize the generalization loss. Following presents further details of the feature optimization algorithm.

Assume $\mathbf{FV} \in \mathbb{R}^{R \times N}$ is the feature set evaluated from CSP with filter bank according to Equation (5.6). Using k-fold cross-validation, calculate the generalization loss given by

$$G_{Loss} = \frac{1}{R} \sum_{i=1}^{R} C(k_i \neq t_i) \tag{5.10}$$

Where k_i, and t_i represents the predicted class, and the true class, respectively. Further, $C(x)$ is a conditional function which gives 1 when k_i is not equal to t_i else 0. The next step is to evaluate the feature weights using Equation (5.9), with λ value set to zero. Select the features with weights more than a threshold value T (we choose T as 5% of the maximum calculated feature weight). Create a new feature set with selected features only and again compute generalization loss given by Equation (5.10) using k-fold cross-validation. If the loss value with selected features is less than that achieved with all the features, it indicates that there is a need to optimize the features and further improve the

Algorithm 1: DTCWT-CSP (NCA)
Input: Input EEG data set, $\mathbf{X} \in \mathbb{R}^{C\times T}$
Output: Best feature subset $\mathbf{FV_n}$
Parameters:
N: total number of features = number of filters in filter bank × m CSP features
λ: Regularization parameter; an array of uniformly distributed values.
L: Length of array of λ values.
$Loss_a$: Generalization error before fitting NCA model.
$Loss_{nca}$: Generalization error after fitting NCA model with $\lambda = 0$.
$\mathrm{Loss_{value}}$: A vector storing values of generalization error for different λ values.
FV: Feature vector.
Algorithm:

1. **for** $j = 1, \ldots \ldots \ldots, N$ ***do***
   a. Apply DTCWT Band pass filter bank on all the trials.
   b. Divide the dataset into train and test sets using k-fold cross validation.
   c. Apply CSP on train set and evaluate the FBCSP features according to Eq. 5 and generate, a feature vector, $\mathbf{FV} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,N} \\ \vdots & \ddots & \vdots \\ a_{R,1} & \cdots & a_{R,N} \end{pmatrix}$
2. Divide the Feature vector $\boldsymbol{FV}$ into train and test sets using cross validation.
3. Evaluate $Loss_a$ defined by eq. 10 for $\boldsymbol{FV}$.
4. **for** $j = 1, \ldots \ldots \ldots \ldots, N$ ***do***
   Compute $\widehat{\mathrm{W}}(\lambda = 0)$ using eq. 9 with $\lambda = 0$
   if $(\widehat{\mathrm{W}}^{(j)}(\lambda = 0) > 0.02 \times \max(\widehat{\mathrm{W}}(\lambda = 0)))$ **then**,
   select j^{th} feature,
   else reject
   $\mathbf{FV_o}(\lambda = 0)$ = new feature vector of selected o features $(o < N)$.
   compute $Loss_{nca}(\lambda = 0)$ defined by eq. 10, use $\mathbf{FV_o}(\lambda = 0)$
5. **if** $(Loss_a < Loss_{nca})$ **then**
   Go to step 8
   else
   for $\lambda = 0, \ldots \ldots \ldots, L$ ***do***
   for $j = 1, \ldots \ldots \ldots \ldots, N$ ***do***
   Compute $\widehat{\mathrm{W}}(\lambda)$ using eq. 9
   if $(\widehat{\mathrm{W}}^{(j)}(\lambda) > 0.05 \times \max(\widehat{\mathrm{W}}(\lambda)))$ **then**,
   select j^{th} feature,
   else reject
   $\mathbf{FV_r}(\lambda)$ = new feature vector of selected r features $(r < N)$.
   compute $\mathrm{Loss_{value}}(\lambda)$ defined by eq. 10, use $\mathbf{FV_r}(\lambda)$
6. find(λ=λ_{best}) at min($\mathrm{Loss_{value}}$)
7. **for** $j = 1, \ldots \ldots \ldots \ldots, N$ ***do***
   $\widehat{\mathrm{W}}(\lambda_{best})$= feature weights computed by eq. 9
   if $(\widehat{\mathrm{W}}^{(j)}(\lambda_{best}) > 0.02 * \max(\widehat{\mathrm{W}}(\lambda_{best})))$ **then**,
   select j^{th} feature,
   else reject
   $\mathbf{FV_n}(\lambda_{best})$ = final feature vector of selected n features $(n < N)$
8. use feature vector **FV** or $\mathbf{FV_n}$ to train an SVM classifier.

classification accuracy otherwise select all the features. In this work, the features are optimized by tuning the regularization parameter using the following steps.

Step 1: Generate a linearly spaced vector of λ values, we used 20 values ranging from 0 to 2.

Step 2: For each value of λ, compute the feature weights using Equation (5.9), then compare the weights with the threshold T to select a subset of features. In parallel, using the selected features compute the generalization loss defined by Equation (5.10) and store its value for each λ value.

Step 3: Plot the stored generalization loss values versus the λ values to select $\lambda = \lambda_{Best}$ at which the generalization loss is minimum.

Step 4: Finally, use λ_{Best} value in Equation (5.9) to compute the final feature weights. Compare the feature weights with the threshold T and select the features. These selected features make the best subset of features. Next, train an SVM classifier using these features and then use the trained SVM classifier to identify motor imagery class of the test data. In the rest of the chapter, we have named this proposed method as DTCWT-CSP (NCA).

