# A Sliding Window Common Spatial Pattern for Enhancing Motor Imagery Classification in EEG-BCI

Pramod Gaur, Harsh Gupta, Anirban Chowdhury, Karl McCreadie, Ram Bilas Pachori, *Senior Member, IEEE,* Hui Wang, *Senior Member, IEEE*

*Abstract*—Accurate binary classification of electroencephalography (EEG) signals is a challenging task for the development of motor imagery (MI) brain computer interface (BCI) systems. In this study two sliding window techniques are proposed to enhance binary classification of motor imagery (MI). The first one calculates the longest consecutive repetition (LCR) of the sequence of prediction of all the sliding windows which is named as SW-LCR. The second calculates the mode of the sequence of prediction of all the sliding windows and is named SW-Mode. Common spatial pattern (CSP) is used for extracting features with linear discriminant analysis (LDA) used for classification of each time window. Both the SW-LCR and SW-Mode are applied on publicly available BCI Competition IV-2a dataset of healthy individuals and on a stroke patients dataset. As compared to the existing state-of-the-art the SW-LCR performed better in the case of healthy individuals and SW-Mode performed better on stroke patients dataset for left vs. right hand MI with lower standard deviation. For both the datasets the classification accuracy (CA) was approximately 80% and kappa ($\kappa$) was 0.6. The results show that the sliding window based prediction of MI using SW-LCR and SW-Mode is robust against inter-trial and inter-session inconsistencies in the time of activation within a trial and thus can lead to reliable performance in a neurorehabilitative BCI setting.

*Index Terms*—Brain-computer interface, EEG, motor-imagery, common spatial patterns, linear discriminant analysis, neurorehabilitation.

## I. INTRODUCTION

A brain-computer interface (BCI) provides a way for people to control external assistive devices and provides a communication pathway for severely motor impaired people having damaged neuromuscular pathways [1], [2] by translating neurophysiological signals into commands used to control external devices [3]. Motor imagery (MI) is a neuronal activity that occurs when a subject voluntarily imagines making a movement without actually performing it, for example, imagination of the movement of the right hand.

BCI is also being used for motor recovery in patients who have suffered from a stroke or other injury which often leaves them with the inability to move. This is done by measuring changes in neuronal activity over the motor cortex during a MI task whilst providing them with a reward in the form of a virtual avatar performing the same movement and/or with a robotic device[4], [5], [6], [7].

It is clear that the topographies and responses corresponding to limb movements obtained from the beta (13-25 Hz) rhythm are distinct in comparison to the mu (8-13 Hz) rhythm. During limb movements, an increase in the oscillatory power is seen in the beta rhythm in the ipsilateral sensorimotor cortex. At the same time, a decrease in oscillatory power is seen in the contralateral sensorimotor cortex of the mu rhythm [8], [9], [10]. The BCI system aims to translate the changes observed in the mu and beta rhythms into meaningful commands. A major issue in BCI systems reported in the literature is the non-stationarity present in the recorded neurophysiological signals such as EEG. This problem occurs when EEG signals originate from multiple sources. Moreover, the recorded EEG signals have a low signal-to-noise ratio (SNR)[11]. There are several other types of artefacts such as electrical power line, electromyogram (EMG) and electrooculogram (EOG) interference which may contribute to low SNR. Therefore, to ensure a highly performing system, a high SNR must be achieved from the recorded EEG. This is done by removing artefacts in the preprocessing stage which helps to increase the feature set separability of multiple MI tasks [2].

Maddirala and Shaik[12] were able to seperate useful EEG signal sources from unwanted signals such as EOG and EMG. A BCI monitoring system with integrated Augmented Reality (AR) glasses was proposed by Angrisani et al.[13] which handles various applications of maintenance and industrial inspection. They built a low-cost wearable and non-invasive monitoring system for single-channel steady-state visual evoked potential (SSVEP) based BCI [13]. Another group developed a wearable EEG-based device used to assess stress conditions in real-time EEG. The amplitude variations in the EEG signals recorded between the left and right region of prefrontal areas were analyzed using a single EEG differential channel [14]. Another research group also proposed a design using off-the-shelf components with SSVEP-based instrument to improve the performance of a real-time application in BCI [15].

Moreover, EEG signal tends to have poor time-frequency

localization which involves a fixed set of a basis functions such as short-time Fourier transform (STFT) and wavelet transform (WT). It also exhibits highly non-stationary characteristics and suffers from high inter-subject variability thus, resulting in often poor task detection accuracy and high error rates in BCI systems[16], [17]. Another study [18] studied BCI Competition IV dataset 2a using Common Spatial Patterns (CSP) for spatial filtering and Linear Discriminant Analysis (LDA) for classification, taking a fixed time segment of 2s from 2.5s to 4.5s for training and testing [19]. The study examined multiple time segments to see whether it was possible to extract more discriminable features. Recently, a multi-/single-channel empirical mode decomposition (EMD) based filtering approach was developed for binary/multiple class classification problems [20], [21], [22]. Similarly, A combination of multivariate EMD and canonical correlation analysis (CCA) was explored[23] to denoise and remove artefacts. Which helped them to enhance the quality of the EEG signal by enhancing the signal to noise ratio.

Common spatial pattern (CSP) algorithm variants have been studied in the BCI literature [24], [25] to obtain highly separable features in terms of spatial patterns. Another research group computed the feature set as the spatial information present in the sample covariance matrix of the recorded EEG signal [26]. Recently, power spectral covariance matrices were used to detect the sleep stages exploiting the frequency information [27]. There are different areas where these features have helped to obtain better classification such as radar image processing [28] and image processing [29] for example. In a recent study the feasibility of using Spiking Neural Networks (SNN) has also been demonstrated in classifying MI from EEG signals [30]. Some other studies explored the utility of wavelet packet transformation [31] and Feature Priority Analysis [32].

One of the major drawbacks in commonly found methods as described above is the handling of inter-session transfer which reduces the classification accuracy significantly due to inherent non-stationarity in the EEG data. A common approach to solving this problem is to find the most generalizable features. However, the caveat is that the selection of time point within the trial period is generally chosen heuristically. But such choices are always suboptimal and suffer heavily due to session-to-session and even trial-to-trial variabilities. In order to minimize the effect of such inconsistencies, in this work, two novel sliding window based CSP techniques have been introduced which try to articulate the longest consecutive repetition (SW-LCR) and mode (SW-Mode) of the predicted classlabels of each window to decode MI tasks performed by the participants. Thus, the LCR and Mode based approaches are independent of the choice of a particular time point which gives the generalized feature extraction techniques such as CSP an added advatange to augment its performance. The results were obtained for both healthy individuals (using BCI Competition IV-2a dataset) and stroke patients (dataset reported in[33]) to validate the efficiency of the proposed approaches. The major contributions of the paper are highlighted as follows:

- The proposed LCR based approach achieved superior performance for left vs. right hand MI classification as compared to the previous methods with reduced inter-subject variability in the case of healthy individuals.
- The proposed Mode based approach significantly ($p <$ 0.05) outperformed the existing benchmark in the case of stroke patients dataset.
- Both the LCR and Mode based approaches achieved comparable performance between the healthy individuals' and stroke patients' datasets with higher accuracy (close to 80%) which makes it a suitable method for handling non-stationarity which could also be applicable for neurorehabilitative BCI designs.

