CHAPTER 6

6 Time window and frequency band optimization using regularized neighbourhood component analysis for Multi-View Motor Imagery EEG classification

6.1 Introduction

In recent years, advancements in medical and computational sciences have developed a communication pathway between the human brain and external devices; such methods are popularly referred as Brain-computer interfaces (BCIs) [1]. As a medical application, (BCI) devices are widely used to assist people with neuromuscular disorders. Among many types of BCIs [226,227,254,255], motor imagery (MI) based BCI uses brain signals associated with the imagination of motor movement-related tasks [9]. Various studies have suggested that when a subject thinks about a specific motor movement, there are significant relative power changes occur in the mu (8-13 Hz) and beta (13-30 Hz) rhythms of EEG acquired over the sensorimotor cortex area of the brain [8,14,229]. Subsequently, these power changes in EEG are processed and classified using pattern recognition methods to control external devices [27,28]. The power changes in EEG occur due to imagination of limb movements, are referred as event-related desynchronization (ERD)/ event-related synchronization (ERS), which can be further processed to control an external device [14]. However, EEG time series are highly contaminated by body motion artifacts and environmental noises, due to which distinguishing between different motor movements is a challenging exercise to perform [256]. Hence it is essential to employ preprocessing methods to suppress artifacts and noises before extracting useful information from the EEG signals [255].

In MI-BCI signal processing, EEG time latencies in motor imagery task differ subject to subject and thus require time window optimization methods. As discussed in section 2.6, with time windows, MI-EEG data become a higher order tensor. To optimize EEG tensor data, we propose a novel multi-task learning model to simultaneously optimize frequency bands and time windows and preserve the multi-view EEG data structure. In some of the recent BCI signal classification studies, learning methods based feature selection approaches such as regularized neighbourhood component analysis (RNCA) [240,257], L-1 norm regularization [28] have been proposed. In others, filter methods such as fisher score [258] and mutual information [259] are used for feature selection. However, filter-based methods are faster but achieve lower classification accuracies compared to learning-based feature selection methods. A few studies have compared various learning-based and filter-based feature selection methods and reported the superiority of learning-based feature selection methods in terms of classification accuracies [240,260]. RNCA is a fast-learning based feature selection method to optimize a single matrix feature vector [128]. In our proposed model, we formulized the RNCA objective function by adding a time constraint to optimize multi-view EEG data. Specifically, MI-related EEG data are segmented into multiple time windows, and then each of them is filtered using a dual-tree complex wavelet transform (DTCWT) decomposition and reconstruction approach. Afterward, CSP features are extracted, and a multi-view feature space is generated. Subsequently, optimal features are selected using the proposed method for MI classification.

The rest of the chapter is organized as follows: Section 6.2 explains the method and materials used in this work. Section 6.3 elucidate the experimental study conducted on the BCI datasets followed by results in section 6.4. Further, a discussion is presented in section 6.5. The last section gives the conclusion.

6.2 Methods

6.2.1 Preprocessing of EEG

The most important step in BCI system design while signal processing EEG time series is to remove the noises and artifacts, affecting the signal's vital information. The nature of EEG is known to be very random and non-stationary that makes it a complex signal to be analyzed [261]. Moreover, MI task-related EEG patterns are highly subject-specific and trials specific in the same subject. Therefore, it is important to perform time-frequency analysis on the EEG to eliminate artifacts, noises and extract useful features. Wavelets are considered one of the best tools to deal with biomedical signals in various studies [262–264]. We opted for Dual-tree complex wavelet transform (DTCWT) for decomposing the EEG signal in this work. As an advanced variant discrete wavelet transform (DWT), DTCWT overcomes the shortcomings of DWT, such as transition band power losses and aliasing. The structure of DTCWT comprises two DWTs connected in parallel to evaluate the real and imaginary part of the transform.

The presence of ERD/ERS patterns are prominent in mu (8-13Hz), and beta (13-30Hz) frequency ranges during a motor imagery task [14]. Many studies have proposed using frequency band optimization to increase the MI classification accuracy [68,257–259]. Basically, in these methods, EEG is filtered at multiple frequency bands between 4 and 30 Hz to capture discriminative features at different sub-bands. In our approach, we devise a DTCWT filter bank to filter the EEG into three sub-bands of frequency range 4-8 Hz, 8-16 Hz, and 16-32 Hz. In the experiment, the EEG is decomposed into DTCWT coefficients of four details and an approximation. Every detail or approximation has a different frequency band. For the first sub-band, we chose the coefficients with sub-band frequency between 4 and 8 Hz to reconstruct the EEG time series, whereas thresholding is applied on the other coefficients by replacing their values with zero. Similarly, we

filtered the EEG into two more sub-bands: 8-16 Hz and 16-32 Hz. An illustrative view of DTCWT filter banks used for the decomposition and reconstruction is shown in Figure 6.1. We eliminated the ocular artifacts using the method explained in [234].

6.2.2 Feature extraction from Multi-view EEG data

For the classification of motor imagery signals, pattern recognition methods are widely investigated in numerous studies [28,177,240]. The first and most important step in pattern recognition methods is feature extraction. In motor imagery classification, features are commonly extracted by spatial, spectral, and temporal analysis. CSP, as spatial analysis, is considered the most effective algorithm for MI classification. Before implementing the CSP, we first define the multi-view nature of motor imagery EEG. In general, the EEG dataset contains a two-dimensional matrix of temporal points and channels. When recording EEG for BCI applications, the time series are further marked by different trials. This naturally adds one more dimension to the dataset and makes BCI dataset a multi-view classification problem. Assume that the i^{th} trial EEG data is represented as $X_{i,C} \in \mathbb{R}^{M \times T}$ where C, M, and T represent the motor imagery class, total number of channels, and temporal points, respectively. Now, for total N trials, we have N number of M×T matrices. Before applying CSP, we assume that each of the matrices is mean subtracted and filtered at a suitable frequency band.

For a given two-class motor imagery classification problem, suppose X_A is the EEG matrix of A^{th} class averaged over all the A^{th} class trials. Similarly X_B is the EEG matrix of B^{th}class averaged over all the B^{th} class trials. CSP projection matrix is generated by first calculating the normalized covariance EEG matrices R_A and R_B given by

$$R_A = \frac{X_A X_A^T}{trace(X_A X_A^T)} \qquad R_B = \frac{X_B X_B^T}{trace(X_B X_B^T)} \tag{6.1}$$

Where X^T computes the transpose of X and $trace(B)$ adds all the diagonal elements in B. In the next step, eigenvector decomposition is performed on the addition of R_A, R_B as given by

$$R = R_A + R_B = U_0 \in U_0^T \tag{6.2}$$

Where U_0 and $\in$ represent eigenvector and diagonal eigenvalue matrix. Subsequently, two covariance matrices are calculated using the formulas given as

$$S_A = PR_AP^T \qquad S_B = PR_BP^T \tag{6.3}$$

Where P is termed as whitening transformation matrix and can be evaluated as $P = \in^{-1/2} U_0$. Moreover, the eigenvector matrix U of S_A and S_B will always be the same with the summation of their eigenvalues be unity. Finally, spatial filters are generated using $W = U^TP$. For ith trial EEG data matrix, $X_{i,C}$ can be transformed by the spatial filters using the following equation

$$Z = WX \tag{6.4}$$

The designed spatial filters have a property that the first and last m columns maximize the variance for one class and minimizes that for the second class. Thus, the final feature space has spatial features defined by

$$f_p = \log\left(\frac{(var(Z_p)}{\sum_{p=1}^{2m} var(z_p)}\right) \tag{6.5}$$

Where $var(.)$ is a function that solves for the variance and p is the row of Z_p ($p = 1, \ldots . 2m$).