5.3 Experimental study

5.3.1 Datasets

The performance of the proposed method is validated on public EEG dataset, BCI Competition IV (Dataset 2b) [241]. The dataset consists of EEG signals acquired from nine subjects (named as B0103T, B0203T,…., and B0903T) while performing one of the motor imagery task from two classes: left-hand and right-hand. Total three sessions were recorded for each subject; however, this work used dataset from the third training session only. Further, Dataset consists of EEG from three channel locations, namely C3, CZ and C4. The sampling frequency is 250Hz. The total number of trials for each subject is 160, (80 for each class). The length of each trial is 7.5s. At time t=3s, a cue is presented on the screen for 1s. The cue indicated one of the class of MI task and subjects performed the MI task for that class for 4.5s. The dataset is pre-processed using a bandpass filter of passband range 0.5 to 100 Hz and a notch filter at 50 Hz.

The second dataset used in this study is Dataset IIIa from BCI Competition III [242]. It comprises of 60-channel EEG from three subjects named “K3”, “K6”, and “L1”, performing cue-based motor imagery task of left hand, right hand, tongue, or foot movement. The sampling frequency of the recording device was 250 Hz. The total number of trials for subject K3 is 180 (45 for each class), whereas that for K6 and L1 is 120 (30 for each class). The length of each trial is 7s. At time t=3s, a cue indicated one of the class of MI task on the screen and subjects performed the MI task for that class for 4s. The dataset was bandpass filtered between 1 and 50 Hz and a notch filter was applied to avoid power line noise.

5.3.2 Competing Methods

The aforementioned datasets are used to conduct an extensive experimental comparison among the following standard algorithms with the parameter settings as explained below.

- **CSP (8-13 Hz):** The EEG signals are bandpass filtered between 8 and 13 Hz, and spatial features are computed using CSP. Since ERD patterns are observed in mu rhythms (8-13Hz) of EEG during a motor imagery task, we opted this method to compare with the proposed algorithm. However, beta band also has MI task related information which can be used to further improve MI signal classification.
- **CSP (8-30 Hz):** The EEG signals are bandpass filtered between 8 and 30 Hz, and spatial features are computed using CSP. In this method, we analysed the ERD/ERS patterns of mu (8-13Hz) and beta (13-30 Hz) rhythms of EEG using a

wider fixed frequency band (8-30Hz), such that the extracted CSP features can be used for classification task. However, use of wider band may result in suboptimal classification results because neural response to MI tasks are subject and frequency specific.

- **FBCSP (MI):** As in [173], six non-overlapped bandpass filters with frequency range 4-40 Hz and bandwidth 6 Hz are designed. The spatial features from each filter are extracted using CSP. For features selection mutual information is used. Only four best features evaluated by mutual information-based feature selection are used for classification. Mutual information-based feature selection method is a filter feature selection approach. Studies related to MI- EEG feature selection [121,240] suggest that the learning-based feature selection methods are efficient than the filter methods. Hence, use of learning-based feature section method can further improve the MI classification performance.
- **DFBCSP (MI):** As in [178], instead of using a wider range bandpass filter we designed a shifted bandpass filter bank with a frequency range of 4-40Hz. In the designed filter bank, each filter has a passband frequency width of 4 Hz and is shifted by 2 Hz. CSP is applied to each filter separately. The top four spectral-spatial features are selected using mutual information-based feature selection approach.
- **Proposed Method:** Spectral- spatial features are optimized using the aforementioned DTCWT-CSP (NCA) method on motor imagery EEG dataset. Instead of using traditional filtering methods, we filtered the MI-EEG using an improved filtering method based on DTCWT and thus spatial features are extracted using CSP. Then, we optimized features using proposed algorithm based on NCA with a regularization parameter to improve the classification

performance of MI signals. Since our algorithm operates in a supervised framework hence it is more effective in selection of relevant MI task related spatial features than the mutual information-based feature selection algorithm used in methods: FBCSP (MI) and DFBCSP (MI).

Note: In the above competing methods (1-4), Chebyshev type II filtered the EEG data. The CSP parameter, m was set to 1 for BCI Competition IV (Dataset 2b) (because this dataset has only three electrodes, hence, 2×m cannot exceed 3, since in this case, the size of CSP matrix, A is 3×3) and 2 for Dataset IIIa from BCI Competition III. Further, an SVM classifier with a linear kernel is used to conduct MI classification task for all the algorithms.

5.3.3 Performance measures

The performance of the proposed method is assessed using the following popularly used measures in BCI systems [243].

- Classification accuracy: is defined as the rate of number of trials being correctly classified with respect to the total number of trials. CA is derived from the confusion matrix as follows

$$CA = \frac{\sum_{i=1}^{C} n_{ii}}{\sum_{i=1}^{C} \sum_{j=1}^{C} n_{ij}}$$

  Where n_{ij} represents elements of the confusion matrix. When $i = j$, the classifier has predicated the true class otherwise prediction is wrong. C is the total number of classes.

- Kappa Coefficient: It is a statistical estimator to test the classification performance in BCIs. It is defined as

$$kappa = \frac{CA - p_e}{1 - p_e}$$

Where p_e is the probability of occurrence of a class, which is 0.5 for two-class MI task classification.

5.4 Results

This section presents the result obtained by all the competing algorithms described in previous section. We develop all the algorithms presented in this work, on a 64-bit version of Matlab R2018a software installed in a computer having 12 GB of RAM and an Intel Core i7 (@ 3.4 GHz) processor and were applied to two different BCI datasets explained in Section II.