The paper is organized as follows: Section II describes the dataset and methods used in this study; Section III discusses the CSP method; Section IV discusses the LDA classifier; Section V discusses the proposed method and algorithm; Section VI presents the results and discussion which includes a comparison of the results reported in the literature using the same datasets [18], [34], [33], [35], whilst Section VII concludes the study.

## II. DATASET

### A. Healthy Individuals' Dataset

The BCI Competition IV Dataset 2a was used for validating the results on healthy individuals [34]. This dataset contains EEG data from nine healthy subjects seated in a comfortable armchair in front of a computer screen. The cue-based BCI paradigm consists of four different MI tasks, namely the imagination of movement of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). Each subject has two sessions namely training and test sessions held on different days. Each session comprises of 72 trials per class making a total 288 trials. At the start of the trial (t = 0 s) a fixation cross is shown on a black screen accompanied by a short acoustic warning tone. At t = 2 s an arrow is shown for 1.25 s and its direction: left, right, down or up which corresponds to one of the four classes: 1, 2, 3 or 4. The subjects were asked to perform the MI task till t = 6 s after which a short break was given. The paradigm is illustrated in Fig. 1(a). EEG signals were recorded monopolarly with the left mastoid serving as reference and the right mastoid as ground using 22 Ag/AgCl electrodes with inter-electrode distances of 3.5 cm. All signals were then band-pass filtered from 0.5 to 100 Hz (with a 50 Hz notch filter enabled). The recorded EEG signals were sampled at a frequency of 250 Hz. For this study, classification of all 6 possible combinations of the four classes was used to show effectiveness of the proposed methods.

### B. Stroke patients' dataset

In order to validate the proposed methodology for neurorehabilitative applications another the dataset reported in [33] was also used. The dataset consists of left hand vs. right hand MI data from 10 hemiparetic stroke patients who received multimodal neurofeedback in terms of visual feedback on a computer screen and proprioceptive feedback using a hand exoskeleton which were contigent to their brain activation. There were 2 training sessions and a online feedback session each consisting of 40 trials randomly distributed equally into
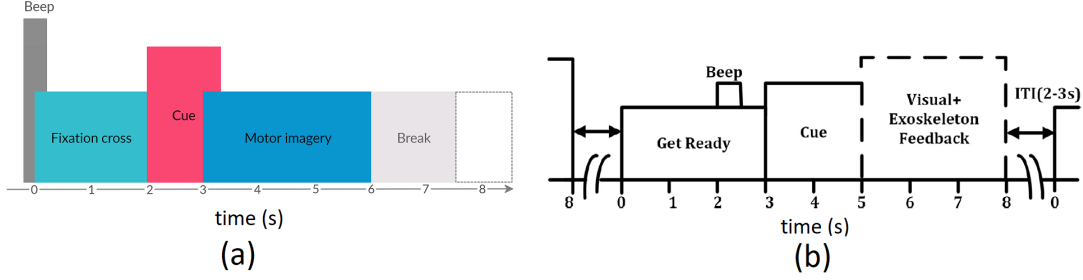
Fig. 1.  Timing diagram of the datasets, (a) BCI competition IV-2a [34], and (b) Stroke patients dataset [33].
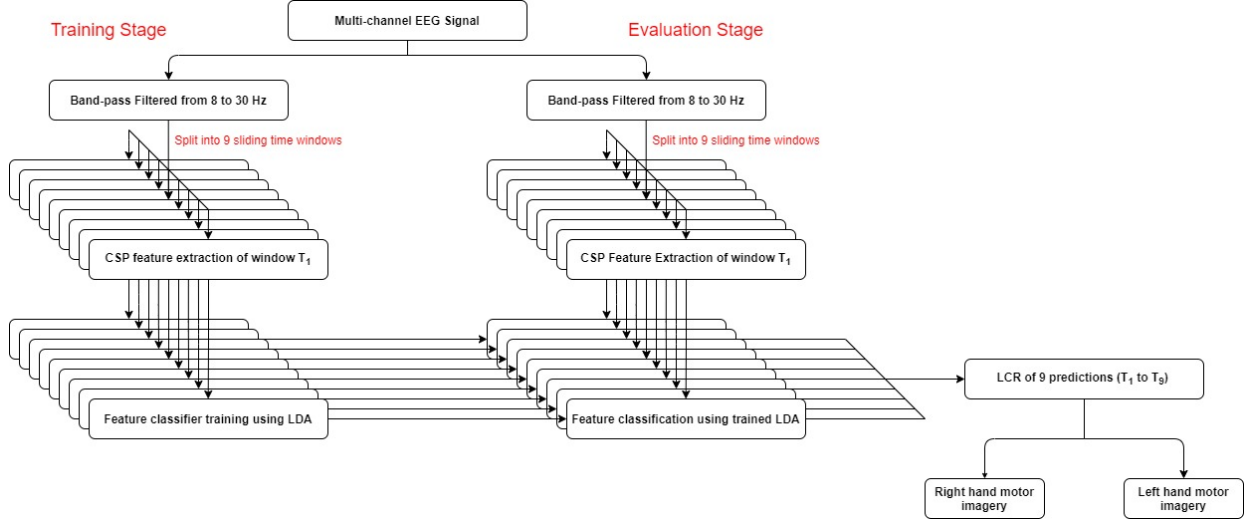


Fig. 2.  Block diagram for the proposed method.

left and right hand tasks, i.e. the chance level of prediction was 50%. A trial starts with a 'Get Ready' cue which lasts for 3 s, followed by a cue for indicating the task (left/right hand task) to perform which lasts for another 5 s during the training session while during the online feedback session this period was divided into two parts: the first 2 s was for displaying the cue and the next 3 s was for displaying the visual feedback and providing the exoskeleton based proprioceptive neurofeedback. There was an inter-trial-interval (ITI) between 2 s to 3 s between two consecutive trials. The timing diagram of the trial is shown in Fig 1(b). The data acquisition was done at 500 Hz sampling rate and initially bandpass filtered between 0.1 Hz to 100 Hz with a notch filter at 50 Hz to cancel the power line noise. There were 12 EEG channels namely, F3, FC3, C3, CP3, P3, FCz, CPz, F4, FC4, C4, CP4, and P4. Only the training data from the first two sessions was used to train the classifier before testing on the data from the online feedback session. A detailed description of the dataset can be found in [33].

