

Article

Investigating Feature Selection Techniques to Enhance the Performance of EEG-Based Motor Imagery Tasks Classification

Md. Humaun Kabir ¹, Shabbir Mahmood ¹, Abdullah Al Shiam ², Abu Saleh Musa Miah ³,
Jungpil Shin ^{3,*} and Md. Khademul Islam Molla ^{4,*}

¹ Department of Computer Science and Engineering, Bangamata Sheikh Fojilatunnesa Mujib Science & Technology University, Jamalpur 2012, Bangladesh

² Department of Computer Science and Engineering, Sheikh Hasina University, Netrokona 2400, Bangladesh

³ School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu 965-8580, Japan

⁴ Department of Computer Science and Engineering, University of Rajshahi, Rajshahi 6205, Bangladesh

* Correspondence: jpschin@u-aizu.ac.jp (J.S.); khademul.cse@ru.ac.bd (M.K.I.M.)

Abstract: Analyzing electroencephalography (EEG) signals with machine learning approaches has become an attractive research domain for linking the brain to the outside world to establish communication in the name of the Brain-Computer Interface (BCI). Many researchers have been working on developing successful motor imagery (MI)-based BCI systems. However, they still face challenges in producing better performance with them because of the irrelevant features and high computational complexity. Selecting discriminative and relevant features to overcome the existing issues is crucial. In our proposed work, different feature selection algorithms have been studied to reduce the dimension of multiband feature space to improve MI task classification performance. In the procedure, we first decomposed the MI-based EEG signal into four sets of the narrowband signal. Then a common spatial pattern (CSP) approach was employed for each narrowband to extract and combine effective features, producing a high-dimensional feature vector. Three feature selection approaches, named correlation-based feature selection (CFS), minimum redundancy and maximum relevance (mRMR), and multi-subspace randomization and collaboration-based unsupervised feature selection (SRCFS), were used in this study to select the relevant and effective features for improving classification accuracy. Among them, the SRCFS feature selection approach demonstrated outstanding performance for MI classification compared to other schemes. The SRCFS is based on the multiple k-nearest neighbour graphs method for learning feature weight based on the Laplacian score and then discarding the irrelevant features based on the weight value, reducing the feature dimension. Finally, the selected features are fed into the support vector machines (SVM), linear discriminative analysis (LDA), and multi-layer perceptron (MLP) for classification. The proposed model is evaluated with two benchmark datasets, namely BCI Competition III dataset IVA and dataset IIIB, which are publicly available and mainly used to recognize the MI tasks. The LDA classifier with the SRCFS feature selection algorithm exhibits better performance. It proves the superiority of our proposed study compared to the other state-of-the-art BCI-based MI task classification systems.

Keywords: BCI; automatic feature selection; CFS; mRMR; SRCFS; CSP; MI classification; SVM; LDA; MLP

MSC: 68T10



Citation: Kabir, M.H.; Mahmood, S.; Al Shiam, A.; Musa Miah, A.S.; Shin, J.; Molla, M.K.I. Investigating Feature Selection Techniques to Enhance the Performance of EEG-Based Motor Imagery Tasks Classification. *Mathematics* **2023**, *11*, 1921. <https://doi.org/10.3390/math11081921>

Academic Editors: Grigoreta-Sofia Cojocar, Adriana-Mihaela Guran and Laura-Silvia Dioşan

Received: 10 March 2023

Revised: 15 April 2023

Accepted: 17 April 2023

Published: 19 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Brain-Computer Interface (BCI) is a promising technology mainly used to help the neuromuscular disorders of paralyzed patients and in motor rehabilitation centres. It also established a linking channel and control capabilities to transform messages between the electronic devices and the brain [1–3]. In recent decades, BCI-related systems have gained exponential importance due to the numerous applications in different sectors, specifically in

the neuro-engineering and neuroscience fields. It has encouraged to use of neuroplasticity in brain stroke patients. In addition, it has made a huge contribution to people with disabilities to help them communicate with other people using emotion [4,5], event-related potential detection [6], and sleep detection [7]. Furthermore, it can collaborate with other individuals with disabilities to articulate their needs, ideas, and thoughts and assist in operating their assistive devices, such as wheelchairs. It also aids in the execution of daily tasks without physical movement by detecting emotions. BCI applications span from communication and rehabilitation to entertainment. Recently, researchers have integrated BCI with artificial intelligence (AI) and created adaptable BCI systems that enable the control of various robotic equipment through brain activity. For example, brain-controlled home automation, robotic arms, and prosthetic arms [3–5,8]. The main reason for using a robotic or prosthetic arm is that brain activity and thinking commands cannot pass through the muscle and peripheral nerves. At the same time, we collect the signal through the electroencephalogram (EEG) sensor and translate it into a digital command to control the assistive devices for locked-in people.