Tables 5.1-5.2 present the comparison of the classification accuracies and the kappa coefficient values achieved by the CSP (8-30Hz), CSP (8-13Hz), FBCSP (MI), DFBCSP (MI), and the proposed DTCWT-CSP (NCA) approaches, for BCI Competition IV (Dataset 2b), and BCI competition III Dataset IIIa. It can be seen from the results Tables that the mean classification accuracy achieved by the proposed method is improved by 6.92%, 6.5%, and 3.65% for BCI Competition IV (Dataset 2b) and 4.2%, 5.87%, and 4.54% for BCI competition III Dataset IIIa compared to that of the CSP, CSP (8-30Hz), CSP (8-13Hz), FBCSP (MI), and DFBCSP (MI), respectively. For BCI Competition IV (Dataset 2b), out of nine subjects, eight times the highest classification accuracy is obtained by the proposed method, whereas for five out of six runs (three subjects and two MI classification tasks -- Right-hand vs left-hand MI and tongue vs foot -- for each subject) greater results are obtained for BCI competition III Dataset IIIa using our approach compared to that using other methods.

The kappa coefficient is the second method that verified the reliability of the proposed method. Tables 5.1 and 5.2 list the obtained kappa coefficient for BCI Competition IV (Dataset 2b) and BCI competition III Dataset IIIa, respectively. It is notable that the proposed method achieved the best mean kappa coefficient of 0.680±0.24, and 0.782±0.15 for BCI Competition IV (Dataset 2b), and BCI competition III Dataset IIIa, respectively. The largest kappa coefficient is evaluated as 0.988 for subject "B0403T" for BCI Competition IV (Dataset 2b) and that as 0.978 for subject "K3b" for BCI competition III Dataset IIIa, utilizing the proposed method. Further, Paired t-test ($p<0.01$) is conducted on both the datasets to verify the statistical significance difference of classification accuracy between the proposed approach and each of other approaches. The results shown in Tables 5.1-5.2 indicate that the proposed method is performing better than the competing methods.

To investigate the effectiveness of NCA as a feature selection method, we compared the classification accuracies achieved and the total number of features selected by standard feature selection methods popularly used in various BCI studies such as ReliefF [244], Mutual information [245], and Genetic algorithm [246] with that by the proposed algorithm (see Table 5.3). We used BCI Competition IV (Dataset 2b) for this comparison. As can be seen in Table 5.3, NCA has not only selected a lesser number of features but also achieved a higher average classification accuracy in comparison with standard feature selection methods. Also, paired t-test is conducted between NCA and each of the competing feature selection algorithm to see statistical significance of the proposed feature selection method. Statistical results show that NCA is performing better than other feature selection methods ($p<0.01$).

Table 5.1 Comparison of classification accuracies (CA) (in %) achieved by the CSP (7-30 Hz), CSP (7-13 HZ), FBCSP (MI), DFBCSP (MI) and the proposed DTCWT-CSP (NCA) method respectively. CA is evaluated using an SVM classifier for BCI Competition IV (Dataset 2b). For each subject, Values in boldness indicate the largest value compared with all others. Further, p-values are obtained by the paired t-test between the results of DTCWT-CSP (NCA) and each of the other methods.

	CSP 8-30		**CSP 8-13**		**FBCSP (MI)**	
	CA	**kappa**	**CA**	**kappa**	**CA**	**kappa**
B0103T	70.6	0.412	73.8	0.476	76.3	0.526
B0203T	62.5	0.25	63.1	0.262	55.6	0.112
B0303T	63.4	0.268	**68.1**	**0.362**	69.4	0.388
B0403T	94.6	0.892	96.3	0.926	91.9	0.838
B0503T	77.5	0.55	72.5	0.45	78.9	0.578
B0603T	74.0	0.48	72.9	0.458	75.6	0.512
B0703T	80.0	0.6	81.0	0.62	84.4	0.688
B0803T	86.3	0.726	85.6	0.712	82.3	0.646
B0903T	85.0	0.7	84.4	0.688	81.2	0.624
Mean	77.1±10.7	0.542±0.21	77.52±10.2	0.550±0.20	77.3±10.2	0.545±0.20
p-value	***p<0.01***		***p<0.01***		***p<0.01***	

Table 5.1 (Continued)

	DFBCSP (MI)		**DTCWT-CSP (NCA)**	
	CA	**kappa**	**CA**	**kappa**
B0103T	80.3	0.606	**85.6**	**0.712**
B0203T	61.1	0.222	**66.3**	**0.326**
B0303T	59.3	0.186	63.7	0.274
B0403T	98.8	0.976	**99.4**	**0.988**
B0503T	86.3	0.726	**91.3**	**0.826**
B0603T	73.1	0.462	**79.2**	**0.584**
B0703T	85.0	0.7	**86.9**	**0.738**
B0803T	93.1	0.862	**94.4**	**0.888**
B0903T	86.3	0.726	**89.4**	**0.788**
Mean	80.37±13.5	0.607±0.27	**84.02±12.2**	**0.680±0.24**
p-value	***p<0.01***		**--**	

Table 5.2 Comparison of classification accuracies (%) achieved by the CSP (7-30 Hz), CSP (7-13 HZ), FBCSP (MI), DFBCSP (MI) and the proposed DTCWT-CSP (NCA) method respectively. CA is evaluated using an SVM classifier for BCI competition III Dataset IIIa. For each subject, Values in boldness indicate the largest value compared with all others. Further, p-values are obtained by the paired t-test between the results of DTCWT-CSP (NCA) and each of the other methods.