## III. COMMON SPATIAL PATTERN (CSP)

CSP aims to learn spatial filters which minimise the variance of a class while maximising the variance of another. It is often helpful to bandpass filter the multichannel EEG signals [19], [36]. The band-power in any given frequency band gives the variance of the filtered EEG signals in the selected band. The

CSP method obtains optimal discrimination for MI based BCI tasks based on band-power features [36]. The CSP method uses the spatial filters $x$ by optimizing the function:

$$Q(x) = \frac{x^T P_1^T P_1 x}{x^T P_2^T P_2 x} = \frac{x^T C_1 x}{x^T C_2 x} \tag{1}$$

where $T$ signifies the transpose of the matrix. $P_i$ gives the training data matrix with sample points as rows and channels as columns. The spatial covariance matrix for a particular class $i$ is $C_i$.

There are many ways to solve this optimization problem but the technique used in this study works by initially visualising that the function $Q(x)$ is unchanged, if the filter $x$ is rescaled. In fact $Q(kx) = Q(x)$, where $k$ gives a real constant indicating that the rescaling of filter $x$ is arbitrary. Therefore, minimising $Q(x)$ is comparable to minimising $x^T C_1 x$ subject to the constraint $x^T C_2 x = 1$ as there is always a possible way to find a re-scaling factor of $x$ such that $x^T C_2 x = 1$. This constrained optimisation problem amounts to minimising the following function using the Lagrange multiplier method:

$$L(\beta, x) = x^T C_1 x - \beta(x^T C_2 x - 1) \tag{2}$$

The derivative of $L$ with regard to $x$ is 0, where $\beta$ is the Lagrange multiplier, and the filters $x$ minimising $L$ are such that :

$$\frac{\partial L}{\partial x} = 2x^T C_1 - 2\beta x^T C_2 = 0$$

$$\Longleftrightarrow C_1 x = \beta C_2 x \Longleftrightarrow C_2^{-1} C_1 x = \beta x \quad (3)$$

Now, this is a standard eigenvalue problem. Hence, the eigenvectors of $Z = C_2^{-1} C_1$ are used to obtain the spatial filters minimising Eq. (1) corresponding to both the largest and the lowest eigenvalues. The features are extracted as the logarithm of EEG signal variance in the selected band after the projection of filters $x$ using the CSP matrix [18].

## IV. LINEAR DISCRIMINANT ANALYSIS (LDA) CLASSIFIER

The LDA based classifier tries to reduce the dimensionality and at the same time protect most of the class discriminatory information [4], [5]. In this paper, the proposed method uses an LDA classifier which is commonly used in EEG-based BCI applications to find the optimum combination of features which provide a better discrimination. For a set of two classes represented by $x_1$ and $x_2$. The classification of the n-dimensional sample points $x = \{x_l, x_2, x_3, ......x_n\}$, $n_1$ samples to the class $w_1$ and $n_2$ samples to the class $w_2$. A line $y = w^t x$ maximizing the discrimination between the two considered classes is tried to archive from a set of all possible lines. In the study, to obtain a good projection vector the measure of separation between the two classes is required. The mean vector of each class in x-space and y-space is given by the following equations:

$$\mu_i = \frac{1}{N_i} \sum_{x \in w_i} x, \quad (4)$$

and

$$v_i = \frac{1}{N_i} \sum_{y \in w_i} y = \frac{1}{N_i} \sum_{y \in w_i} w^t x = w^t \mu_i \quad (5)$$

The distance between two projected means is known as the objective function, which is given by the following equation:

$$J(w) = |v_1 - v_2| = |w^t(\mu_1 - \mu_2)| \quad (6)$$

The standard deviation between classes has not been considered. So, the distance measured between projected means may not always be a good measure. To overcome this limitation Fisher's LDA classifier has been proposed. This method enhances the LDA classifier by determining a decision boundary or a hyperplane in the feature space to classify the features into distinct classes. It determines the separation boundary between two given distributions in terms of the ratio of two group variances as given by following equation:

$$J(w) = \frac{\sigma_{between}^2}{\sigma_{within}^2} = \frac{w^t(\mu_1 - \mu_2)^2}{w^t S_1 w + w^t S_2 w} \quad (7)$$

where $S_1$, $S_2$ are the variances of the feature distribution between two classes $w_1$, $w_2$ respectively and $\mu_1$, $\mu_2$ are the mean of classes. The maximum separation between two classes can be shown by:

$$w^* = (S_1 + S_2)^{-1}(\mu_1 - \mu_2) \quad (8)$$

The $w^*$ is the weight vector which provides the optimum direction for the projection of the data. The following equation is used by the decision boundary in Fisher's LDA to classify the feature vector $d(m)$ as:

$$p(m) = d(m)w^t + b \quad (9)$$

where b is the threshold or bias. The features are assigned to one of the classes based on the sign of the $p(m)$ [10].
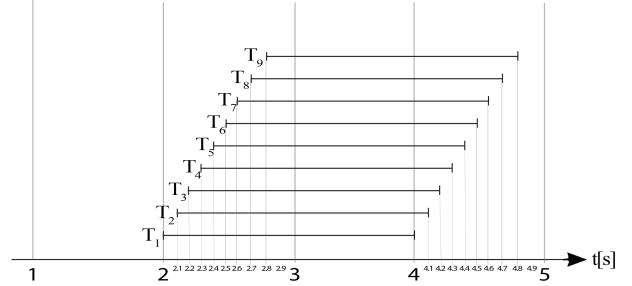


Fig. 3. Sliding time windows $T_1$ to $T_9$.

## V. METHODOLOGY

### A. SW-LCR

In this study, the Sliding Window based Longest Consecutive Repetition (SW-LCR) is used to assign the final prediction. LCR is an element in a sequence which has the longest sub-sequence containing just itself, i.e. in '121111222', '1111' is the longest sub-sequence containing only 1. For example, in sequence 112221121, 2 is the LCR. If there is more than one candidate for LCR then, the earliest one is chosen. For example in sequence 121222111, 2 as LCR is chosen. The block diagram for the proposed methodology is shown in Fig. 2.