There are various ways to measure and capture brain activity in a non-alive approach: EEG, magnetoencephalogram (MEG), and functional magnetic response imaging (fMRI) are most of them. Among them, the BCI system with EEG signal is the most cost-effective and can be implemented with minimal clinical risk because the non-invasive approach does not require any operation; however, it needs some electrodes on the scalp [9–11]. Here, the person needs to imagine a specific muscle movement or limb movement without any patient action (motor action). That imagination makes a great oscillatory action with rhythmic tremors which is known as different kinds of event-related function ERD or ERS, which can be recognized with a machine learning algorithm [12,13]. The main goal of the BCI-based application is to identify actual human activity during the MI task aiming to translate human thinking to the corresponding digital command, which can be controlled by different kinds of machines. To implement the goal, researchers have been working to extract effective features and search the compatible machine learning algorithm for classification.

Various feature extraction methods have been applied to the EEG signal for motor imagery (MI) task classification; among them, common spatial pattern (CSP) is one of the most used feature extraction algorithms [14]. The main concept of the CSP method is to employ the optimal spatial filter on the training EEG datasets, which produces the weight matrix for each electrode and measures the electrode information's significance. Later, researchers replaced the spatial pattern of the CSP with common patterns such as frequency domain, time domain, or combined time-frequency domain to produce the effective features for the MI-based EEG signals [15]. The primary issue with these methods is that they employ Common Spatial Pattern (CSP) on a broad frequency range, such as 1–30 Hz. Due to the intricate nature of the EEG signal, narrow-band signals perform better than full-band frequencies. Researchers have proposed that the EEG signal is composed of various types of rhythms and bands, such as delta, theta, alpha, beta, gamma, and mu. Among these, alpha, beta, and gamma exhibit significant rhythmic properties of the EEG signals [16–18]. Luo et al. first applied a subband-based feature extraction technique with the CSP to include the narrow-band rhythmic properties in the system [19]. The primary issue with this study is that it has increased the computational complexity exponentially due to the multiband increase in the number of signals, which is virtually n times. Additionally, initially, researchers collected the imagination data with a minimum number of electrodes, which could be 1, 2, or 3. However, recently researchers have collected signals with many electrodes, creating a challenging situation for implementing a portable, inexpensive, and fast BCI system for daily activities. Furthermore, this large amount of electrode information produces redundant and noisy data, which adds significant computational complexity [20]. So the feature selection procedure is inevitable for the EEG-based MI classification task; however, no one used the following work [16–19].

As we said, the multiband processed features have been extracted from the individual band and combined to produce the final features; thus, it derived a very higher dimension [21] and it affected the classification algorithm by reducing the performance [22]. Various kinds of supervised and unsupervised feature selection algorithms are available in data science and other machine learning-related research domains [23]. Molla et al. employed a supervised-based feature selection algorithm, neighbourhood component analysis (NCA). They extracted spatial features by using the CSP and then combined the four band features, resulting in a large dimension of features. Finally, they used NCA to select the potential number of features that are less than or equal to 50% of the original feature. The main drawback of their concept is that they selected the feature based on the weighted value and less than or equal to 50%, which may result in difficulties in producing high performance because of the inefficiency of the feature.

To overcome the problems mentioned above, we proposed CFS, mRMR, and SRCFS feature selection approaches along with the SVM, LDA, and MLP classifiers where the LDA and SRCFS-based MI tasks classification system outperforms using EEG signals. The main idea of the SRCFS method is to divide the features into multi-subspace and produce a Laplacian score, which is considered a weight value for each channel using the multi k nearest neighbour technique. Based on the Laplacian score, we selected features from 50% of the original number of features here. We have also implemented the traditional feature selection methods such as f-test, random forest, and logistic lasso and it is proved that our proposed system is far better than the traditional methods.

2. Related Works

There are numerous studies that have been conducted to develop MI classification systems based on the EEG signal. In the year 1875, the first EEG signal was collected by Richard Caton from the animal brain, and later, in 1929, the EEG signal was collected from the human brain first by Hans Berger [24]. Recently, steady-state visual-evoked potential (SSVEP)-based BCI has been developed to assist paralyzed patients by recognizing SSVEP-based commands [25]. EEG mainly records the biological electrical activity of the human brain using many electrodes that are essential for many human-oriented applications to make life easier, especially for people with complete paralysis or extreme disability [26].