Advance variants of CSP such as FBCSP, and DFBCSP have shown that the effectiveness of CSP gets further improved when EEG is filtered in multiple frequency bands before

the implementation of CSP. Therefore, we apply the CSP feature extraction method from Equation (6.1) on the EEG filtered at different frequency bands. In this work, we devise a filter bank to filter the EEG by first decomposing the EEG using dual-tree complex wavelet transform into details and approximation coefficients and then reconstructing the EEG using the details coefficients from specific frequency ranges. We used coefficients of three details for the reconstruction of the EEG. Thus, this procedure filters the EEG into three distinct frequency bands: ranges 4-8Hz, 8-15Hz, and 16-32 Hz. Now, the CSP features are extracted from the EEG at multiple frequency bands and feature space, FV is created as given by

$$FV = \begin{pmatrix} f_{1,1} & \cdots & f_{1,Q} \\ \vdots & \ddots & \vdots \\ f_{N,1} & \cdots & f_{N,Q} \end{pmatrix} \tag{6.6}$$

Where N is the total number of trials combining both the classes and Q is given by $2m \times J$, where J is the total number of filters in the filter bank.

Besides frequency band optimization, time window optimization is equally important because the brain response to a motor imagery task is different in different subjects and varies with frequency band and time. In an MI paradigm, the MI task is generally performed for 4s, where the starting period (from 0s to 1s) is mainly considered as the imagination preparation stage and the later period (from 3.5s to 4s) is considered as a post imagination period [183]. Hence, it is crucial to select a time window covering the significant ERD/ERS patterns and reject the irrelevant time interval. In studies [28,184,265],

EEG data is segmented into multiple time windows using sliding time window approach to select relevant time windows. Further, in sliding time window methods of EEG, generally, a width of 2s with 0.5s shift has given optimum results. Keeping this in view,

we segmented the EEG data into K time windows of length 2s each. Considering t=0s as the start of the motor imagery task and t=4s as the end of the motor imagery task, the first time window is from -0.5 s to 1.5 s, followed by five more time windows (0s to 2s, 0.5s to 2.5s……, 2s to 4s), where each is shifted by 0.5s. Hence, we have total six time windows ($K = 6$). Finally, using the above explained feature space extraction method given in Equation (6.6), we evaluated multi-view feature space $FV^{(t)}$ where t is the time window index. Figure 6.1 describes the complete multi-view feature extraction scheme.

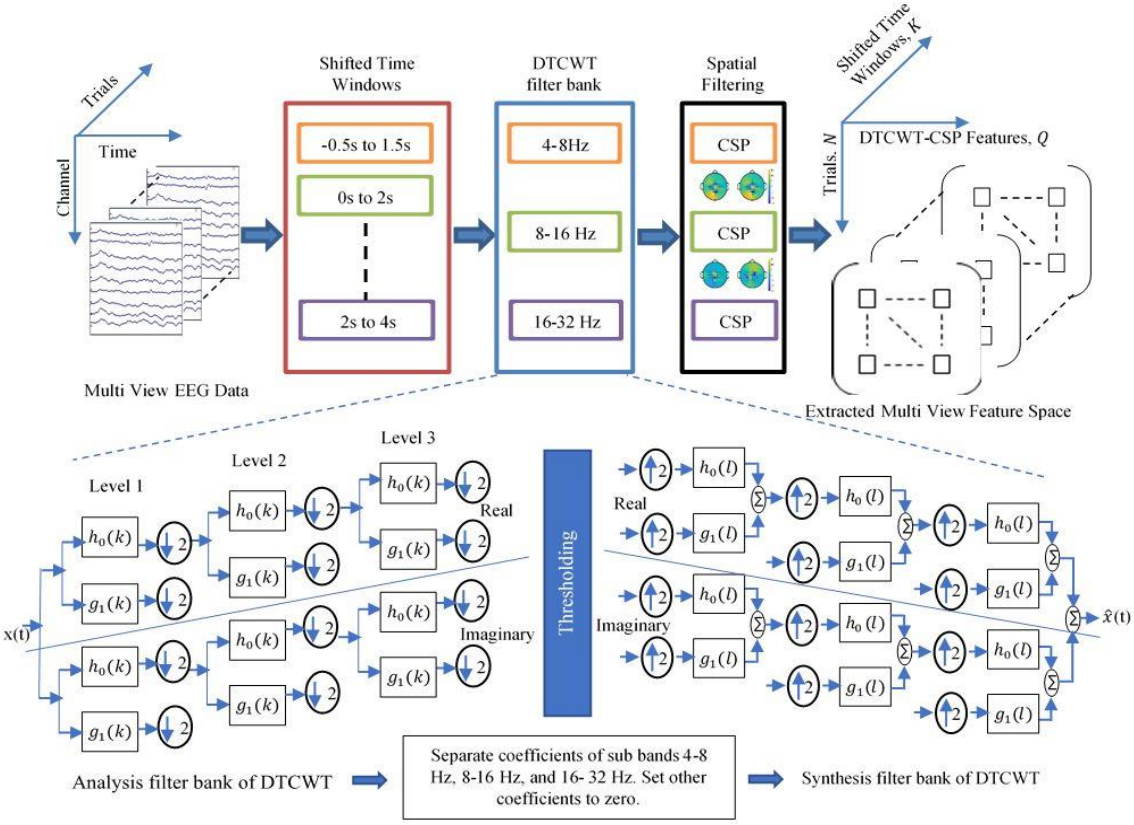


Figure 6.1 Illustration of multi-view feature extraction from MI-related EEG dataset for classification. Naturally, EEG data has three modes: channels, time, and trials. In the first stage, EEG time series of each trial is segmented into K multiple time windows. Next, each time window is filtered at three frequency bands using DTCWT filter bank. Afterward, CSP is executed on filtered EEG to evaluate MI-related features. The final feature space has K matrices. Columns of each such matrix have DTCWT CSP features, and rows represent trials.

6.2.3 Multi-view Feature selection approach: Regularized Neighbourhood component analysis

Regularized neighbourhood component analysis (RNCA) learns the mahalanobis distance with a regularizing parameter λ. RNCA estimates the weights of all the features present in feature space by optimizing an average leave one out classification accuracy objective function [128]. The optimization is done by evaluating the best regularization parameter using cross-validation on the train set. In this algorithm, a weight vector for feature vector is learned using nearest neighbour learning classifier. Let us consider a multiclass feature space, $FV = \{(x_i, y_i), i = 1,2 \dots, n\}$ from which optimal features need to be selected for classification. Where $x_i \in \mathbb{R}^{n \times r}$ are the feature vectors, and n and r representing total number of trials and extracted features, respectively. $y_i = \{1,2, \dots c\}$ represents class label, c is the total number of class. In RNCA, the mahalanobis distance between a sample x_i and a selected reference sample x_j is calculated. The selection of a reference sample x_j for a sample x_i has a probability P_{ij} , which is higher if the distance between the two samples is less. The formula for evaluating this distance D_w is given by

$$D_w(x_i, x_j) = \sum_{m=1}^{r} w_m^2 |x_{im} - x_{jm}| \tag{6.7}$$

Where the evaluated weight of mth feature is w_m. Further, an exponentially decaying kernel function establishes a relationship between the probability P_{ij} and the weighted distance D_w as given by

$$P_{ij} = \begin{cases} \dfrac{G\left(D_w(x_i, x_j)\right)}{\sum_{j=1, j \neq i}^{n} G\left(D_w(x_i, x_j)\right)} & for\ i \neq j \\ 0 & for\ i = j \end{cases} \tag{6.8}$$