A total of 9 overlapping time windows of 2 s, each represented as $T_1$ to $T_9$ (Fig. 3) were extracted while performing MI tasks. There was a difference of 0.1 s between each successive sliding window, wherein the first window ($T_1$) started at 2.0 s after the start of the trial. Further, the second window ($T_2$) was from 2.1 to 4.1 s, similarly the ninth window ($T_9$) was from 2.8 to 4.8 s. EEG data was band-pass filtered between 8 to 30 Hz. Then CSP features were extracted taking three pairs of spatial filters from 100% of the training data. LDA was used to classify CSP features in each sliding time window. A sequence of 9 predicted classes were obtained corresponding to each sliding time window and the LCR of the sequence was then calculated to assign the final prediction.

### B. SW-Mode

Along with LCR another variant was used called SW-Mode. In this approach, out of the 9 overlapping windows as described for SW-LCR, the number of occurrence of class 1 and class 2 are counted. The class containing the maximum count is attributed to that particular trial. This Mode based approach is specially important when the person is unable to concentrate on the task for a long time such as in the case of stroke patients although the overall focus is on the cued task

despite fluctuations in concentration. Such a scenario could generate sequences for example, '122121211', where the LCR is '2' but the count is greater for '1'. Thus, Mode can reduce the misclassifications arising due to such a scenario where LCR would be unable to capture this. This notion supported by the results shown in section VI, where we can see that the SW-Mode performed significantly better than SW-LCR in the case of stroke patients.

A major advantage of the proposed SW-LCR and SW-Mode based strategies is that it can improve the classification accuracy for single-trial prediction irrespective of any feature extraction or classification algorithm used in the background. This is because the proposed sliding window based method looks for the features at different time segments within the trial rather than focusing on a single time point which may vary from trial to trial and session to session. Thus sliding windows captures the idea that change in the mental state is not instantaneous and spread across the whole trial. This is particularly important when the user has a lesser capacity to engage with the task such as for stroke patients for whom the response time after the cue may vary significantly. Looking at all the small windows of time within a trial and then deciding on the particular class of the trial gives more consistent results. However, the disadvantage is that in real time BCI the proposed sliding window techniques can add latency in the issued feedback.

## VI. RESULTS AND DISCUSSION

There are four MI tasks and the proposed method is used to classify the CSP features into two classes with six combinations of two MI tasks as follows: left and right (L&R), left and foot (L&F), left and tongue (L&T), right and foot (R&F), right and tongue (R&T), and foot and tongue (F&T). The classification accuracy is calculated for all nine subjects for each evaluation session.

Table I and Table II report the classification accuracy obtained on BCI Competition IV datatset 2a by applying the proposed SW-LCR and SW-Mode methods respectively. For L&R, L&R, and L&T MI tasks, SW-LCR provided better average classification accuracy (ACA) of 80.02%, 83.64%, and 86.19% respectively than in SW-Mode. On the other hand, SW-Mode achieved better ACA of 85.03%, 84.03%, and 73.38% respectively in R&F, R&T, and F&T than in SW-LCR. The corresponding kappa values against these classification accuracies has been reported in Table III and Table IV for SW-LCR and SW-Mode respectively.

Table V demonstrates left and right MI classification accuracy of the proposed methods as compared to competing comparable methods M1, M2, M3 and M4. We have chosen L&R over other binary class combinations for comparison as this is widely used in many BCI applications. Another reason is that the stroke patients' data used in this paper is also based on left vs. right hand task and it would be easier to compare the results for healthy vs. patients in that manner.

Method 1 (M1) shows results[21] based on a tangent space based transfer learning technique by Gaur et al. The ACA of the SW-LCR ($80.02\% \pm 13.45$) is higher when compared with

TABLE I
CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION (STD) FOR THE PROPOSED SW-LCR CLASSIFICATION METHOD APPLIED TO BCI COMPETITION IV DATASET 2A.

| Subject | Accuracy with proposed SW-LCR method (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | L&R | L&F | L&T | R&F | R&T | F&T | Avg |
| A01 | 86.81 | 97.22 | 97.22 | 97.22 | 100 | 69.44 | 91.32 |
| A02 | 64.58 | 63.89 | 65.97 | 80.56 | 66.67 | 73.61 | 69.21 |
| A03 | 95.83 | 93.06 | 94.44 | 93.06 | 94.44 | 69.44 | 90.05 |
| A04 | 67.36 | 82.64 | 88.19 | 89.58 | 86.81 | 62.5 | 79.51 |
| A05 | 68.06 | 65.97 | 74.31 | 70.83 | 65.97 | 68.06 | 68.87 |
| A06 | 67.36 | 70.83 | 72.22 | 64.58 | 71.53 | 70.14 | 69.44 |
| A07 | 80.56 | 97.92 | 93.75 | 93.75 | 93.75 | 79.86 | 89.93 |
| A08 | 97.22 | 84.72 | 92.36 | 88.19 | 89.58 | 78.47 | 88.43 |
| A09 | 92.36 | 96.53 | 97.22 | 84.03 | 82.64 | 85.42 | 89.70 |
| Average | 80.02 | 83.64 | 86.19 | 84.64 | 83.49 | 72.99 | |
| Std | 13.45 | 13.75 | 12.02 | 10.95 | 12.65 | 7.07 | |

TABLE II
CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION (STD) FOR THE PROPOSED SW-MODE CLASSIFICATION METHOD APPLIED TO BCI COMPETITION IV DATASET 2A.

| Subject | Accuracy with proposed SW-Mode method (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | L&R | L&F | L&T | R&F | R&T | F&T | Avg |
| A01 | 86.11 | 96.53 | 96.53 | 97.22 | 100 | 70.83 | 91.20 |
| A02 | 64.58 | 64.58 | 65.97 | 80.56 | 66.67 | 75 | 69.56 |
| A03 | 95.83 | 93.06 | 94.44 | 93.06 | 95.14 | 70.14 | 90.28 |
| A04 | 64.58 | 84.72 | 88.89 | 90.28 | 89.58 | 63.19 | 80.21 |
| A05 | 68.06 | 65.28 | 75 | 72.22 | 66.67 | 67.36 | 69.10 |
| A06 | 68.75 | 70.14 | 70.83 | 64.58 | 71.53 | 70.14 | 69.33 |
| A07 | 81.94 | 97.92 | 93.06 | 93.75 | 93.75 | 79.86 | 90.05 |
| A08 | 97.22 | 84.03 | 92.36 | 88.89 | 90.28 | 78.47 | 88.54 |
| A09 | 90.97 | 97.22 | 97.22 | 84.72 | 82.64 | 85.42 | 89.70 |
| Average | 79.78 | 83.72 | 86.03 | 85.03 | 84.03 | 73.38 | |
| Std | 13.48 | 13.82 | 12.04 | 10.8 | 12.77 | 6.92 | |