To classify the EEG-based classification, Pfurtscheller et al. first applied LDA with adaptive autoregressive (AAR) for classifying left- and right-hand MI-EEG [27]. Many researchers have employed the common spatial pattern (CSP) as an optimal spatial filter to extract a weighted score of each electrode based on a significant score that proves the importance of each electrode [17,18]. The main drawback of these methods is that they consider only a broader range of frequencies in EEG signals, but a narrow signal is more effective compared to a broader signal. Usually, researchers divide the broader EEG signal into different subbands, namely mu, beta, alpha, beta, and gamma rhythm [28]. Pfurtscheller et al. showed that narrowband frequency, specifically the mu and beta rhythms, contain essential information for voluntary movement, and these two rhythms should be considered when implementing the EEG-based MI task classification [16,29]. There are many methodologies that have been proposed for considering each narrow band rhythm such as subband CSP, discriminant filter bank with CSP [30,31], sparse filter-band CSP, and filter bank CSP [21]. However, combining multiband features into a feature yields a large feature vector size, increasing the computational complexity and reducing the system's performance.

To solve the problem, it is inevitable to reduce feature dimension and size to improve performance. Both supervised, and unsupervised algorithms are mainly used to select the effective feature from the large feature dimension [22]. All features in the feature vector might not be relevant and important for the MI task classification, which can be considered a garble for the classification algorithm and degrades the method's performance [32]. Molla et al. divided the EEG signal into multiple sub-bands and then extracted features from each subband, producing large feature dimensions. Lastly, they employed Graph Eigen Decomposition (GED) to reduce the dimensionality of the feature vector to improve

the performance and achieved 99.39% accuracy for epileptic seizure detection [33]. Siuly et al. proposed a Logistic Regression with a cross-correlation technique for classifying the EEG-based MI tasks [34]. In the procedure, they first extracted features with the CSP and then reduced the feature dimension with the hybrid unsupervised feature selection technique. Ali et al. proposed a CSP approach to extract the feature and then rank that feature with the mutual information score. Finally, they applied LDA to classify the MI task and achieved good performance [35]. Kevrich et al. applied empirical mode decomposition, wavelet packet decomposition, and discrete wavelet transforms to generate the narrowband of the EEG signal from a broader frequency [36]. They converted the feature vector into a group of features to justify the performance of the specific set of features. Finally, they claimed that the multiscale principal component analysis (PCA) feature achieved better performance accuracy, which was produced by the highest averaging technique.

Siuly et al. employed an updated CC-LR algorithm to improve the MI tasks classification accuracy where they focused on the specific electrode features and evaluated their method with the BCI III dataset [37]. Song et al. applied a supervised feature selection algorithm that included regression and classification as a unified framework [38]. Goldberger et al. employed a supervised-based neighbourhood component analysis (NCA) feature selection algorithm [39].

Chen et al. proposed a feature selection approach called conditional covariance minimization (CCM) which employs kernel-based measures of independence to find a subset of covariates that is maximally predictive of the response. They carried out numerous experiments using synthetic and real-world data and found that it outperforms other state-of-the-art approaches including Minimum Redundancy Maximum Relevance (mRMR), Backward Elimination Hilbert-Schmidt Independence Criterion (BAHSIC), and Mutual Information (MI) [40]. Constantinopoulos et al. presented a Bayesian method for mixture model training that addresses the feature selection and the model selection problems at the same time. This approach combines a mixture model formulation considering the saliency of the features and a Bayesian approach to mixture learning that can automatically determine the number of components and the saliency of features. Authors proved that this algorithm outperforms the MML-based approaches [41]. A deep learning-based method—Graph Convolutional Network Feature Selector (GRACES) has been implemented to select important features for the high-dimensional and low-sample size (HDLSS) data in [42]. Chen et al. demonstrated empirical evidence that GRACES can achieve a superb and stable performance on both synthetic and real-world HDLSS datasets by utilizing GCN along with different overfitting-reducing strategies including multiple dropouts, the introduction of Gaussian noises, and F-correction.

Molla et al. employed a CSP feature extraction approach and then used a nearest-neighbour-based discriminative features selection method to select the potential feature and discard the garbage feature to improve MI classification using multichannel EEG signal [43]. Finally, they applied a machine learning algorithm SVM and evaluated their method with the BCI Competition III dataset IIIB, and IVA obtained superior performance compared to the recently developed algorithms. Based on their algorithm, they selected 50% of the feature from the extracted feature. To overcome the lacking, we proposed an unsupervised-based sequential feature selection algorithm, which is able to achieve higher accuracy than the existing performance available in the literature.

3. Dataset Description

To evaluate our model, we used two benchmark datasets for MI classification. These are BCI Competition III Dataset IVA, and BCI Competition III Dataset IIIB are described in Sections 3.1 and 3.2 consequently.