	CSP(8-30Hz)		**CSP(8-13Hz)**		**FBCSP (MI)**	
	CA	**kappa**	**CA**	**kappa**	**CA**	**kappa**
Left hand vs Right hand						
K3b	90.0	0.8	90.3	0.806	96.0	0.92
K6b	75.0	0.5	75.0	0.5	65.8	0.316
L1b	**90.0**	**0.8**	89.1	0.782	**94.2**	0.884
Tongue vs foot						
K3b	91.1	0.822	90.0	0.8	86.8	0.736
K6b	78.3	0.566	71.7	0.434	82.3	0.646
L1b	85.0	0.7	83.3	0.666	74.7	0.494
Mean	84.90±6.81	0.698±0.13	83.23±8.13	0.667±0.16	83.3±11.6	0.666±0.23
p-value	***p<0.01***		***p<0.01***		***p<0.01***	

Table 5.2 (Continued)

	DFBCSP (MI)		**DTCWT-CSP (NCA)**	
	CA	**kappa**	**CA**	**kappa**
Left hand vs Right hand				
K3b	96.7	0.934	**98.9**	**0.978**
K6b	68.3	0.366	**76.7**	**0.534**
L1b	84.7	0.694	88.8	0.776
Tongue vs foot				
K3b	91.1	0.822	**92.2**	**0.844**
K6b	88.3	0.766	**92.3**	**0.846**
L1b	78.3	0.566	**85.7**	**0.714**
Mean	84.56±10.1	0.691±0.2	**89.1±7.5**	**0.782±0.15**
p-value	***p<0.01***		**--**	

Table 5.3 Lists the classification accuracies (CA) (in %) achieved by feature selection methods: ReliefF, Mutual Information, Genetic Algorithm, and NCA for BCI Competition IV (Dataset 2b). The number of features selected (FS) by the algorithms are written in brackets. The feature extraction method used for all these algorithms is DTCWT-CSP and SVM classifier is trained to evaluate classification accuracies using k-fold cross validation. Values in boldness indicate the largest value compared with all others. Further, p-values are obtained using paired t-test between the results of DTCWT-CSP (NCA) and each of the other methods.

	ReliefF	**Mutual Information**	**Genetic Algorithm**	**NCA**
	CA (FS)	**CA (FS)**	**CA (FS)**	**CA (FS)**
B0103T	81.9 (3)	80.6 (4)	83.1 (3)	**85.6 (3)**
B0203T	56.3 (5)	60.6 (4)	60.1 (5)	**66.3 (1)**
B0303T	58.8 (6)	53.8(4)	59.7 (3)	**63.7 (2)**
B0403T	98.8 (5)	98.6 (4)	98.1 (5)	**99.4 (3)**
B0503T	72.5 (5)	71.3 (4)	86.7 (5)	**91.3 (3)**
B0603T	73.8 (4)	70 (4)	74.3 (6)	**79.2 (2)**
B0703T	84.4 (5)	84.4 (4)	85.8 (4)	**86.9 (2)**
B0803T	75 (2)	93.1 (4)	85.6 (4)	**94.4 (1)**
B0903T	85.6 (2)	83.8 (4)	86.3 (3)	**89.4 (1)**
Mean	76.3 (4.1)	77.4 (4)	80.0(4.2)	**84.02 (2)**
p-value	***p<0.01***	***p<0.01***	***p<0.01***	

Figure 5.2 shows the dispersion of features on a 2-D plot between the two best features selected by FBCSP (MI), DFBCSP (MI) and DTCWT-CSP (NCA) approaches for subject “B0503T” of BCI Competition IV (Dataset 2b) (left hand vs right hand MI), “k6b” of BCI competition III Dataset IIIa (left hand vs right hand MI), and “l1b” of BCI competition III Dataset IIIa (tongue vs foot MI). It can be seen that the features dispersion is more discriminative using the proposed approach in all three cases.