Where G is the exponentially decaying kernel function given by $G(a) = exp(-\frac{a}{\sigma})$ and σ represents the width of the kernel. The probability that a sample x_i is correctly classified as the true class is defined by

$$P_i = \sum_{j=1, j\neq i}^{n} P_{ij} Y_{ij} \tag{6.9}$$

Where Y_{ij} is an indicator to show if a true class is predicted, it returns 1 for $y_i = y_j$ and 0 for $y_i \neq y_j$. From Equation (6.9), the average leave-one-out (LOO) probability of true classification is defined as

$$F(W) = \frac{1}{n}\sum_{i=1}^{n} P_i \tag{6.10}$$

In RNCA, a term regularized parameter, λ is added in Equation (6.9) to avoid overfitting of the model, and Equation (6.10) can be rewritten as

$$F(W) = \frac{1}{n}\sum_{i=1}^{n} P_i - \lambda \sum_{m=1}^{r} {W_m}^2 \tag{6.11}$$

Since $F(W)$ is the average LOO true classification probability, RNCA algorithm targets to maximize $F(W)$ for a given λ to evaluate feature weight vector W

$$\begin{aligned} \widehat{W} &= \underset{w}{\operatorname{argmax}}\, F(W) = \underset{w}{\operatorname{argmax}} \left(\frac{1}{n}\sum_{i=1}^{n} P_i - \lambda \sum_{m=1}^{r} {W_m}^2 \right) \\ &= \underset{w}{\operatorname{argmin}} \left(-\frac{1}{n}\sum_{i=1}^{n} P_i + \lambda \sum_{m=1}^{r} {W_m}^2 \right) \\ &= \underset{w}{\operatorname{argmin}} \left(-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} P_{ij} Y_{ij} + \lambda \sum_{m=1}^{r} {W_m}^2 \right) \end{aligned} \tag{6.12}$$

Note that a constant (p=1) added to the objective function does not affect the argument of minimum. From Equation (6.9), we have, $\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} P_{ij} = 1$

$$\begin{aligned}
\widehat{W} &= \underset{w}{\operatorname{argmin}} \left(1 - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} P_{ij} Y_{ij} + \lambda \sum_{m=1}^{r} {W_m}^2 \right) \\
&= \underset{w}{\operatorname{argmin}} \left(\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} P_{ij} - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} P_{ij} Y_{ij} + \lambda \sum_{m=1}^{r} {W_m}^2 \right) \\
&= \underset{w}{\operatorname{argmin}} \left(\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} P_{ij} \left(1 - Y_{ij}\right) + \lambda \sum_{m=1}^{r} {W_m}^2 \right) \\
&= \underset{w}{\operatorname{argmin}} \left(\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} P_j \, I(y_i \neq y_j) + \lambda \sum_{m=1}^{r} {W_m}^2 \right) \qquad (6.13)
\end{aligned}$$

Where condition-based indicator value, $I(.)$ returns one if y_i is not equal to y_j else it assigns zero. The argument of minimum will give the weights to each feature, present in the feature space FV. However, as can be seen from Equation (6.13), the derived objective function can assign weights to features of a single matrix feature space only. For the optimization of a multiple matrix feature space, $FV^{(t)}$ that has three dimensions, the objective function in Equation (6.13) is formulated as

$$\widehat{W^{(t)}} = \underset{w}{\operatorname{argmin}} \left(\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} {P_{ij}}^{(t)} I^{(t)}(k_i \neq t_i) + \lambda^{(t)} \sum_{m=1}^{r} \left({W_m}^{(t)} \right)^2 \right) \qquad (6.14)$$

Equation (6.14) is maximized by selecting the $\lambda (= \lambda_{best})$ at which the average leave one out classification accuracy is highest or generalization loss is lowest. The tuning of the λ is done by running a certain number of iterations. Further, $\widehat{W}$ is calculated using the conjugate gradient approach. Lastly, values of $\widehat{W}$ are the weights assigned by the objective function to the features. All the weights are thus compared with a threshold value to select optimal features. The threshold chosen in this work is 5% of the maximum evaluated weight. The flowchart shown in Figure 6.2 elucidates how the λ_{best} is chosen

to select optimal features by RNCA for a single time window EEG feature space. This is further explained in steps as below.

Step 1: Start the iteration and create a uniformly distributed array of λ values. In this work, 100 linearly spaced values are used, ranging between 0 and 2. Hence, the iteration size is also 100.

Step 2: Fit the λ value of the first iteration in Equation (6.13). Then, evaluate feature weights, select a subset of features by comparing with the threshold and calculate the generalization loss as given by

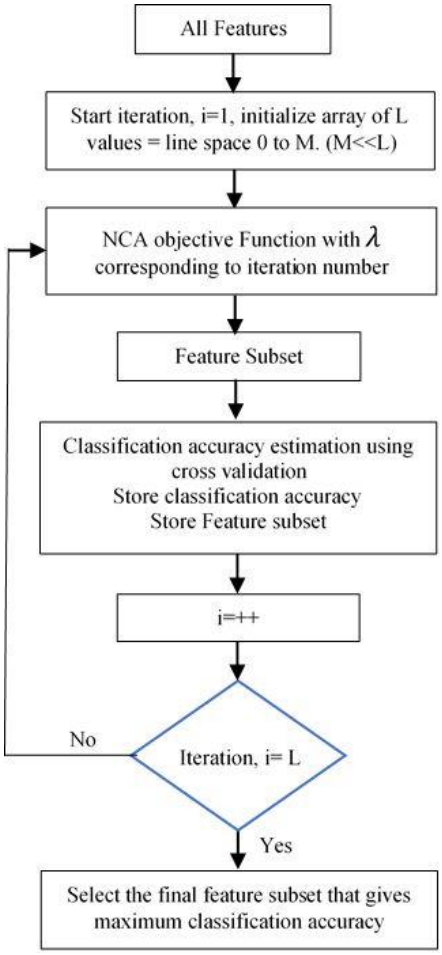


Figure 6.2 Flow chart presents the approach of tuning the regularization parameter within the NCA framework for optimal feature selection.

$$err = \frac{1}{n}\sum_{i=1}^{N} I(k_i \neq t_i) \tag{6.15}$$

Where $k_i$and t_i are predicted and true classes, respectively. The Indicator value I(.) returns one if k_i is not equal to t_i else it assigns zero. Next store err for each changing value of λ. Repeat the procedure until maximum iteration size is achieved.

Step 3: Plot err vs. λ values. Chose the best λ, λ_{best} at which err is minimum.

Step 4: Select the best feature subset using λ_{best} in Equation (6.13).

In multi-view data classification, preserving the structure of the data is a complex exercise to perform. The aforementioned multi-view features extracted from motor imagery related EEG must be optimized by a feature selection method so that the basic structure of the features does not get altered. Figure 6.3 presents a scheme to classify multi-view features. In this approach, frequency band CSP features of each time window are first normalized. Normalizing the data makes the frequency band CSP features from different time windows comparable because EEG characteristics (such as amplitude, phase, and frequency) changes with time. Afterward, among all the CSP extracted features at multiple frequency bands from each sliding time window, the best feature subset is evaluated using the proposed multi-view learning approach based on RNCA. In other words, we selected K feature subsets from the K shifted time windows by the proposed method. Subsequently, selected features are combined together to train a classifier. Once the classifier is trained on the training dataset using only the selected optimal features, only these selected features are extracted from the test dataset and fed into the classifier to perform the classification task. Unlike FBCSP and DFBCSP algorithms that concentrate solely on frequency band selection, the proposed method provides a more sophisticated approach to optimize robust CSP features at multiple frequency bands and time windows under a multi-view learning framework.