3.1. BCI Competition III Dataset IVA

In this study, we consider conducting experiments using publicly available MI data, which is available online with a detailed description that can be found at [44]. This recorded

signal was collected from 5 healthy people, namely aa, al, av, aw, and ay where 118 EEG electrodes were used to record the signal. Each person performed four tasks which are considered here MI tasks, namely right foot, right hand, left hand, and limb. In this study, we have considered only binary classifications, which are left and right classes. The electrodes are placed on the scalp of the subject by following the instruction of the international 10–20 system. The subject is in a relaxed mode during the signal recording, and the subject is asked to imagine specific motor imagery tasks: left and right-hand movements. Each trial is recorded in intervals of 1.25 s to 2.25 s. The recorded signals were filtered with a filter, namely a bandpass filter in the frequency range from 0.05 Hz to 200 Hz, and digitized at 1000 Hz with 16-bit precision. After that, the filtered signal is downsampled at 100 Hz and used in the experiment for the duration of 0.5 s to 3 s in each cue.

3.2. BCI Competition III Dataset IIIB

Another dataset we used here to evaluate our model, BCI competition III dataset IIIB is recorded from the three subjects, namely O3, S4, and X11. This dataset was recorded with the three electrodes which are placed on the subject scalp based on the international 10–20 system. A trial signal consists of a seven-second duration recorded signal. Different trials are collected from the different subjects, such as 320 trials collected from the O3 subjects, and 1080 trials collected from S4 and X11, respectively. This recorded signal was sampled with a ratio of 125 Hz then it was filtered with a notch filter in the range of 0.5 to 30 Hz [45]. Since the experiment was conducted in the virtual reality (VR) paradigm for the O3 subject, we have discarded this subject for performance evaluation of our proposed method (see the Figure 1).

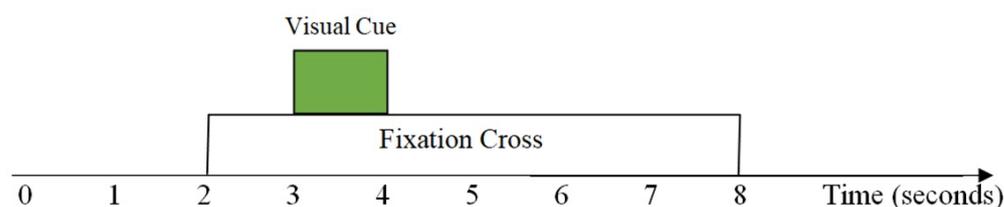


Figure 1. The timing sequence of BCI experiments when only the MI section from each dataset is used.

4. Proposed Method

The working flow architecture of the proposed method is given below in Figure 2, where we included the key contributions of this research and the implementation sequence of the study.

- Step-1** Preprocessing of multichannel EEG signal
- Step-2** Decompose each trial of EEG signal into subbands through filter bank analysis
- Step-3** Extract the spatial from each subband by applying CSP
- Step-4** Combine the features obtained from the individual subband to derive a feature vector
- Step-5** Potential features are selected with feature selection algorithms named CFS, mRMR, and SRCFS, which are used as the final reduced feature vector for the classifier
- Step-6** SVM, LDA and MLP classifiers are employed for the reduced features to distinguish the activities of MI EEG signals

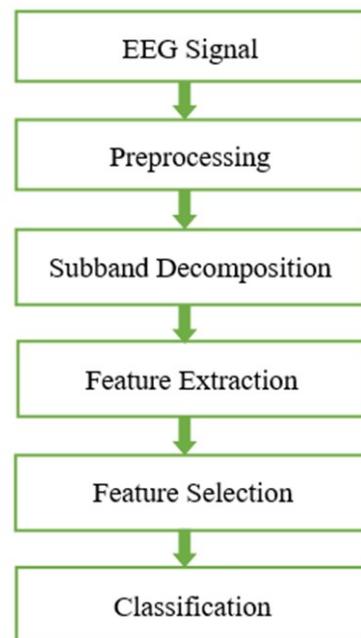


Figure 2. Working flow architecture of the proposed study.

4.1. Preprocessing

We applied a bandpass filter to remove noise from the raw EEG signal because raw EEG usually consists of different kinds of artefacts like eye blinking, sudden sound, muscle movement, body movement, environmental noises, etc. Furthermore, some narrowband EEG signal components are more sensitive to specific MI tasks. As a result, it is not surprising that using sub-bands rather than the entire EEG bandwidth results in more accurate MI task classification. According to a related study, the majority of brain activity associated with MI tasks occurs between 7 Hz and 36 Hz [46,47]. This study divides the broader 8–35 Hz frequency band EEG signal into multiple narrowband signals to calculate the exact feature information of the EEG signal. We have mainly decomposed the signal into four equivalent narrowband signals, namely Mu-band (8–13 Hz), low-beta (13–22 Hz), high-beta (22–35 Hz), and full-band (8–35 Hz) for our experimental purposes [43].