TABLE III
KAPPA VALUE FOR THE PROPOSED SW-LCR CLASSIFICATION METHOD APPLIED TO BCI COMPETITION IV DATASET 2A.

| Subject | Accuracy with proposed SW-LCR method (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | L&R | L&F | L&T | R&F | R&T | F&T | Avg |
| A01 | 0.74 | 0.94 | 0.94 | 0.94 | 1 | 0.39 | 0.83 |
| A02 | 0.29 | 0.28 | 0.32 | 0.61 | 0.33 | 0.47 | 0.38 |
| A03 | 0.92 | 0.86 | 0.89 | 0.86 | 0.89 | 0.39 | 0.80 |
| A04 | 0.35 | 0.65 | 0.76 | 0.79 | 0.74 | 0.25 | 0.59 |
| A05 | 0.36 | 0.32 | 0.49 | 0.42 | 0.32 | 0.36 | 0.38 |
| A06 | 0.35 | 0.42 | 0.44 | 0.29 | 0.43 | 0.4 | 0.39 |
| A07 | 0.61 | 0.96 | 0.88 | 0.88 | 0.88 | 0.6 | 0.80 |
| A08 | 0.94 | 0.69 | 0.85 | 0.76 | 0.79 | 0.57 | 0.77 |
| A09 | 0.85 | 0.93 | 0.94 | 0.68 | 0.65 | 0.71 | 0.79 |
| Average | 0.6 | 0.67 | 0.72 | 0.69 | 0.67 | 0.46 | |

TABLE IV
KAPPA VALUE FOR THE SW-MODE CLASSIFICATION METHOD APPLIED TO BCI COMPETITION IV DATASET 2A.

| Subject | Accuracy with proposed SW-Mode method (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | L&R | L&F | L&T | R&F | R&T | F&T | Avg |
| A01 | 0.72 | 0.93 | 0.93 | 0.94 | 1 | 0.42 | 0.82 |
| A02 | 0.29 | 0.29 | 0.32 | 0.61 | 0.33 | 0.5 | 0.39 |
| A03 | 0.92 | 0.86 | 0.89 | 0.86 | 0.9 | 0.4 | 0.81 |
| A04 | 0.29 | 0.69 | 0.78 | 0.81 | 0.79 | 0.26 | 0.60 |
| A05 | 0.36 | 0.31 | 0.5 | 0.44 | 0.33 | 0.35 | 0.38 |
| A06 | 0.38 | 0.4 | 0.42 | 0.29 | 0.43 | 0.4 | 0.39 |
| A07 | 0.64 | 0.96 | 0.86 | 0.88 | 0.88 | 0.6 | 0.80 |
| A08 | 0.94 | 0.68 | 0.85 | 0.78 | 0.81 | 0.57 | 0.77 |
| A09 | 0.82 | 0.94 | 0.94 | 0.69 | 0.65 | 0.71 | 0.79 |
| Average | 0.6 | 0.67 | 0.72 | 0.7 | 0.68 | 0.47 | |

M1 ($75.52\% \pm 14.39$). Notably, six of the nine subjects have shown improvement, with an improved ACA of 4.5%. Also the standard deviation (Std) of the accuracy of all subjects has been reduced by roughly 1%. Subject A07 shows an improvement of $>20.5\%$, whereas A05 shows improvement of $>11\%$, and A09 improves $>8.5\%$. Three subjects (A02, A03 and A08) show improvement of $>3\%$, with subjects A01 and A06 declining in performance by 2.1% and A04 by 7%.

Method 2 (M2) shows results obtained from the evaluation session from a different study by Gaur et al.[20] which uses a subject-specific multivariate empirical mode decomposition based filtering method (SS-MEMDBF) for preprocessing and implements a Riemannian geometry framework designed separately for each subject for classification. The average classification accuracy of the SW-LCR ($80.02\% \pm 13.45$) is comparable to method M2 ($79.93\% \pm 14.99$). Notably, five out of the nine subjects have shown improvement, with an improved ACA of 0.1% and also the Std of the accuracy of all subjects has been reduced by roughly 1.5%. Subject A05 shows improvement of 9.5%, whilst subject A02 improved by 4%. Three subjects (A03, A07 and A08) show improvement $<2\%$, whilst subjects A06 and A09 declined in performance by $<1.5\%$. A01 has declined by $>4.5\%$ and A04 by $>9\%$. The performance of the SW-LCR is comparable to that of M2 in 4 out of 6 combinations although the average accuracy is slightly lower than M2. But more importantly, it can be observed that the SW-LCR can lead to more uniformity in performance across all the subjects as the standard deviation is lower as compared to M2. This is due to the fact that the LCR based sliding window approach can enhance the performance of low performing subjects such as A02 and A05 as they suffer from lower SNR and higher non-stationarity in inter-session transfer. The use of fixed time points such as 0.5s to 2.5s in the case of M2 is in fact disadvantageous to handle such inter-session non-stationarities as the time window is selected on the basis of the training session which is prone to be shifted during the evaluation session. On the contrary, the SW-LCR does not depend on such an assumption and observes a larger time scale through successive windowing and the label of that particular trial is decided from the outcomes of those windows using LCR. Thus the small shifts in the time point of activation between training and evaluation sessions are accounted for. M2 also comes with an overhead of a subject-specific multivariate empirical mode decomposition based filtering method (SS-MEMDBF) for preprocessing making it difficult to implement online due to higher computational complexity. Another advantage of the LCR based sliding window approach is that it can be used with any feature extraction or classification algorithms and therefore, limitations which come with the use of a particular feature type (such as in M2) can be avoided.