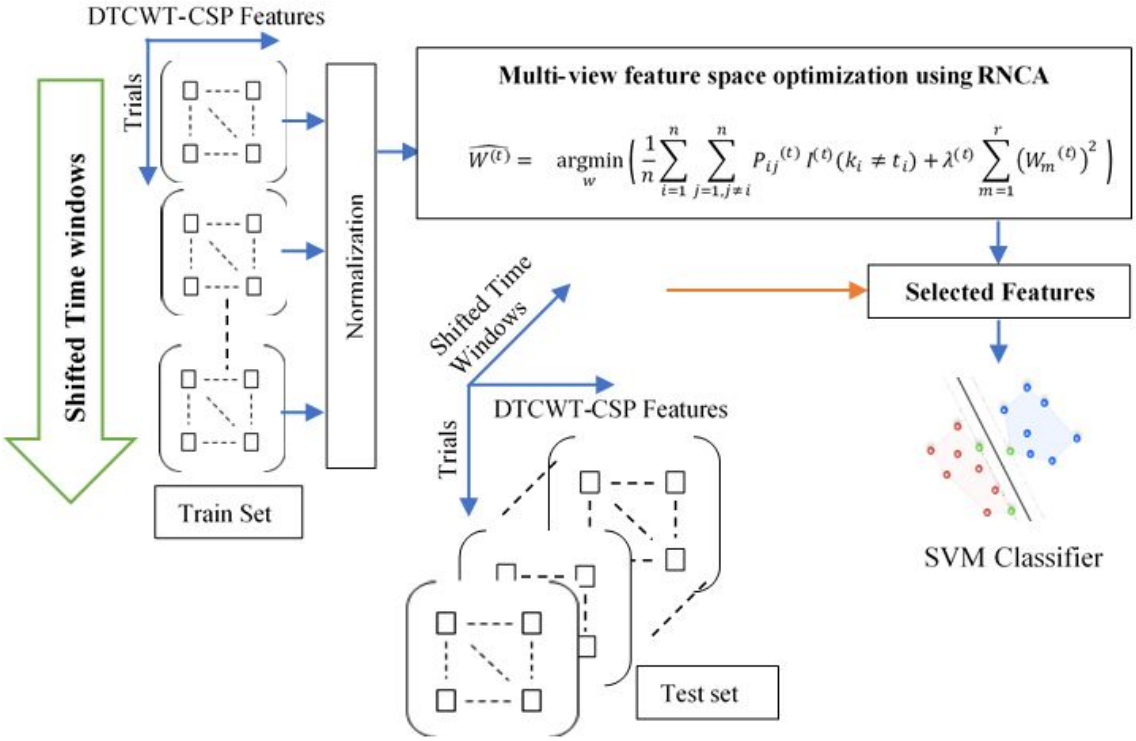


Figure 6.3 The workflow elucidates the proposed feature selection method to optimize time windows and frequency band CSP features of MI-related EEG. Optimization is performed under a structured-based multi-view learning environment where RNCA is applied on each of the matrices of feature space (extracted in Figure 6.1) for optimal feature selection. Subsequently, all the selected features are combinedly used to train an SVM classifier. During the test phase, similar multi-view features are extracted from the Test data and predict the MI class.

6.2.4 Classification

In this work, the two-class MI classification task is performed using a support vector machine (SVM). In SVM framework, a discriminant hyperplane globally maximizes the marginal distance between the two classes. In most BCI studies, SVM is chosen for classification because it is a relatively faster classifier and can effectively be used for a lesser number of trials [219].

6.3 Experimental study

6.3.1 Dataset description

The performance of the proposed algorithm is validated on three publicly available BCI datasets. Details of these datasets are as follows

A motor imagery EEG dataset popularly named BCI competition IV dataset 2a is the first dataset used in this work [68]. It consists of 22 channels EEG acquired from nine different subjects indulged in four class motor imagery tasks. The four tasks were thinking about the movement of either left hand, right hand, foot, or tongue following the presented cue on the screen. Further, it is bandpass filtered between 0.5 Hz and 100 Hz. It consists of 72 trials for each class from one subject.

The second dataset is BCI competition III dataset IIIa [266]. This dataset contains 60 channels of EEG recording of three subjects, namely "K3b", "K6b", and "L1b". The dataset was preprocessed by applying bandpass filtering between 1 and 50 Hz with a 60Hz power line noise eliminator, notch filter. The sampling frequency of the recording device was 250 Hz. The total number of trials were 180, 120, and 120 for subjects "K3b", "K6b", and "L1b" respectively. During each trial, cue-based motor imagery tasks were performed by the subjects for four classes, i.e., the imagination of movement of either left hand, right hand, foot, or tongue.

BCI competition IV dataset 2b is utilized as the third dataset for this study [16]. EEG from three channel locations C3, Cz, and C4 was recorded in this dataset while nine subjects performed cue-based two-class motor imagery tasks. The two tasks were to imagine either left hand or right-hand movement. The recorded EEG was sampled at 250 Hz and filtered between 0.5 Hz and 100 Hz.

6.3.2 Performance evaluation

The proposed method's performance is measured by conducting a comparative study among following standard feature extraction algorithms and the proposed method.

- **CSP:** The EEG time series data are first filtered using a chebyshev type II filter between 4 Hz and 40 Hz, and then CSP features are extracted from a single time window, 0.5s to 3.5s [267]. CSP algorithm generally uses a fixed frequency range and time window, due to which suboptimal classification accuracy is achieved [268]**.**
- **CSP with shifted time windows (CSP_{STW}):** In this method, CSP features are extracted from multiple overlapped time windows of 2s each, shifted by 0.5s. The first time window starts 0.5s before the beginning of the cue. The number of time windows used is six. The EEG is pre-filtered in the frequency band 4-40 Hz. All the extracted features are used together for classification.
- **FBCSP:** We used experimental settings given in [269]**.** Multiple frequency band filters are applied to the EEG data. The frequency ranges of filters are 4-10Hz, 10-16Hz, 16-22Hz, 22-28Hz, 28-34 Hz, and 34-40Hz. After filtering, features are extracted using CSP. A single time window of 3s after 0.5s from the beginning of the cue is taken. Subsequently, optimal features are selected for classification using a mutual information-based feature selection algorithm.
- **DFBCSP:** As in [164], a filter bank with sixteen overlapped filters, each having a passband of 4 Hz and shifted by 2 Hz, is designed. The overall range of the filter bank is 4-40Hz. Subsequently, from each trial, 3s recording during the motor imagery task is taken, and the overlapped filters are applied. Then, CSP features are extracted from each of the frequency band filtered EEG. Among all the extracted features, the best feature subset is selected by fisher score.

Next, we applied a shifted time window approach to analyze the effect of time window selection on existing conventional methods such as FBCSP and DFBCSP.

- **FBCSP with shifted time windows ($FBCSP_{STW}$):** FBCSP feature extraction method is applied on the EEG of multiple overlapped time windows. Six time windows are extracted in a similar way as in the CSP_{STW} method. Subsequently, we extracted the FBCSP features from each of the time windows. Next, we combined all the features extracted from multiple time windows at multiple frequency bands to create a two-dimensional feature space. Finally, the best feature subset is estimated by a mutual information-based feature selection algorithm.
- **DFBCSP with shifted time windows ($DFBCSP_{STW}$):** Similar to CSP_{STW}, multiple time windows are extracted from the raw EEG data. Then, DFBCSP features are extracted from each of the time windows. All the features extracted from multi time windows at multi frequency bands are combined to create a feature space in the next step. Further, a subset of best features is chosen for classification as estimated by a fisher score-based feature selection approach.
- **Proposed method:** EEG trials are segmented at multiple time windows and frequency bands, and then CSP features are extracted from each segment. The best subset of features is selected by the RNCA feature selection method for classification from all the extracted features.

6.4 Results

This section presents the classification performance of the proposed feature extraction and selection method in comparison with standard feature extraction method such as CSP, CSP_{STW}, FBCSP, $FBCSP_{STW}$, DFBCSP, and $DFBCSP_{STW}$. For all the algorithms, an SVM

classifier is used to perform the classification tasks. The software module used for the development of all the algorithms was MATLAB 2019b installed on a computer with 16 GB of RAM and Intel Core i7 (@ 3.4 GHz) processor. All the algorithms were applied to the three publicly available BCI datasets, which are discussed in the previous section.