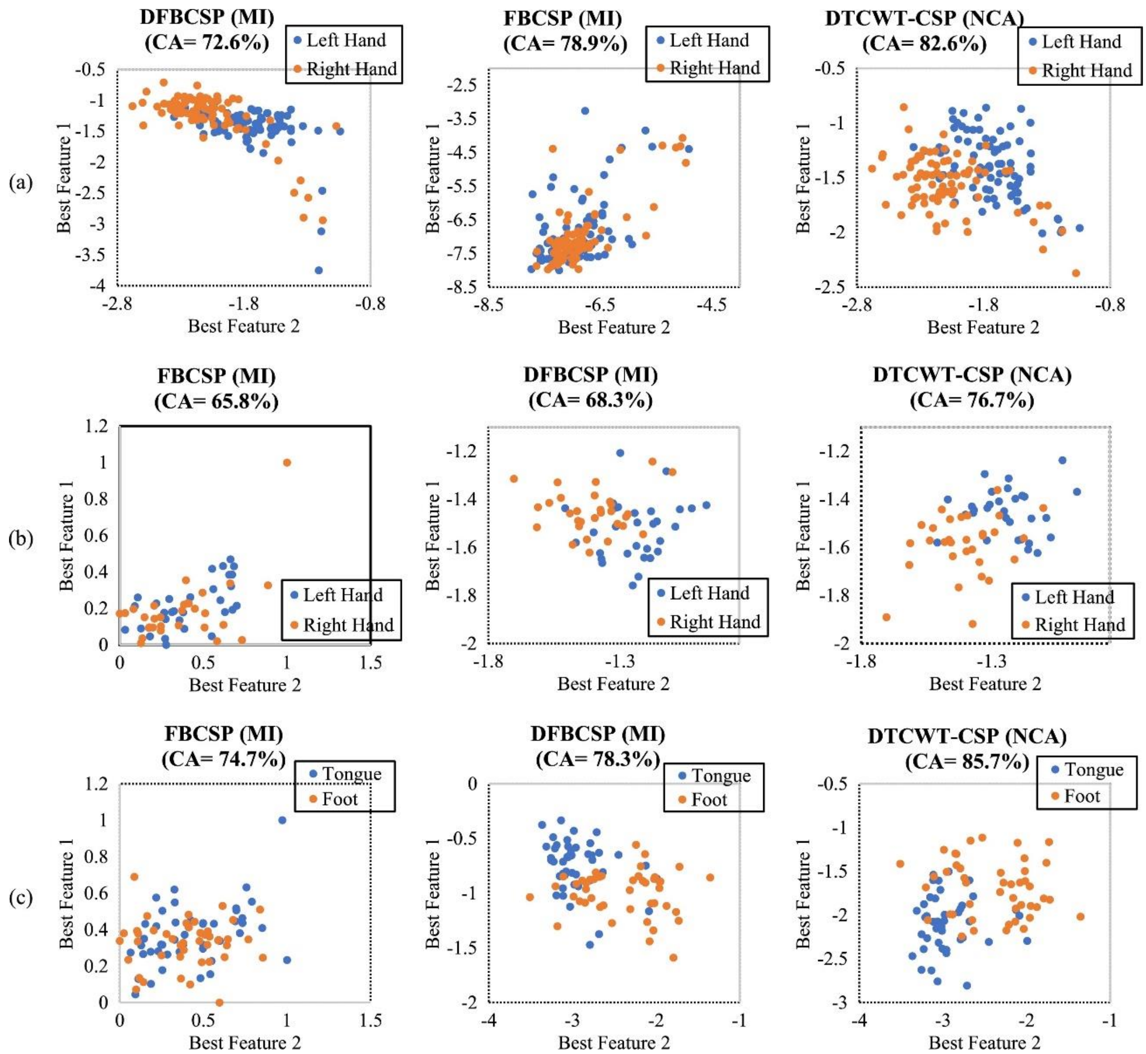

Figure 5.2 Dispersion of the best two features in 2-D, selected by the FBCSP (MI), DFBCSP (MI), and the proposed DTCWT-CSP (NCA) for (a) subject "B0503T" from BCI Competition IV (Dataset 2b), (b) subject "k6b" from BCI competition III Dataset IIIa right hand vs. left hand MI, and (c) subject "L1b" from BCI competition III Dataset IIIa for tongue vs. foot MI.

5.5 Discussion

5.5.1 Time- frequency analysis of DTCWT-filtered EEG

Motor imagery classification performance highly depends on feature extraction methods. So far, common spatial patterns and its advance variants are investigated for MI classification in most of the literature. A few studies [177,178,247] have suggested that the performance of CSP is prone to selection of proper frequency bands. Spatial features are extracted at different frequency bands. For division of raw EEG into various sub bands, digital filters are employed. However, the temporal structure of the filtered EEG depends on filter parameters: roll off, cut off frequency, ripple, stop band attenuation, order and type. Before implementing a filter on EEG data, it is recommended to analyse its effects on the data [232]. In this study, we designed a filter bank using DTCWT for division of MI EEG into sub bands, instead of using digital filtering methods. Subsequently, we compared the SMR recognition effectiveness of the designed filter with three IIR filters: Butterworth, elliptical and Chebyshev type II, using time-frequency analysis and averaged sub band power.

Figure 5.3 shows time-frequency plots of average of all the left-hand epochs and right-hand epochs of BCI Competition IV (Dataset 2b) (subject B0803T). Where each epoch represents a single trial for either of the class. In this dataset, there are 160 trials and 80 epochs for each class. In Figure 5.3, the first red line at time, t =3 s represents the start of the cue and the MI task. The cue was presented for 1s. Before averaging, we filtered the epochs using DTCWT based filter bank, and three IIR filters: Butterworth, elliptical and Chebyshev type II in frequency band between 8 and 32 Hz. Time-frequency plots represent significant SMRs during the MI task in 8-16 Hz (see Figure 5.2). Notice that the power in C3 channel is higher than that in C4 channel during left -hand MI task. Whereas, power in C4 channel is higher than that in C3 channel during right-hand MI. Further, in Figure 5.3, coloured topographic head plots show the averaged power in the frequency sub band (8-16Hz) during MI task (3-7.5 s). It is notable that the SMRs are more significant in first sub band of DTCWT based filtered EEG (8-16 Hz) than that in IIR (8-16Hz) filtered EEG.