4.2. Feature Extraction

In this study, to extract the effective features from the narrowband signals, we have employed a well-known feature extraction method in multichannel EEG-based BCI the CSP [14,48,49]. The main concept of the algorithm is to minimize the variance among the intra-class features and maximize the variance among the inter-class. In addition, the CSP method finally projects the high-dimensional data into a low dimension, which is known as spatial feature subspace, by using a projection matrix. We have used the CSP algorithm as a spatial filter for making high-variance features between the right-hand and right-foot classes, resulting in peak variances between those classes. Let $E_{c_1}^i$ and $E_{c_2}^i$ be EEG signal of i^{th} trial, c_1 and c_2 represent the class 1 and class 2. The projection matrix W_{CSP} is computed by first calculating the normalized spatial covariance matrix for both classes as follows in Equations (1) and (2).

$$C_L = \frac{E_{c_1} E_{c_1}'}{\text{trace}(E_{c_1} E_{c_1}')} \quad (1)$$

$$C_R = \frac{E_{c_2} E_{c_2}'}{\text{trace}(E_{c_2} E_{c_2}')} \quad (2)$$

where E' is the transpose of E . The averaged normalized covariances \bar{C}_L and \bar{C}_R are then computed by averaging all segments within each class. Equation (3) denotes the total composite spatial covariance.

$$C_c = \bar{C}_L + \bar{C}_R \tag{3}$$

The following is the factorization of this covariance matrix into its eigenvalues and eigenvectors.

$$C_c = U_c \lambda_c U_c' \tag{4}$$

Here, the eigenvector matrix and diagonal eigenvalue matrix are denoted by U_c and λ_c , respectively, which are organized in descending order. Following the above formula, we can calculate the whitening transformation using the following Equation (5).

$$P = \sqrt{\lambda_c^{-1}} U_c' \tag{5}$$

where whitening transformation is denoted by P . The covariance matrices of the two classes are transformed by Equation (5). The projection matrix W_{CSP} is defined by

$$W_{CSP} = P' B = [w_1, w_2, \dots, w_{(ch-1)} w_{ch}] \in R^{(ch \times ch)} \tag{6}$$

where ch is the channel and B is an orthonormal matrix.

A matrix $W_{CSP} = [w_1, w_2, \dots, w_{2m}] \in R^{(2 \times k)}$, including the spatial filters, represents k largest and smallest eigenvalues formed by the eigenvectors by solving the Equation (6). The final feature can be written as $f = [f_1, f_2, \dots, f_{2k}]$.

$$f_j = \log(\text{var}(W'_{CSP} E)), j = 1, 2, \dots, 2k \tag{7}$$

Here, variance is represented by $\text{var}(\cdot)$, and \log transformation is used for normalizing the elements of f_j .

4.3. Feature Selection

Since EEG signals are complex and collected using multiple electrodes, they often contain irrelevant information. Discarding such information is one of the most crucial steps in BCI. Features have a direct impact on how well a BCI system performs, and recent studies have focused on improving currently used methods or creating new ones. The extracted multiband feature dimensions are large and contain less effective features, which is not helpful for classification and increases computational complexity, resulting in reduced performance. In fact, machine learning algorithm performance is typically diminished by specific features. Feature selection techniques are divided into two groups: filter approaches and wrapper approaches [37]. Feature selection techniques can be divided into two groups: filter approaches and wrapper approaches. Filter approaches rely on predetermined criteria and are independent of the learning criteria. They create subsets that are assessed using a search algorithm. Wrapper approaches, on the other hand, require the use of a learning algorithm, and the performance of the selected feature subsets is evaluated using this algorithm.

In this study, we investigated three feature selection approaches: CFS, mRMR, and SRCFS. These methods have been recently developed and successfully applied in MI classification. We found that SRCFS outperformed the other two methods in terms of classification accuracy. In addition, the HSIC Lasso [50] and three conventional feature selection schemes named f-test, random forests, and logistic lasso have been investigated to evaluate the performance of our proposed system.

4.3.1. Correlation-Based Feature Selection (CFS)

The working idea of the CFS algorithm is to calculate a subset of the feature by following the initial hypothesis, which is mainly correlated with the output classes not correlated with themselves [22]. The usefulness of the features in class prediction and their connection with other features serve as the validation criteria. The subset calculation process of this algorithm can be written as the following formula,

$$CFS_s = \frac{f(\bar{r}_{tq})}{\sqrt{f + f(f - 1)\bar{r}_{qq}}} \tag{8}$$

Here, the mean of correlation among the inter-class and the mean of correlation among the intra-class are denoted by \bar{r}_{tq} and \bar{r}_{qq} , respectively. In addition, the heuristic merit of each subset is denoted by f . The denominator measures the degree of redundancy among the features that make up the feature subset, and the numerator measures how predictive the feature subset is. The technique thus detects aspects that are superfluous or redundant. The search algorithm we utilized included backward exclusion and forward selection, and it was called Best First.