The most commonly used evaluation criterion to measure the performance of BCI algorithms is classification accuracy. It is calculated from the confusion matrix as given by

$$classification\ accuracy = p_0 = \frac{\sum_{i=1}^{C} a_{ii}}{\sum_{i=1}^{C}\sum_{j=1}^{C} a_{ij}} \tag{6.16}$$

Where, a_{ij} is an element of the confusion matrix with i and j representing different classes. C is the total number of classes. Hence, the classification accuracy is the ratio of true predicted classes to the total number of trials.

Classification results are presented in Tables 6.1-6.4, where a comparison of classification accuracies (in %) estimated by the CSP, CSP_{STW}, FBCSP, $\text{FBCSP}_{\text{STW}}$, DFBCSP, and $\text{DFBCSP}_{\text{STW}}$, and the proposed algorithm is done for three BCI datasets (BCI competition IV dataset 2a, BCI Competition IV dataset 2a, and BCI competition III dataset IIIa). Table 6.1 lists the classification accuracies (in %) achieved for BCI competition IV dataset 2a to distinguish between two class motor imagery trials such as left hand vs. right hand, left hand vs. foot, and right-hand vs. foot. Similarly, Table 6.2 presents the classification accuracies for BCI competition III dataset IIIa. Whereas Table 6.3 elucidates the results obtained on BCI Competition IV dataset 2a for only left-hand vs. right-hand motor imagery classes. The results show that time window optimization to the spectrum optimized CSP features improve the MI signal classification performance. More importantly, the proposed algorithm outperforms all the competing algorithms.

Table 6.1 Classification accuracies (CA) (%) achieved using CSP, CSP_{STW}, FBCSP, $FBCSP_{STW}$, DFBCSP, and $DFBCSP_{STW}$, and the proposed algorithm on BCI competition IV dataset 2a (between two classes Left Hand vs. Right Hand, Left Hand Vs. Foot, and Right Hand vs. Foot). SVM classifier is learned and CA is calculated using 10-fold Crossvalidation. The highest CA obtained is marked in boldface for each subject. In addition, p-values are calculated using paired t-test between the proposed method and each of competing methods.

Left Hand Vs. Right Hand							
	CSP	$\boldsymbol{CSP_{STW}}$	**FBCSP**	$\boldsymbol{FBCSP_{STW}}$	**DFBCSP**	$\boldsymbol{DFBCSP_{STW}}$	**Proposed algorithm**
A01T	71.1	71.5	69.6	74.9	73.3	79.1	**85.6**
A02T	63.9	59.6	66.2	**67.2**	61.6	64.9	66.7
A03T	89.4	91.8	91.1	92.2	90.1	91.8	**97.2**
A04T	61.1	66.7	65.1	64.2	68.9	70.7	**77.1**
A05T	61.8	60.4	62.6	66.3	65.8	68.6	**82.6**
A06T	70.1	**73.1**	66.3	68.1	64.6	63.2	69.4
A07T	63.9	70.8	69.3	71.3	70.5	76.8	**79.1**
A08T	92.8	**95.8**	89.4	87	90.6	92.8	95.4
A09T	74.3	84.7	81.3	80.2	82.4	**86.2**	86.1
Mean	72.04	74.93	73.43	74.6	74.2	77.12	**82.13**
p-value	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	-
Left Hand Vs. Foot							
	CSP	$\boldsymbol{CSP_{STW}}$	**FBCSP**	$\boldsymbol{FBCSP_{STW}}$	**DFBCSP**	$\boldsymbol{DFBCSP_{STW}}$	**Proposed algorithm**
A01T	79.2	85.4	68.3	90.3	71.5	87.5	**92.4**
A02T	75.7	91.7	69.6	78.9	69.0	91.3	**96.5**
A03T	92.7	91.7	83.6	91.2	86.7	86.8	**93.7**
A04T	63.2	64.6	59.3	66.1	61.1	70.1	**72.9**
A05T	61.8	63.2	62.1	67.9	67.8	71.3	**77.1**
A06T	75.0	71.5	66.6	74.3	69.3	73.6	**80.5**
A07T	89.6	93.1	78.0	81.9	83.3	88.2	**96.5**
A08T	75.7	**84.7**	73.2	79.1	75.6	77.1	80.6
A09T	79.2	86.1	61.1	84.1	70.1	84.8	**87.5**
Mean	76.9	81.3	69.1	79.3	72.7	81.2	**86.4**
p-value	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	-
Right Hand Vs. Foot							
	CSP	$\boldsymbol{CSP_{STW}}$	**FBCSP**	$\boldsymbol{FBCSP_{STW}}$	**DFBCSP**	$\boldsymbol{DFBCSP_{STW}}$	**Proposed algorithm**
A01	93.1	94.3	86.1	86.8	89.6	95.8	**95.1**
A02	87.5	93.2	70.2	78.8	67.1	80.1	**91.7**
A03	91.0	91.3	84.1	91.1	89.9	82.9	**95.1**
A04	51.0	66.0	56.3	**74.1**	64.1	68.1	71.6
A05	61.8	63.2	61.1	77.1	66.9	78.1	**81.6**
A06	70.1	65.3	60.8	71.2	65.4	68.1	**74.3**
A07	72.9	88.9	76.3	80.1	79.1	85.4	**97.2**
A08	59.7	77.8	62.4	76.1	63.7	77.1	**85.4**
A09	53.5	59.0	68.6	72.1	65.8	73.2	**79.2**
Mean	71.2	77.7	69.5	78.6	72.4	78.7	**85.7**
p-value	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	-

Table 6.2 Classification accuracies (CA) (%) achieved using CSP, CSP_{STW}, FBCSP, $FBCSP_{STW}$, DFBCSP, and $DFBCSP_{STW}$, and the proposed algorithm on BCI competition III dataset IIIa. SVM classifier is learned and CA is calculated using 10-fold Crossvalidation. The highest CA obtained is marked in boldface for each subject. In addition, p-values are calculated using paired t-test between the proposed method and each of competing methods.

	CSP	$\boldsymbol{CSP_{STW}}$	**FBCSP**	$\boldsymbol{FBCSP_{STW}}$	**DFBCSP**	$\boldsymbol{DFBCSP_{STW}}$	**Proposed algorithm**
Left hand vs Right hand							
K3b	90.0	93.2	96.0	95.8	96.7	96.4	**98.6**
K6b	75.0	73.3	65.8	75.0	68.3	81.1	**89.3**
L1b	90.0	83.1	**94.2**	88.3	84.7	91.6	93.3
Left hand vs foot							
K3b	96.7	96.8	91.2	96.2	93.3	**97.8**	95.6
K6b	73.3	77.1	76.7	81.7	71.7	78.4	**83.3**
L1b	76.7	75.2	78.3	85.0	81.7	86.1	**86.7**
Right hand vs foot							
K3b	98.9	95.8	95.6	98.1	95.6	98.9	**99.3**
K6b	70.0	80.7	71.1	77.6	**83.3**	79.1	81.7
L1b	68.3	87.1	89.1	80.0	90.0	95.1	**97.7**
Mean	82.1	84.7	84.2	86.4	85.0	89.4	**91.7**
p-value	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	-

Table 6.3 Classification accuracies (CA) (%) achieved using CSP, CSP_{STW}, FBCSP, $FBCSP_{STW}$, DFBCSP, and $DFBCSP_{STW}$, and the proposed algorithm on BCI Competition IV dataset 2a. SVM classifier is learned and CA is calculated using 10-fold Crossvalidation. The highest CA obtained is marked in boldface for each subject. In addition, p-values are calculated using paired t-test between the proposed method and each of competing methods.