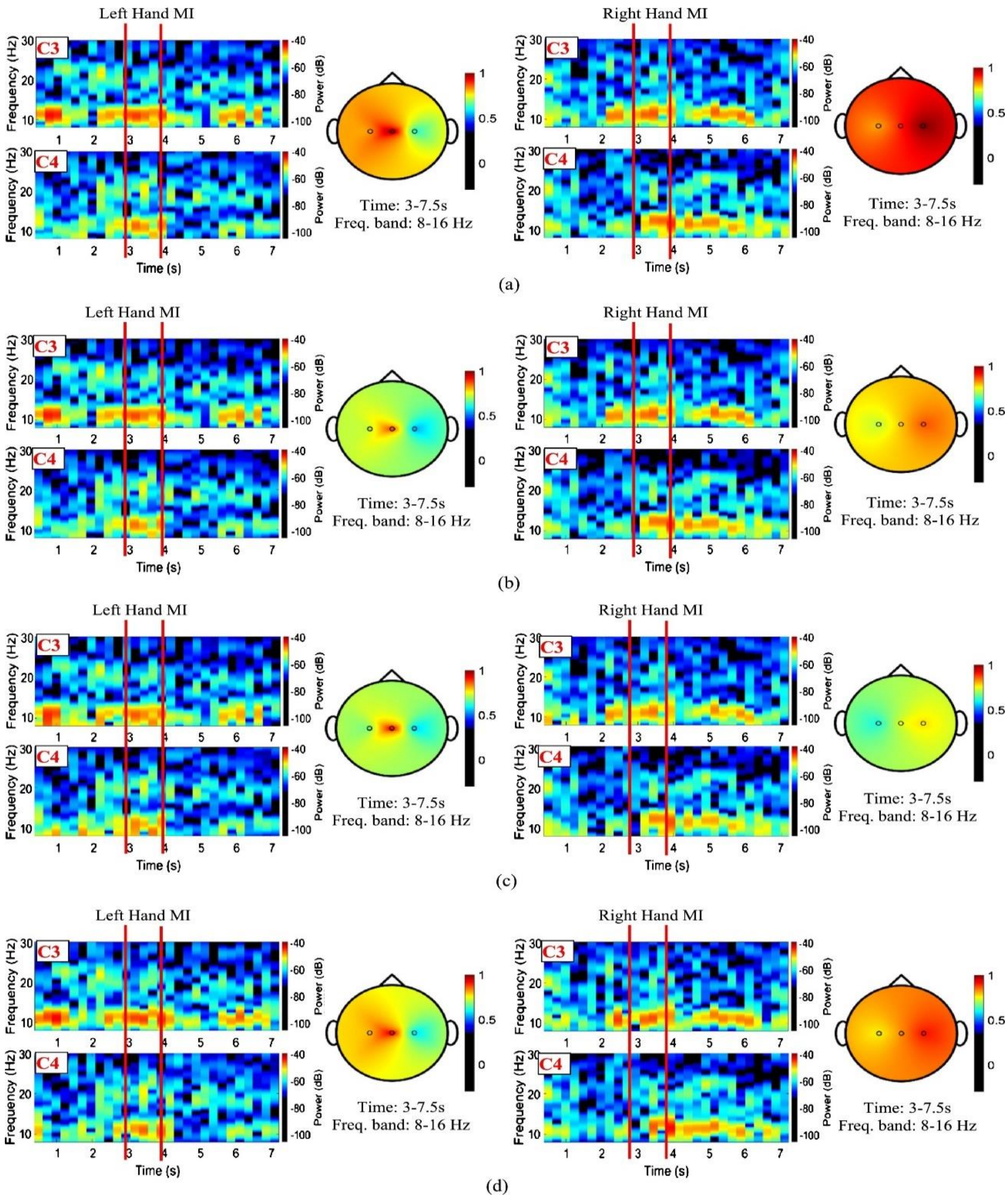

Figure 5.3 Time-frequency plots and averaged subband power topographic head plots estimated after filtering using (a) DTCWT based filter bank, (b) Butterworth filter, (c) Elliptical Filter, and (d) Chebyshev Type II filter. Filter of frequency band 8-32Hz filtered the EEG for time-frequency analysis. Whereas topographic head plots present averaged band power at 8-16Hz during MI task. Red lines indicate start and end of the cue at t=3s and t=4s, respectively. MI tasks started with the cue at t=3s.

5.5.2 Spectro-spatial feature optimization

In recent studies, features - extracted after the spectral and spatial filtering - are optimized by feature selection algorithms based on mutual information, fisher score, and L1-norm regularization to improve the MI task classification performance [178,248,249]. In [250], the researcher compared the mutual information, and fisher ratio-based feature selection methods applied on DFBCSP and showed the superiority of the mutual information over fisher score. In this work, we investigated the use of NCA with a regularization term to optimize the spectral-spatial features.

In most of the literature, for feature selection using mutual information, four best features are used, but in proposed NCA based method we selected features based on the features weight values compared to a threshold. The proposed algorithm evaluated the feature weights, as listed in Tables 5.4 and 5.5 for each subject from BCI Competition IV (Dataset 2b), and BCI competition III Dataset IIIa, respectively. Boldface values indicate the selected features. The proposed algorithm selected an average of 3.83 features (out of

Table 5.4 Feature weights estimated by the proposed algorithm for Dataset2. Boldface values indicate selected features. Threshold value for the selection of a feature is 5% of the maximum feature weight in each subject.