Method 3 (M3) implements CSP and uses detection of the covariate shift and adaptive learning [37]. The ACA of the SW-LCR ($80.02\% \pm 13.45$) is better when compared with method M3 ($74.92\% \pm 15.43$). Notably, eight out of the nine subjects have shown improvement, an ACA improvement $>5\%$ and also the Std of the accuracy of all subjects has been reduced by $<2\%$. Subject A07 shows improvement of $>19.4\%$, whilst A05 and A02 have shown improvement of $>6.9\%$. A09 shows

TABLE V
CLASSIFICATION ACCURACY (%) FOR LEFT VS. RIGHT HAND MI TASK USING SW-LCR AND SW-MODE COMPARED TO METHODS M1, M2, M3 AND M4 ON BCI COMPETITION IV DATASET 2A.

| Subject | SW-LCR | SW-Mode | M1 | M2 | M3 | M4 |
|---------|--------|---------|-----|-----|-----|-----|
| A01 | 86.81 | 86.11 | 88.65 | **91.49** | 90.28 | 88.89 |
| A02 | **64.58** | **64.58** | 61.27 | 60.56 | 57.64 | 51.39 |
| A03 | 95.83 | 95.83 | 91.24 | 94.16 | 95.14 | **96.53** |
| A04 | 67.36 | 64.58 | 74.14 | **76.72** | 65.97 | 70.14 |
| A05 | **68.06** | **68.06** | 57.04 | 58.52 | 61.11 | 54.86 |
| A06 | 67.36 | 68.75 | 69.44 | 68.52 | 65.28 | **71.53** |
| A07 | 80.56 | **81.94** | 60 | 78.57 | 61.11 | 81.25 |
| A08 | **97.22** | **97.22** | 94.03 | 97.01 | 91.67 | 93.75 |
| A09 | 92.36 | 90.97 | 83.85 | **93.85** | 86.11 | 93.75 |
| Average | **80.02** | 79.78 | 75.52 | 79.93 | 74.92 | 78.01 |
| Std | **13.45** | 13.48 | 14.39 | 14.99 | 15.43 | 17.01 |

TABLE VI
CLASSIFICATION ACCURACIES (%) AND KAPPA VALUES FOR THE **STROKE PATIENTS' DATASET** USING **SW-LCR**

| Subject | 5-CV Training Accuracy | Training kappa | Testing Accuracy | Testing kappa |
|---------|------------------------|----------------|------------------|---------------|
| A01 | 68.75 | 0.38 | 85 | 0.7 |
| A02 | 87.5 | 0.75 | 80 | 0.6 |
| A03 | 87.5 | 0.75 | 62.5 | 0.25 |
| A04 | 81.25 | 0.63 | 80 | 0.6 |
| A05 | 93.75 | 0.88 | 60 | 0.2 |
| A06 | 100 | 1 | 52.5 | 0.05 |
| A07 | 100 | 1 | 80 | 0.6 |
| A08 | 93.75 | 0.88 | 82.5 | 0.65 |
| A09 | 81.25 | 0.63 | 70 | 0.4 |
| A10 | 87.5 | 0.75 | 82.5 | 0.65 |
| Average | 88.13 | 0.77 | 73.5 | 0.47 |
| Std | 9.52 | 0.19 | 11.44 | 0.23 |

improvement of $>6.2\%$, whereas A08 shows improvement of $>5.5\%$. A06 shows improvement of $>2\%$, with two subjects (A03 and A04) showing an improvement of $<1.5\%$ with A01 declining by $<3.5\%$.

Method 4 (M4) uses CSP on band-pass filtered EEG between 8 and 30 Hz before calculating the log variance from three pairs of filters for extraction of features and uses LDA for a binary classification problem [18]. The ACA of the SW-LCR ($80.02\% \pm 13.45$) is higher when compared with method M4 ($78.01\% \pm 17.01$). Notably, three out of the nine subjects have shown improvement, with the ACA improving by $>2\%$ and also the Std of the accuracy of all subjects has been reduced by $>3.5\%$. A05 and A02 have shown improvement of $>13.1\%$, whilst A08 shows improvement of $>3.4\%$. A06 has declined by $<4.2\%$. The remaining five subjects are within 3% of the results.

Apart from M1, M2, M3, and M4, there are a few emerging techniques which were also recently applied on the BCI Competition IV-2a dataset. For example, a Spiking neural network [30] based technique yielded an ACA of 75.62% for left vs. right hand MI task which is lower than the proposed methods of SW-LCR (ACA=80.02%) and SW-Mode (ACA=79.78%). Among the deep learning based approaches EEGNet achieved an ACA of 68.98% and NSL-EEGNet [38] achieved 70.6% which are significantly ($p < 0.05$) lower than the proposed SW-LCR and SW-Mode based approaches.

The results on the stroke patients' dataset are given in Table VI and Table VII for the SW-LCR and SW-Mode

TABLE VII
CLASSIFICATION ACCURACIES (%) AND KAPPA VALUES FOR THE **STROKE PATIENTS' DATASET** USING **SW-MODE**

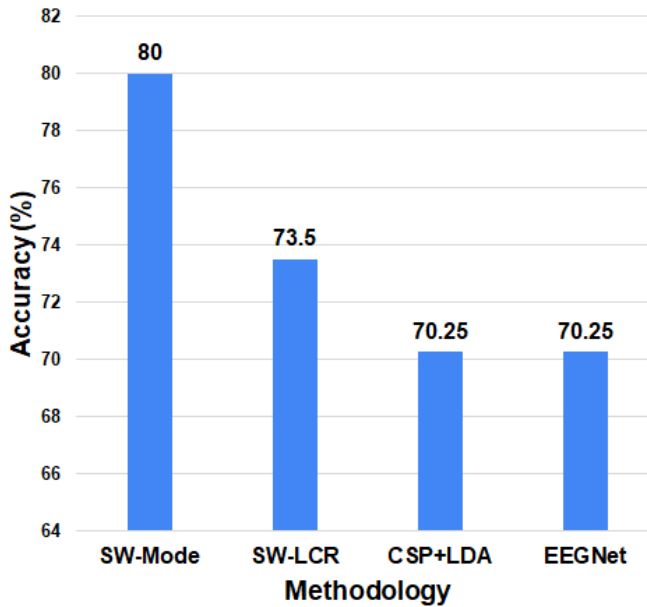| Subject | 5-CV Training Accuracy | Training kappa | Testing Accuracy | Testing kappa |
|---|---|---|---|---|
| A01 | 81.25 | 0.61 | 87.5 | 0.75 |
| A02 | 100 | 1 | 87.5 | 0.75 |
| A03 | 100 | 1 | 65 | 0.3 |
| A04 | 75 | 0.5 | 70 | 0.4 |
| A05 | 87.5 | 0.74 | 75 | 0.5 |
| A06 | 93.75 | 0.87 | 65 | 0.3 |
| A07 | 87.5 | 0.75 | 90 | 0.8 |
| A08 | 75 | 0.49 | 90 | 0.8 |
| A09 | 81.25 | 0.61 | 75 | 0.5 |
| A10 | 93.75 | 0.87 | 95 | 0.9 |
| Average | 87.5 | 0.74 | 80 | 0.6 |
| Std | 9.32 | 0.19 | 11.24 | 0.22 |