4.3.2. Minimum Redundancy and Maximum Relevance (mRMR)

A heuristic resembling CFS is used by the lowest redundancy and maximum relevance algorithm. The metric employed in this instance to verify the significance of the features is mutual information, which leads to a ranking of the features based on how well they cooperate with other features and the class. The most pertinent feature shares the least mutual information with the other features and the most with the class. This is achieved by increasing the value of the following expression,

$$F_{mRMR} = \frac{\frac{1}{n_f} \sum I(c, f)}{\frac{1}{n_f^2} \sum I(f_1, f_2)}. \tag{9}$$

Here, the number of features, the mutual information between two classes, and the mutual information between two features are denoted by n_f , $\sum I(c, f)$, and $\sum I(f_1, f_2)$, respectively. After the ranking phase, this approach creates a subset with a varying number of features and orders it with the ranking score [51]. The machine learning algorithms finally validate these feature groups based on the ranking score.

4.3.3. Multi-Subspace Randomization and Collaboration-Based Unsupervised Feature Selection (SRCFS)

The SRCFS is a powerful framework for unsupervised feature selection in huge datasets where this algorithm conceals the original high-dimensional feature in several sub-groups [38,52]. Primarily, this algorithm creates a huge number of random subgroup features and after scoring each subspace it concatenated all the subgroups into a single feature vector based on the score of each group. Suppose, the feature partition variable is denoted with $F^{(i)}$ for the i^{th} basic feature partition, and random subspace for j^{th} position can be denoted with $F^{(i,j)}$ of the $F^{(i)}$ partition. We can express the feature partition formula according to the following Equation (10). Then $F^{(i)}$ can be represented as follows,

$$F^{(i)} = \{F^{(i,1)}, F^{(i,2)}, \dots, F^{(i,z)}\}. \tag{10}$$

Here, $F^{(i)}$, and $F^{(i,1)}$ denote the feature partition and subspace in the partition, respectively. The quantity of random subspaces in F is given by z where an ideal condition would be for all subspaces to have the same size because the three must be equal to all random subspaces. Individual feature partition is created repeatedly, which can form

a composed feature which is known as a final feature F and can be expressed with the following Equation (11).

$$F = \{F^{(1)}, F^{(2)}, \dots, F^{(g)}\}. \quad (11)$$

The total number of basic partitions and i^{th} basic partition are denoted with g and $F^{(i)}$, respectively. In each partition, there is an unknown number of subspaces which can be denoted with $g.z$, but the number of subspaces in each partition must be equal. It actually calculates g number of Laplacian scores where every partition must produce an individual score, which produced a final Laplacian score vector. The average Laplacian score can be calculated using the following formula which is the average of the Laplacian score for the basic partition F .

$$L_z(f) = \frac{1}{g} \sum_{i=1}^g L_s(F) \quad (12)$$

Here, $L_z(f) \in \mathbb{R}$ represents the full Laplacian score vector that be obtained by concatenating the Laplacian score vectors for all of its z random subspaces. To reflect the structure information of all $g.z$ numbers of random subspaces, we build $g.z$ numbers of KNN graphs. The combining information of the KNN Graph and the local preserving power of each subspace can lastly be used to compute the main score which is used to rank the feature and selected potential features called Laplacian scores of the features in each subspace.

4.4. Classification Using LDA, SVM and MLP

In this study, we used three well-known and mature machine learning-based classification algorithms, namely LDA, SVM, and MLP, to classify the left-hand and right-hand human motor imagery EEG signal. The goal is to find out and evaluate which one can be produced the best outcomes. LDA, also known as the Fisher linear discriminant, is a simple and well-known technique for categorizing BCI data. A linear binary classifier maps a p -dimensional input vector x to a hyperplane that divides the input space into two half spaces, each of which denotes a class (+1 or -1). The SVM is a relatively new classification method developed by Vapnik. It has a strong mathematical base in statistical learning theory and has demonstrated great performance in a variety of practical issues, particularly in BCI. To translate a higher-dimension row of training data, it uses a nonlinear map. Within this new dimension, it looks for the linear optimal dividing hyperplane (also known as a “choice border” separating the tuples of one class from another). A proper nonlinear mapping can always be used to split data from two classes into a suitably large dimension via a hyperplane. Support vectors are used by the SVM to find this hyperplane (“essential” training tuples) and margins (identified by the support vectors). SVM classifier with radial basis function (RBF) kernel is used to assess the proposed technique. A detailed description of these two methods can be found in [53,54]. MLP is a popular machine learning algorithm and a powerful tool for classifying brain activities. The inputs to the MLP are typically features extracted from EEG or other neuroimaging data. These features are then passed through multiple layers of interconnected nodes, with each node performing mathematical calculations on the input data. The output layer of the MLP represents the predicted class label for the input data. During training, the MLP’s weights are changed to minimize the difference between the expected and actual output using techniques such as backpropagation. Their performance, however, is heavily influenced by the quality and significance of the input data, as well as the size and complexity of the MLP architecture [55,56]. The size of the hidden layers used in our experiment is ten.