		Left-Hand vs Right-Hand			
Frequency band	**Feature index**	**B0103T**	**B0203T**	**B0303T**	**B0403T**
8-16 Hz	**1**	**3.08**	9.86e-05	9.91e-05	**3.7**
	2	**3.60**	**0.37**	0.00013	**3.57**
16-24 Hz	**3**	4.21e-05	0.00033	**0.1157**	**0.91**
	4	**2.60**	0.00043	7.919e-05	0.0004
24-32 Hz	**5**	5.37e-06	1.71e-05	0.0001	0.0001
	6	3.61e-05	0.00024	**2.072**	0.0005
	No. of features selected	**3**	**1**	**2**	**3**

Table 5.4 (Continued)

		Left-Hand vs Right-Hand				
Frequency band	**Feature index**	**B0503T**	**B0603T**	**B0703T**	**B0803T**	**B0903T**
8-16 Hz	**1**	**0.68**	**4.07**	4.71e-05	7.55e-05	5.19e-05
	2	9.77e-06	6.63e-06	**3.65**	**4.98**	0.0001
16-24 Hz	**3**	1.95e-05	2.60e-06	0.0061	8.65e-05	0.0004
	4	**4.52**	5.76	**0.72**	0.0002	**4.093**
24-32 Hz	**5**	3.37e-05	3.48e-05	9.58e-05	5.16e-05	0.0004
	6	**4.10**	**4.66**	0.0001	1.82e-05	0.0003
	No. of features selected	**3**	**2**	**2**	**1**	**1**

Table 5.5 Feature weights estimated by the proposed algorithm for Dataset1. Boldface values indicate selected features. Threshold value for the selection of a feature is 5% of the maximum feature weight in each subject.

		Left-Hand vs Right-Hand		
Frequency band	**Feature index**	**K3b**	**K6b**	**L1b**
8-16 Hz	**1**	**1.40**	1.40e-06	4.46e-05
	2	2.35e-05	3.84e-06	1.50e-05
	3	4.86e-05	**2.34**	**2.63**
	4	**3.38**	2.60e-06	2.93e-05
16-24 Hz	**5**	3.82e-05	**2.20**	3.08e-05
	6	**2.57**	5.66e-06	2.14e-05
	7	**4.45**	1.14e-05	**2.75**
	8	1.77e-06	4.01e-07	**2.40**
24-32 Hz	**9**	3.19e-06	1.31e-06	1.18e-05
	10	6.03e-05	2.68e-06	8.12e-06
	11	**2.70**	5.22e-06	1.78e-06
	12	1.20e-05	**1.65**	2.26e-05
	No. of features selected	**5**	**3**	**3**

Table 5.5 (Continued)

		Tongue vs Foot		
Frequency band	**Feature index**	**K3b**	**K6b**	**L1b**
8-16 Hz	**1**	2.75e-06	**2.07**	2.74e-05
	2	**2.45**	7.27e-06	**2.69**
	3	2.73e-06	**1.80**	2.47e-05
	4	1.35e-05	**2.00**	2.63e-05
16-24 Hz	**5**	**1.85**	**2.28**	6.29e-06
	6	1.87e-05	**2.56**	2.24e-05
	7	1.59e-06	1.55e-06	1.09e-05
	8	**1.39**	2.47e-06	**0.58**
24-32 Hz	**9**	0.55e-05	6.58e-07	2.83e-06
	10	1.39e-05	1.46e-05	1.45e-05
	11	1.53e-05	**1.78**	5.31e-05
	12	1.67e-05	0.035	**2.41**
	No. of features selected	**3**	**6**	**3**

twelve features) from BCI competition III Dataset IIIa and 2 features (out of six features) from BCI Competition IV (Dataset 2b), which are less than four best features that we choose in mutual information method (as suggested in literature). Therefore, the proposed algorithm is producing higher classification accuracy with lesser number of features. The reduced number of features increases the classification speed of the classifier and reduces the computational cost. To show the effectiveness of the selected features on the MI classification, in Figure 5.4, we presented all the twelve spatial filters of subject "k3b" from BCI competition III Dataset IIIa (left-hand vs right-hand) and six spatial filters of subjects “B0103T” and “B0403T” from BCI Competition IV (Dataset 2b). Spatial filters selected by the proposed algorithm are outlined by green colour. It can be observed that in most of the cases, for instance, spatial filters 4, 6, 7 and 11 of subject ‘K3b’ from BCI competition III Dataset IIIa, show significant SMRs over sensory motor cortex and these filters are successfully selected by our proposed approach. In few cases, for instance, spatial filter 3 of subject “B0603T” from BCI Competition IV (Dataset 2b) which shows significant SMR, is not selected by our approach. However, in this case also, the proposed algorithm selected other significant filters (1, 2, and 4). Hyperparameter, length of array of lambda values (L) (see algorithm 1), controls the feature selection performance of the proposed algorithm. In this approach, we chose this value (L=20, line spaced between 0 and 2) based on experiments. To further improve feature selection capability of the proposed method, hyperparameter optimization algorithms can be used to achieve an optimal value of L. However, this will increase the cost of computation. The L value we used is capable enough of extracting the most significant features in most of the cases.

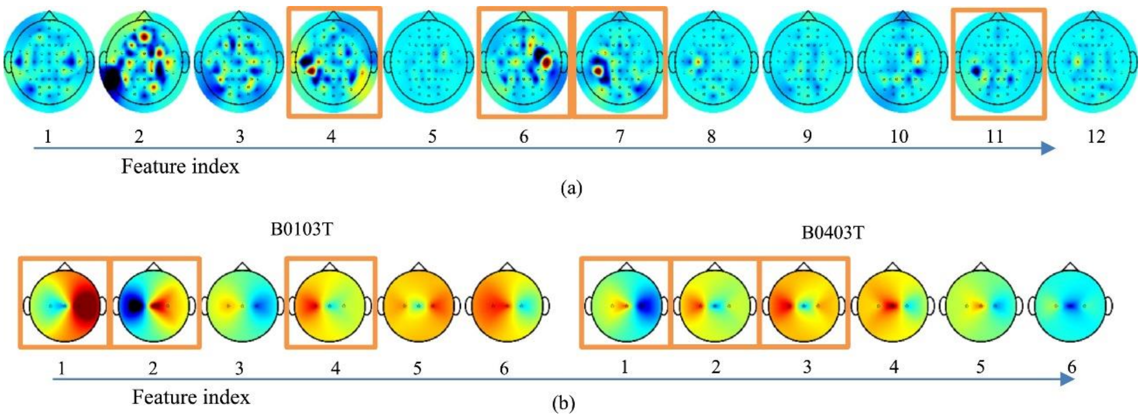

Figure 5.4 Spatial filters at feature indices for (a) Subject “K3b” of BCI competition III Dataset IIIa and (b) subjects “B0103T” and “B0603T” of BCI Competition IV (Dataset 2b). Corresponding feature weights estimated by the proposed method are listed in Table IV and V.