	CSP	$\boldsymbol{CSP_{STW}}$	**FBCSP**	$\boldsymbol{FBCSP_{STW}}$	**DFBCSP**	$\boldsymbol{DFBCSP_{STW}}$	**Proposed algorithm**
B0103T	70.6	70.1	76.3	79.2	80.3	83.1	**88.1**
B0203T	62.5	**62.8**	56.6	58.7	56.1	57.1	61.9
B0303T	63.4	**65.3**	64.4	59.1	59.3	64.5	60.5
B0403T	94.6	95.4	91.9	98.7	98.8	98.2	**99.4**
B0503T	77.5	75.0	78.9	86.2	86.3	87.4	**95.0**
B0603T	74.0	76.3	75.6	**83.7**	73.1	73.6	83.5
B0703T	80.0	82.7	84.4	85.0	85.0	83.7	**88.8**
B0803T	86.3	89.0	88.3	**94.5**	93.1	92.5	93.5
B0903T	85.0	83.1	87.2	86.2	86.3	81.3	**89.4**
Mean	77.1	77.7	78.2	81.3	79.8	80.2	**84.5**
p-value	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	*p<0.01*	-

In addition, we used paired t-test ($p<0.01$) with a Bonferroni correction to evaluate p-values between the proposed method and each of the competing methods to investigate the statistical significance difference of classification accuracy. The null hypothesis assumes that all the competing methods have identical performance. The results shown in Table 6.1-6.3 indicate that the performance of the proposed method is statistically significant than the competing methods ($p<0.01$).

6.5 Discussion

6.5.1 Selected spatial patterns at multiple frequency bands and time windows

Feature extraction is the most crucial step in the design of the BCI system software module. Among different features, spatial features are widely investigated in numerous studies for MI signal classification [270–272]. The commonly used method for spatial feature extraction is CSP. While many extensions to conventional CSP have incorporated the importance of filtering the EEG into several frequency bands but only a few have shown the effect of multiple shifted time windows on the MI classification performance. Since the brain's response time to motor imagery tasks is unknown and varies subject to subject, the importance of optimizing the time windows increases.

To explain the proposed algorithm's effectiveness in the frequency band and time window optimization, we created topoplots of spatial filters.

Performance of the proposed method for time window optimization: To present time window optimization performance, we made topoplots of spatial filters from BCI competition IV dataset 2a for subject "A05T" as an example (see Figure 6.4). Figure 6.4 (a) elucidates the feature space learned by the proposed method for subject A05T using a pictorial representation. Overall, feature space has six time windows where for each time window, we have three frequency bands, each of which contains 4 CSP features. The

proposed method's selected features are marked by blue color and rejected by no color. Notice that the selected features are from different time windows at different frequency bands. Since we set CSP parameter, m is equal to 2; there are four (2m) features in each frequency band. Also, in Figure 6.4(a), features from CSP feature index 11 at all the time windows are marked by an orange outline. The corresponding feature weights calculated by the proposed algorithm of the orange marked features are shown in Figure 6.4 (b). It can be noted that features of CSP index 11 in time windows -0.5s to 1.5s, 0.5s to 2.5s, and 2s to 4s are selected by the proposed method. To further analyze the proposed algorithm's working in selecting only the optimal features, the topoplots of the estimated spatial filters from third frequency band (CSP feature index 11) at all the six time windows are shown in Figure 6.4(c). An evident change in ERD/ERS patterns in the sensory-motor cortex area is observed as the time window changes, which shows that the neural response during motor imagery tasks changes with the time window. In time windows -0.5s to 1.5s, 0.5s to 2.5s, and 2s to 4s spatial features are significant over sensory-motor cortex area and shows the presence of ERD/ERS. In contrast, in other time windows, they are not so significant and hence rejected by the proposed algorithm. Hence, the proposed algorithm is selecting time windows with the most discriminatory features to classify motor imagery tasks. Another important point to mention here is that the feature weight assigned by the proposed method for the time window 2s to 4s is highest (Figure 6.4(b)). This is verified from topoplots of Figure 6.4(c), it can be seen that the ERD/ERS patterns in sensorimotor areas are more significant in time window 2 to 4s than other time windows. However, the spatial filter of time window 0.5s to 2.5s also shows significant ERD/ERS patterns, but interference from other unrelated brain regions is also

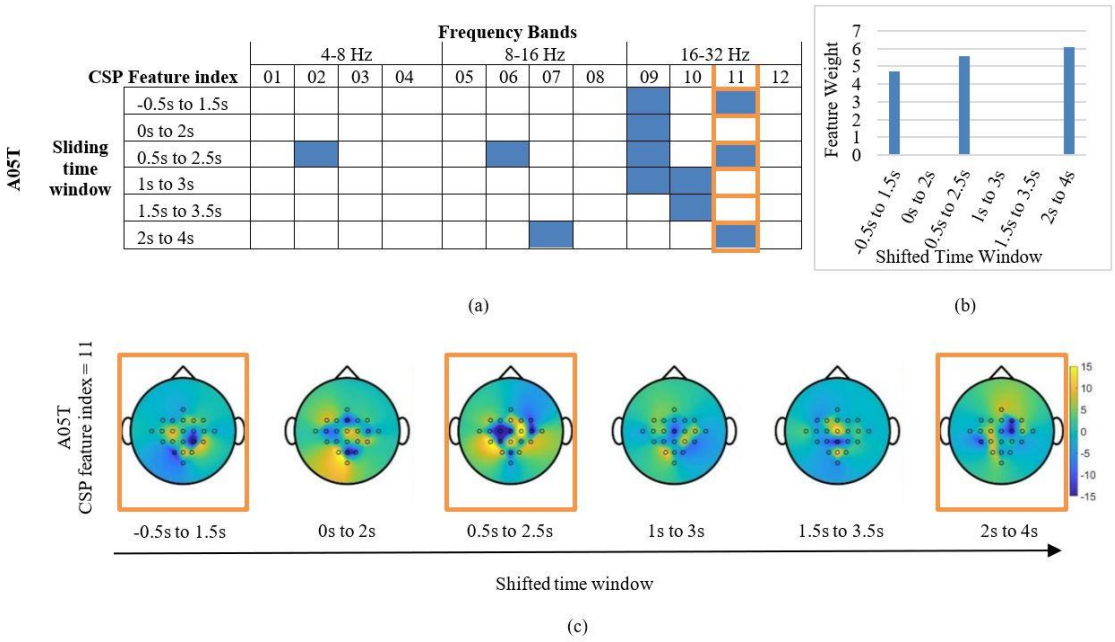


Figure 6.4 Performance of the proposed method for time window optimization: a), Pictorial representation of the feature space learned by the proposed method for subject A05T. In each time window, we have three frequency bands, each of which contains 4 CSP features. Blue boxes show the selected features at different time windows and frequency bands. b) Presents the feature weights assigned to all the six time windows at CSP feature index 11 (i.e., features marked by orange outline in 4.a). c) spatial filters all time windows (i.e., features marked by orange outline in 4.a). From spatial filters, it can be observed that the presence of ERD/ERS in sensorimotor cortex is strong at time windows -0.5s to 1.5s, 0.5s to 2.5s, and 2s to 4s, and that's why the proposed algorithm has chosen CSP features of these time windows.

present; that's why the proposed algorithm assigns a lower feature weight to time window 0.5s to 2.5s than that to time window 2s to 4s.

Performance of the proposed method for frequency band optimization: We investigated the frequency band optimization ability of the proposed algorithm on subject "k6b" of BCI competition III dataset IIIa as shown in Figure 6.5. Figure 6.5 (a) presents how the proposed method has learned features at different time windows and frequency bands for subject k6b. The selected features are shown using blue color and rejected using no color. CSP feature indices from 1 to 6 at time window 0.5s to 2.5s as marked by red outline in Figure 6.5(a) are selected to create topoplots of spatial filters in Figure 6.5(c).

Feature weights estimated by the proposed method for CSP indices 1 to 6 are shown in Figure 6.5(b). It can be observed from Figure. 6.5(c) that the spatial features are inconsistent with changing frequency bands. The proposed algorithm has selected the CSP feature indices 3, 4, and 6 (marked by orange outline in Figure 6.5(c)). To verify that the proposed method's selected features are the most significant, notice in Figure 6.5(c) that the ERD/ERS patterns in the sensorimotor area are more significant in CSP feature indices 3, 4, and 6 than in other CSP indices.