5. Results and Discussion

To evaluate the model, we used here two well-known publicly available EEG-based MI task datasets. For each of the trials of the dataset, we decomposed into four narrowband signals to extract the exact information contained in the signal. The CSP approach is used

to extract features from each narrow band and combine each feature to produce a final feature vector which generates a high-dimensional feature vector. The discriminative features are chosen using the CFS, mRMR, and SRCFS-based techniques. As a result, the collected features are utilized to train three classifiers, SVM, LDA, and MLP, separately. Then test data are used to assess the performance of the classifiers. Each 2.5 second trial for every person is taken out of the EEG data. Each frequency band is subjected to the CSP in order to extract the spatial information. From each subband, four pairs of spatial filters producing eight features are chosen from dataset BCI III-IVA and two pairs of spatial filters are chosen from BCI III-IIIB. For each trial, 32 (4×8) and 8 (4×2) dimensional feature vectors are created by combining the CSP features collected from each of the four bands from dataset BCI III-IVA and BCI III-IIIB, respectively. The high-dimensional feature space is then subjected to the CFS, mRMR, and SRCFS-based feature selection techniques. They give each feature a weight based on the label of the training data. The features are ranked based on the weights established by each of the feature selection approaches. The number of top-ranked features is chosen for classification.

5.1. Experimental Setting

We evaluated the proposed model with 5-fold cross-validation formula where we took the individual subject dataset feature and randomly divided the feature into five folded. After that, we randomly trained the model with four folded and tested the model with the rest one-fold features and preserved the accuracy for the first fold feature. We repeatedly preserved the accuracy five times and finally, we average the performance score and produced the final average performance score. We computed the accuracy (%) matrix using the following formula, which is also known as a best performance calculation procedure.

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \times 100 \quad (13)$$

where, T_p , T_n , F_p , and F_n represents true positive, true negative, false positive, and false negative, respectively. The accuracy values from the several experiments conducted mainly show the effectiveness of the proposed approach. Two different feature selection methods CFS, and mRMR have been employed and the result was compared with the SRCFS-based feature selection method. To evaluate the classifier performance, SVM and MLP are employed along with LDA. We have also calculated some statistical performance metrics like AUROC, F1 scores, and computational time of different subjects on two datasets to ensure the robustness and effectiveness of the proposed approach.

5.2. Performance Result with BCI Competition III Dataset IVA

Figure 3 demonstrates the performance comparison of different feature selection methods where SVM, LDA, and MLP are used, respectively. These figures proved that the SRCFS feature selection method's performance is better in most cases than others.

Figure 4 demonstrates that SRCFS with LDA outperforms the other for dataset BCI III-IVA. The result also showed that the feature selection technique has certain benefits in terms of enhancing classification performance. Without feature selection, the mean accuracy (across all subjects) is substantially lower than the other approaches that use feature selection methods. The method without feature selection uses extra features that are irrelevant and lowers the classifier's performance as a result.

Figure 5 compares the accuracy of the proposed method with different combinations of feature selection and classifier as a function of the number of selected features. It has been found that utilizing 16 well-chosen features from dataset BCI III-IVA enables the classification of objects with the highest degree of accuracy.