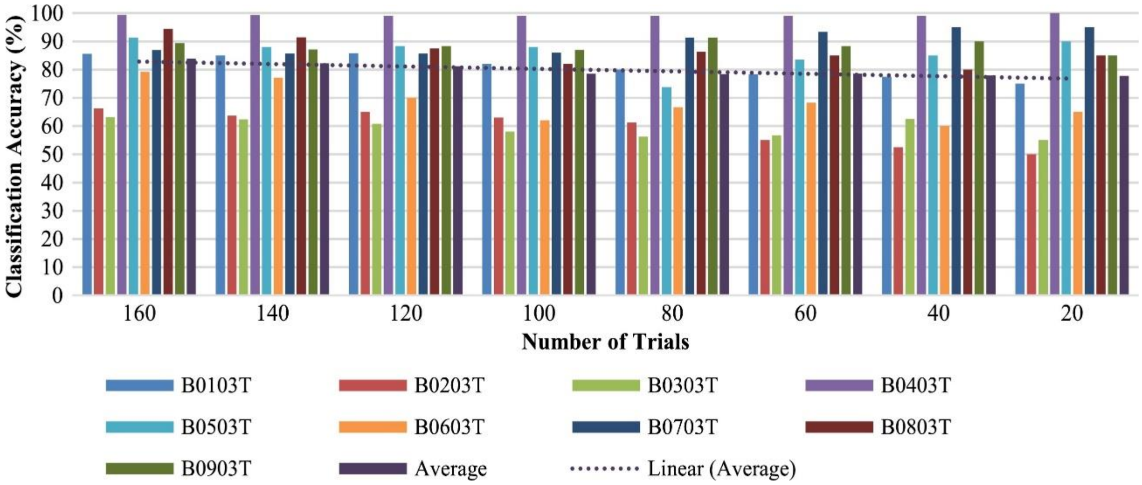

Figure 5.5. Classification accuracies obtained by each subject from BCI Competition IV (Dataset 2b) for varying number of trials. Note: the trend line shows the linear degradation in the average classification accuracy as the length of the trials decreases.

5.5.3 Training with different length of trial data

MI data recording from the subjects is a very time-consuming process due to the involvement of multiple training sessions. Each subject undergoes for long and tiring training sessions to collect the MI EEG data Sometimes it is not possible to collect a good amount of data. Hence, it is important to analyse the performance of MI classification algorithm on smaller dataset. In this work, we further investigated the effectiveness of the proposed algorithm by evaluating the classification accuracies achieved by the proposed method for varying length of trials. Figure 5.5 shows the classification accuracies achieved by each subject from BCI Competition IV (Dataset 2b) for different number of trials. The average classification accuracy for 20 trials is 77.78%, which is degraded by 6.24% from that achieved for 160 trials. This degradation shows that our algorithms is slightly affected by decreasing the number of trials.

5.5.4 Extension

Recent models presented in [169,251], added one more dimension to the structure of the MI data by using the sliding time windows. The sliding time window approach is performed to identify the beginning of the neural response to the MI task because the human brain responses differently in time for different subjects. Although, sliding time window improves the classification accuracy but adds more complexity and makes the solving model very time-consuming. However, our proposed optimization algorithm is only tested to optimize the spectral-spatial features, present algorithm can be further improved by also considering automatic selection of the best relevant time window. This will add more dimensions to the present model and require structural tensor data analysis. A few studies have suggested the use of structural tensor data decomposition and multiway learning to optimize a higher order dimensional problem [187,252,253]. Hence, EEG classification can further be benefited by 1) considering the temporal analysis in the presented algorithm, and 2) utilizing the supervised multiway learning approaches to optimize the MI structural data.

5.6 Conclusion

In this chapter, we proposed a novel algorithm, DTCWT-CSP (NCA), for the optimization of spectral-spatial MI features to enhance the classification performance of a two-class MI tasks. Specifically, time-frequency analysis using the designed DTCWT based filter provided more band power compared to traditional bandpass filters such as Butterworth, Elliptical, and Chebyshev. We then divided the MI EEG data into multiple sub bands and extracted spatial features using conventional CSP. Next, we applied a feature weighting algorithm based on neighbourhood component analysis under a supervised framework to optimize extracted spectral-spatial features. An SVM with linear kernel is trained using the optimized EEG feature set and perform the MI classification tasks. An experimental study is conducted using two public motor imagery datasets, BCI Competition IV (Dataset 2b), and BCI completion III Dataset IIIa, to validate the effectiveness of the proposed work in comparison with the various competing algorithms. The proposed method obtained superior classification accuracy and kappa coefficient compared to standard methods. Achieved results suggest that the proposed algorithm can be implemented for the design of an enhanced performing MI BCI device.