Fig. 4. Comparison of classification accuracies of the proposed methods with previous methods applied on the stroke patients' dataset.
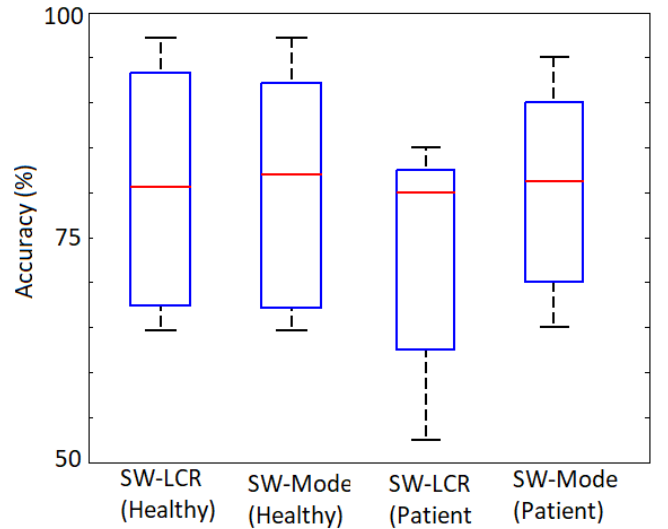


Fig. 5. Comparison of classification accuracy distributions of the proposed methods between the healthy and stroke patients' datasets.

73.5% with kappa 0.47). A deep learning based approach was also applied on the same dataset previously [35] using the popular EEGNet architecture. They obtained an average classification accuracy of 70.25% across all the subjects which is also outperformed significantly ($p < 0.05$) by the proposed method in the case of the SW-Mode based approach (average accuracy 80%). These comparisons are depicted in Fig. 4 with a bar graph.

It is also interesting to note that the proposed method gave comparable performance in healthy (BCI Competition IV Dataset 2a) and stroke patients' datasets which is an indication of the potential robustness of the algorithm. The average accuracy and kappa in the case of healthy individuals is around 80.02% with kappa at 0.6 while in the case of patients the average accuracy is at 80% (using SW-Mode) with the same kappa value (0.6). The comparison of the distribution of accuracies across different subjects for the proposed methods (SW-LCR and SW-Mode) in the case of healthy and patients are shown in the boxplot given in Fig. 5. However, it can be seen that SW-LCR performed better in the case of healthy individuals than in the case of stroke patients while SW-Mode was better for stroke patients. This might be related to lack of focus during the task often observed in stroke patients as they can get tired quickly and fail to generate longest consecutive repetition of predicted labels needed for good SW-LCR results. On the other hand, as SW-Mode depends on the most frequent predicted label within a trial rather than depending on the longest sequence, it can handle the issue of lack of focus more effectively.

In this study the proposed method (M1) was compared with three other methods denoted in the manuscript as M2, M3, and M4. Among these methods M1, M3, and M4 are based on common spatial patterns (CSP) with some variations, while M2 is based on multivariate empirical mode decomposition (MEMD) based filtering and Riemannian geometry (RG). In comparison to the CSP based methods the computational time

based approaches respectively. The reason we have chosen this dataset is because the stroke patients data has more non-stationarity due to altered neurodynamics and lack of engagement during the task. It can be seen from Table VI that in the case of SW-LCR the average classification accuracy in training (5-fold cross validation) is 88.13%±9.52 while in testing it is 73.5%±11.44. The average kappa value in training (5-fold cross validation) is 0.77±0.19, while for testing it is 0.47±0.23. Again, for the SW-Mode based approach (Table VII, the average classification accuracy in training (5-fold cross validation) is 87.5%±9.32 while in testing it is 80%±11.24. The average kappa value in training (5-fold cross validation) is 0.74±0.19, while for testing it is 0.6±0.22. It is important to note that the previous benchmark accuracy on the same dataset as reported in [33] was 70.25% with a kappa of 0.41, which is significantly ($p < 0.05$) outperformed by the proposed method in the case of the SW-Mode based approach (average accuracy 80% with kappa 0.6) and in the case of SW-LCR also the performance is higher (average accuracy

for MEMD is enormous which impedes its utility for real time application [39]. For example, it takes around one hour to execute the empirical mode decomposition (EMD) with 30,000 data points using a standard modern personal computer as reported in [40]. This creates additional dependancy on parallel computing architecture such as Cuda or hardware resources such as FPGA [41]. The typical time complexity of MEMD is $O(n \log n)$ where $n$ is the data length [40]. Additionally, M2 requires the computation of Riemannian Geometry (RG) with complexity $O(n^3)$ ($n$ =number of EEG electrodes), which makes the real time implementation more difficult. The other methods M1, M3, M4 computes the features using pre-computed CSP weight matrix generated during the calibration process and requires only matrix multiplication of the bandpass filtered EEG data with the CSP weight matrix, the processing time of which is in miliseconds. For example, to make every single prediction in M1 required around 62.5 ms which is good enough for issuing intuitive neurofeedback. The hardware used in the experiments was a laptop with Intel® Core™ i5-8250U as CPU and 8GB DDR4 RAM.

## VII. CONCLUSION

EEG data is often recorded across session as well as across subject (sometimes from those with brain injuries) and tends to be highly non-stationary with inter-subject variability. In this work, two sliding window based CSP technique, namely SW-LCR and SW-Mode were introduced to classify EEG signals into multiple MI tasks. The proposed methods take advantage of examining multiple time segments within a trial rather than depending on a single time point which is more robust against inter-trial and inter-session variabilities. With the proposed method, inter-subject variability has been reduced to some extent as evidenced by the lower standard deviation. Moreover, the SW-Mode has performed significantly better than the existing state-of-the-art on the stroke patients' dataset while SW-LCR performed better on healthy individuals. The performance of the proposed methods are also comparable for both healthy individuals and stroke patients with accuracy around 80% with a higher kappa value of 0.6, which is an important observation since the EEG data from stroke patients suffers from lower SNR and higher non-stationarity due to the presence of altered neurodynamics. Thus SW-LCR and SW-Mode has the potential to augment the performance of MI-BCI and can pave the way for reliable neurorehabilitative BCI applications.

## REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, T. M. Vaughan *et al.*, "Brain-computer interface technology: a review of the first international meeting," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 164–173, 2000.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[3] J. Wolpaw and E. W. Wolpaw, *Brain-computer interfaces: principles and practice*. OUP USA, 2012.

[4] F. Lotte, "A tutorial on EEG signal-processing techniques for mental-state recognition in brain–computer interfaces," in *Guide to Brain-Computer Music Interfacing*. Springer, 2014, pp. 133–161.

[5] F. Lotte, M. Congedo, A. Lécuyer, and F. Lamarche, "A review of classification algorithms for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 4, 2007.