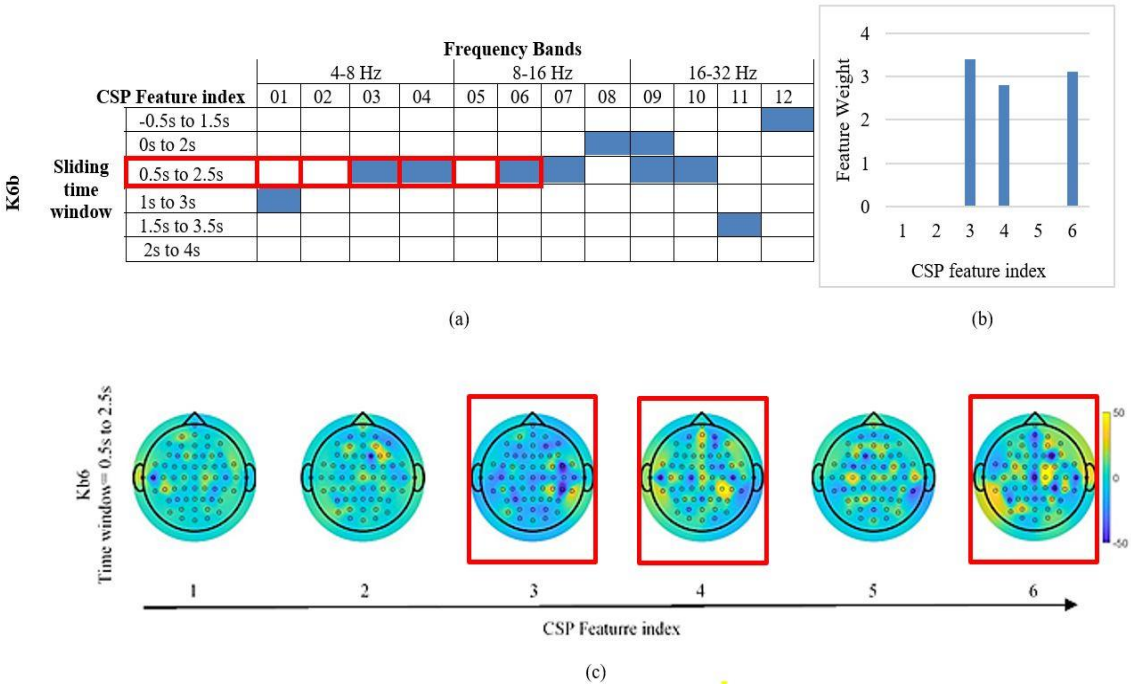


Figure 6.5 Performance of the proposed method for frequency band optimization: a), Pictorial representation of the feature space learned by the proposed method for subject k6b. In each time window, we have three frequency bands, each of which contains 4 CSP features. Blue boxes show the selected features at different time windows and frequency bands for subject k6b. b) Presents the feature weights assigned to six frequency band csp features at time window 0.5s to 2.5s (i.e. features marked by red outline in 5.a). c) spatial filters of six frequency band csp features at time window 0.5s to 2.5s (i.e. features marked by red outline in 5.a). From spatial filters, it can be observed that the presence of ERD/ERS in sensorimotor cortex is strong for CSP feature index 3,4, and 6 and since the proposed algorithm selected only these features proves its robustness in optimizing frequency bands.

6.5.2 Feature dimensionality reduction

The proposed feature selection method assigns a weight to each of the extracted features and then select or reject it by comparing its weight with a threshold. In the proposed work, the dimension D of the final feature space is given by

$$D = 2 * m * J * K \tag{6.17}$$

Where m, J, and K represents CSP parameter, number of filters in the designed DTCWT filter bank, and number of shifted time windows, respectively. In our approach, four CSP features are extracted from each of the three frequency bands, and six time windows are used. Hence the dimensionality of the generated feature space is seventy-two, which is reduced by the proposed method. In this section effectiveness of the proposed method in dimensionality reduction is assessed using the three MI datasets presented in this work. Figure 6.6 shows the average number of features selected by the proposed method. The average is calculated for all the subjects from three BCI datasets for different motor imagery tasks. It can be observed that the dimension of the feature space is significantly reduced by the proposed method. Also, the number of features selected by the proposed

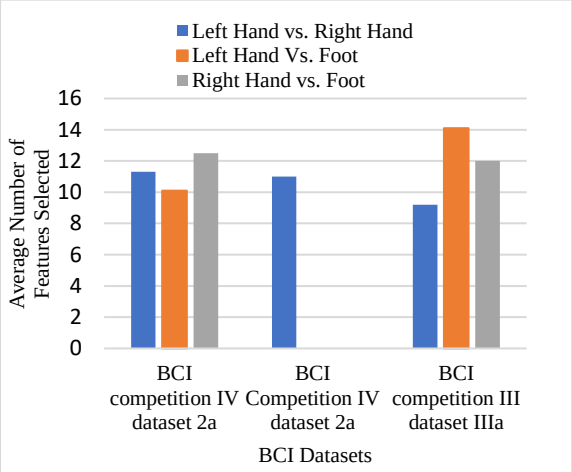
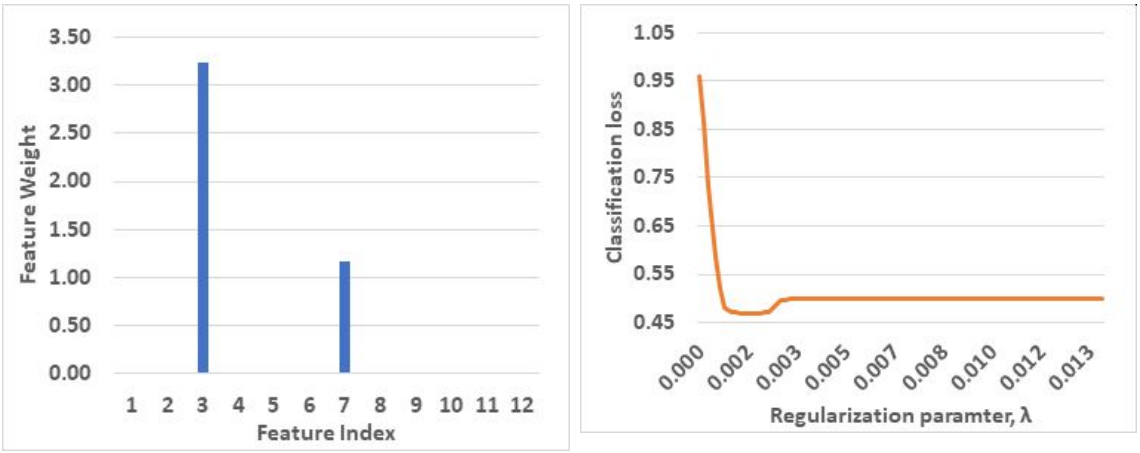


Figure 6.6 Average number of features selected by the RNCA for three BCI datasets for different motor imagery tasks classification.



(a) *(b)*

Figure 6.7 Feature optimization at a particular time window for subject k3b from dataset 2 (right vs. left-hand MI task) (a) Estimation of the regularization parameter $\lambda=\lambda_{\text{best}}$ at minimum loss value. (b) Feature weights calculated using λ_{best}. Although these graphs show feature selection in single time window, with similar tuning and weight assigning technique CSP features at different time windows are optimized.

method are less compare to other aforementioned time window optimization methods such as CSP_{STW}, $FBCSP_{STW}$, and $DFBCSP_{STW}$. Since in these methods, we selected four frequency band optimized CSP features from each time window. Hence, the dimension of the feature space used in these methods is twenty-four.