[6] A. Chowdhury, Y. K. Meena, H. Raza, B. Bhushan, A. K. Uttam, N. Pandey, A. A. Hashmi, A. Bajpai, A. Dutta, and G. Prasad, "Active physical practice followed by mental practice using bci-driven hand exoskeleton: A pilot trial for clinical effectiveness and usability," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 6, pp. 1786–1795, 2018.

[7] A. Chowdhury, A. Dutta, and G. Prasad, "Corticomuscular co-activation based hybrid brain-computer interface for motor recovery monitoring," *IEEE Access*, vol. 8, pp. 174 542–174 557, 2020.

[8] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 4, pp. 317–326, 2008.

[9] V. Gandhi, G. Prasad, D. Coyle, L. Behera, and T. M. McGinnity, "Quantum Neural Network-Based EEG Filtering for a Brain–Computer Interface," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 278–288, 2014.

[10] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "An empirical mode decomposition based filtering method for classification of motor-imagery EEG signals for enhancing brain-computer interface," in *International Joint Conference on Neural Networks*, 2015, pp. 1–7.

[11] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[12] A. K. Maddirala and R. A. Shaik, "Separation of sources from single-channel EEG signals using independent component analysis," *IEEE Transactions on Instrumentation and measurement*, vol. 67, no. 2, pp. 382–393, 2017.

[13] L. Angrisani, P. Arpaia, A. Esposito, and N. Moccaldi, "A Wearable Brain–Computer Interface Instrument for Augmented Reality-Based Inspection in Industry 4.0," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1530–1539, 2019.

[14] P. Arpaia, N. Moccaldi, R. Prevete, I. Sannino, and A. Tedesco, "A wearable EEG instrument for real-time frontal asymmetry monitoring in worker stress analysis," *IEEE Transactions on Instrumentation and Measurement*, 2020.

[15] L. Angrisani, P. Arpaia, D. Casinelli, and N. Moccaldi, "A single-channel SSVEP-based instrument with off-the-shelf components for trainingless brain-computer interfaces," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 10, pp. 3616–3625, 2018.

[16] Bahar Hatipoglu Yilmaz, Cagatay Murat Yilmaz, and t. y. v. n. p. Cemal Kose, journal=Medical and Biological Engineering and Computing.

[17] S. R. Sreeja, Himanshu, and j. t. y. v. n. p. Samanta, Debasis.

[18] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.

[19] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *Signal Processing Magazine, IEEE*, vol. 25, no. 1, pp. 41–56, 2008.

[20] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry," *Expert Systems with Applications*, vol. 95, pp. 201–211, 2018.

[21] P. Gaur, K. McCreadie, R. B. Pachori, H. Wang, and G. Prasad, "Tangent Space Features-Based Transfer Learning Classification Model for Two-Class Motor Imagery Brain–Computer Interface," *International Journal of Neural Systems*, vol. 29, no. 10, p. 1950025, 2019.

[22] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "An Automatic Subject Specific Intrinsic Mode Function Selection for Enhancing Two-Class EEG-Based Motor Imagery-Brain Computer Interface," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6938–6947, 2019.

[23] X. Chen, X. Xu, A. Liu, M. J. McKeown, and Z. J. Wang, "The use of multivariate EMD and CCA for denoising muscle artifacts from few-channel EEG recordings," *IEEE transactions on instrumentation and measurement*, vol. 67, no. 2, pp. 359–370, 2017.

[24] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, 2012.

[25] H. Zhang, H. Yang, and C. Guan, "Bayesian learning for spatial filtering in an EEG-based brain–computer interface," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 7, pp. 1049–1060, 2013.

[26] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," *IEEE*

*Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2012.

[27] Y. Li, K. M. Wong *et al.*, "EEG signal classification based on a Riemannian distance measure," in *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*. IEEE, 2009, pp. 268–273.

[28] F. Barbaresco, "Innovative tools for radar signal processing based on cartans geometry of SPD matrices & information geometry," in *2008 IEEE Radar Conference*. IEEE, 2008, pp. 1–6.

[29] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on riemannian manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1713–1727, 2008.

[30] C. D. Virgilio G., J. H. Sossa A., J. M. Antelis, and L. E. Falcn, "Spiking neural networks applied to the classification of motor tasks in eeg signals," *Neural Networks*, vol. 122, pp. 130 – 143, 2020.

[31] M. A. Lee, F. Khanam, M. Ahmad, and M. S. Uddin, "Multiclass eeg signal classification utilizing rnyi min-entropy-based feature selection from wavelet packet transformation," *Brain Informatics*, vol. 7, no. 1, p. 7, 2020.

[32] S. Li and H. Feng, "Eeg signal classification method based on feature priority analysis and cnn," in *2019 International Conference on Communications, Information System and Computer Engineering (CISCE)*, 2019, pp. 403–406.

[33] A. Chowdhury, H. Raza, Y. K. Meena, A. Dutta, and G. Prasad, "Online covariate shift detection-based adaptive braincomputer interface to trigger hand exoskeleton feedback for neuro-rehabilitation," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 4, pp. 1070–1080, 2018.

[34] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI Competition 2008–Graz data set A," *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, pp. 136–142, 2008.

[35] H. Raza, A. Chowdhury, and S. Bhattacharyya, "Deep learning based prediction of eeg motor imagery of stroke patients for neuro-rehabilitation application," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–8.

[36] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[37] H. Raza, H. Cecotti, Y. Li, and G. Prasad, "Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface," *Soft Computing*, pp. 1–12, 2015.

[38] H. Raza, A. Chowdhury, S. Bhattacharyya, and S. Samothrakis, "Single-trial eeg classification with eegnet and neural structured learning for improving bci performance," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–8.

[39] Y.-H. Wang, C.-H. Yeh, H.-W. V. Young, K. Hu, and M.-T. Lo, "On the computational complexity of the empirical mode decomposition algorithm," *Physica A: Statistical Mechanics and its Applications*, vol. 400, pp. 159 – 167, 2014.

[40] L.-C. Wu, H.-H. Chen, J.-T. Horng, C. Lin, N. E. Huang, Y.-C. Cheng, and K.-F. Cheng, "A novel preprocessing method using hilbert huang transform for maldi-tof and seldi-tof mass spectrometry data," *PLOS ONE*, vol. 5, no. 8, pp. 1–15, 08 2010.

[41] M. Lee, K. Shyu, P. Lee, C. Huang, and Y. Chiu, "Hardware implementation of emd using dsp and fpga for online signal processing," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 6, pp. 2473–2481, 2011.