To explain how the algorithm is tuning the regularization parameter for optimal feature selection in a single time window, we used EEG data of subject "A02T" from BCI competition IV dataset 2a for left-hand vs. right-hand classification. In Figure 6.7(a), the λ_{best} at minimum classification loss, is chosen to solve the RNCA objective function for feature subset selection from a particular time window. Figure 6.7(b) shows the weights estimated by the RNCA. The criteria for selecting a particular feature are set by a threshold, taken as 5 % of the maximum feature weight. Similarly, we applied RNCA on

each time window, and then all the selected features are fed into a linear kernel SVM for classification.

6.5.3 Performance with different trial length

Particular motor imagery task-oriented training of the subjects is usu ally exhausting and time-consuming [273]. During the experiment, the subjects undergo mantle and physical fatigue such as neck pain due to sitting on an armchair for longer hours, eye stress by looking at the cue-displaying screen, and mantle stress. Hence, it is impossible to perform enough trials in few cases, especially when recording from a paralyzed subject for rehabilitation. Therefore, the performing algorithm efficiency mustn't vary with the number of trials. To investigate the effect of using a lesser number of trials on the performance of the proposed algorithm, we varied the trial length of the five subjects "B0203T", "B0403T", "B0603T", "B0803T", and "B0903T" of BCI Competition IV dataset 2a and evaluated classification accuracies (see Figure 6.8). It can be observed that the classification accuracies achieved by the proposed method are degrading as the length of the trials is reducing. This limits the performance of the proposed algorithm to some

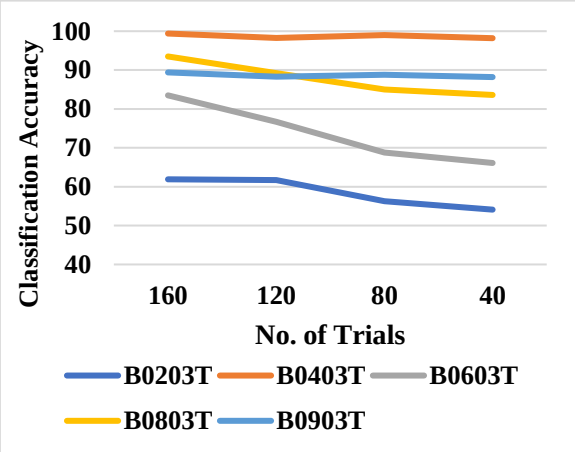


Figure 6.8 Change in classification accuracies according to the varying length of trials.

extent. However, the obtained classification accuracies do not degrade much for four subjects other than B0603T.

Further, we tuned the λ using a fixed length (L=100) uniformly distributed array between 0 and 2. In this work, we opted L based on experiment. This array length L should be optimized according to the changing number of trials to further enhance the classification accuracy. As an extension to the present study, hyperparameter optimization can select the best L, but that will also increase the computational cost.

6.5.4 Limitations and Future scope

Figure 6.9 shows the average computational time consumed during the optimization of CSP features by the proposed algorithm and other competing methods. It can be noted that the computational time of the proposed method is more than that of the CSP, CSP_{STW}, FBCSP and DFBCSP. The computational time taken by these algorithms is less because these algorithms can optimize either frequency bands (FBCSP and DFBCSP) or time windows (CSP_{STW}) only, but our algorithm is optimizing both frequency bands and time windows simultaneously. Other competing algorithms such as $DFBCSP_{STW}$ and $FBCSP_{STW}$ are optimizing frequency bands and time windows but are consuming more computational time than the proposed algorithm. Further, this time is the training run time for optimal feature selection, which is a one-time process and does not affect the test time or real-time classification. Hence, the proposed algorithm enhanced the motor imagery classification performance without degrading the computation efficiency for any BCI based system.

Further, we explored the effect of time window optimization in conventional methods such as FBCSP and DFBCSP by applying shifted time window approach. For feature creation in $DFBCSP_{STW}$ and $FBCSP_{STW}$, we unfolded the time window segmented data

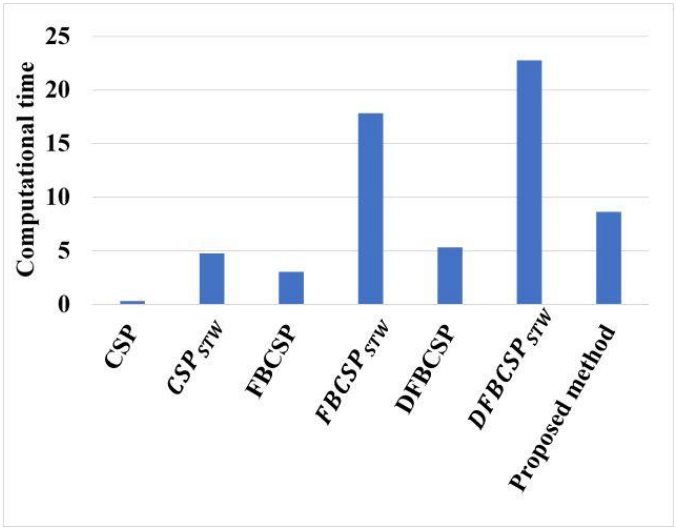


Figure 6.9 Computational time consumed by CSP, CSP_{STW}, FBCSP, $FBCSP_{STW}$, DFBCSP, and $DFBCSP_{STW}$, and the proposed algorithm for best feature space generation.

and created a larger matrix, and then applied the same feature selection approaches as used in DFBCSP and FBCSP, such as fisher's ratio (for $DFBCSP_{STW}$) and mutual information (for $FBCSP_{STW}$). However, this conversion of a multi-view matrix into a single large matrix causes the loss of internal structure of high-order data. It is important to preserve the internal structure of the EEG data during the analysis to minimize the loss of useful information [274]. Our proposed algorithm optimized the multi-view MI data in a multi-task learning framework without losing the internal structure of the multi-view MI data. The study presented in [251] proposed a method, namely, Temporally Constrained Sparse Group Spatial Pattern algorithm (TSGSP), to jointly optimize time windows and frequency bands by formulating a multitask-learning framework. As reported in [251], TSGSP is computationally expensive because of the inner loop runs for three hyperparameters optimization. Instead, our proposed method is based on RNCA, which is relatively faster because we only need to regularize parameter λ for feature optimization.

Besides frequency bands and time windows optimization, trial optimization is also important. Since ERD/ERS patterns changes when the same subject performs the same MI task again. Hence, use of a fixed set of selected time windows and frequency bands CSP features for different trials can reduce the efficiency of the MI classification. In the work of J. Feng et al. [169], a correlation-based relationship between each trail and a reference signal is explored to select significant time window from each trial. With a greater number of constraints (i.e., frequency bands, time windows, and trials), optimizing the higher order feature space requires higher-order multi-task learning approaches. Many subspace regularization algorithms have been finding wide applications in image processing, event-related potential (ERP) analysis in brain signals, BCI rehabilitation systems, and electromyogram (EMG) classification for prosthetic control [275–283]. Besides, tensor algebra and linked feature analysis have shown potential to deal with complex relations in feature space of biomedical signal [274,284–287]. Although we considered only three-dimensional feature space optimization in this study, we consider implementing more advanced structure-preserving multi-view optimization approaches to deal with higher dimensional feature space.

6.6 Conclusion

This work presents a novel feature optimization algorithm based on RNCA for simultaneous selection of the most relevant CSP feature at multiple frequency bands and time windows. In the experiment, multiple overlapped time windows are extracted from MI-related EEG using sliding time window segmentation approach. Then, DTCWT filters each time window into specific frequency subbands, and CSP features are extracted. This created a three-dimensional feature space, which is optimized by the proposed feature selection algorithm. The optimized CSP features trained an SVM classifier for the classification of MI tasks. Further, a comparative study is conducted to investigate the

proposed method's effectiveness in comparison with standard feature optimization algorithms. Obtained superior results suggest that the proposed algorithm shows potential to identify true MI tasks in a practical BCI device.