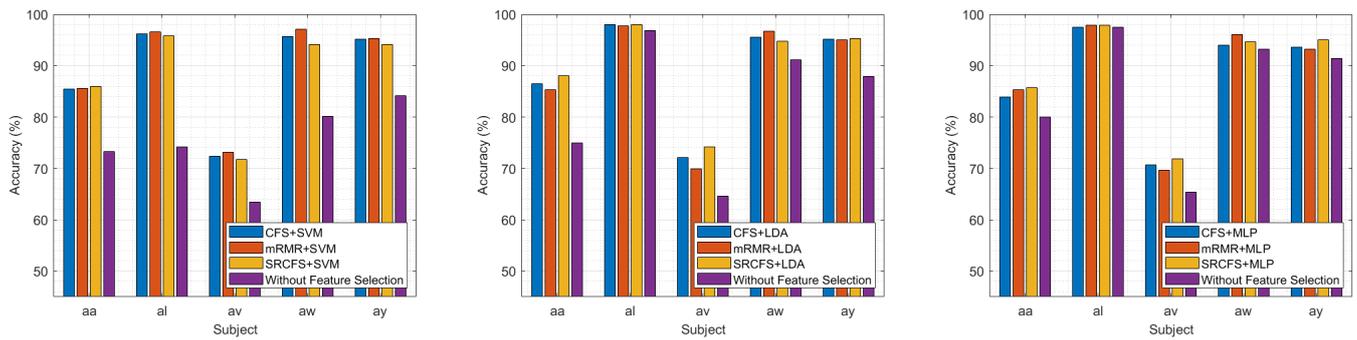


Figure 3. The motor imagery (MI) classification performance comparison among CFS, mRMR, SRCFS feature selection methods and without feature selection. The left, middle, and right subplots represent the accuracies of different subjects for the BCI III-IVA dataset, where SVM, LDA, and MLP classifier has been used, respectively.

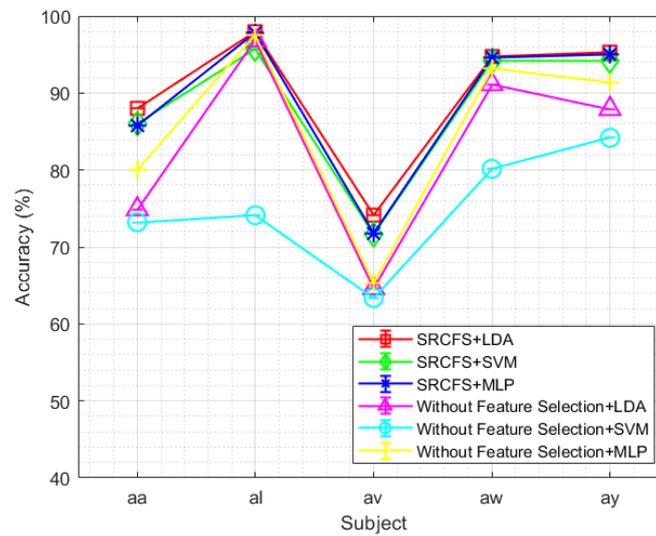


Figure 4. The motor imagery classification performance comparison between LDA, SVM, and MLP classifier using SRCFS feature selection method and without feature selection. The figure represents the accuracies of different subjects for the BCI III-IVA dataset.

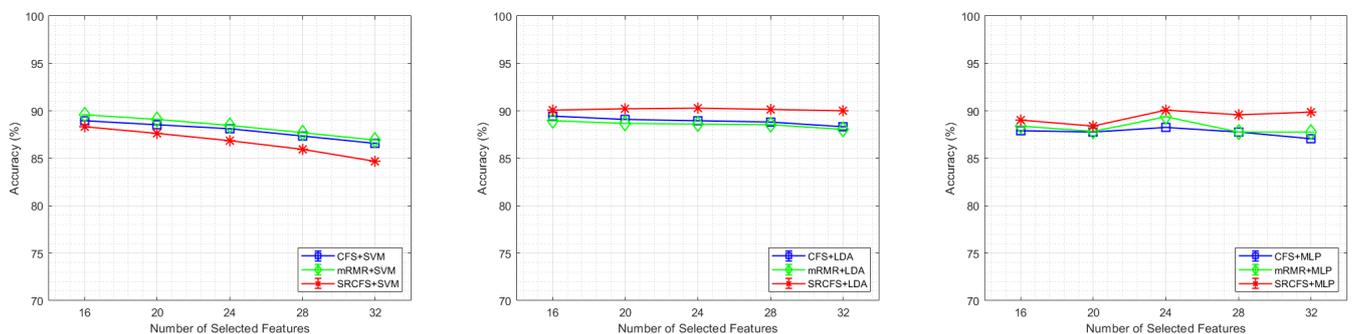


Figure 5. The motor imagery classification performance comparison using CFS, mRMR, and SRCFS feature selection methods with SVM, LDA, and MLP classifiers for different numbers of selected features. The left, middle, and right subplots represent the accuracies of the BCI III-IVA dataset for different numbers of features (50% to 100%) selected by the feature selection algorithm where SVM, LDA and MLP classifiers have been used.

5.3. Performance Result with BCI Competition III Dataset IIIB

Figure 6 demonstrates the performance comparison of different feature selection methods with SVM, LDA, and MLP classifiers, respectively. These figures show that the SRCFCS feature selection method’s performance is stable. This dataset has been used to verify the extensive generalizability property of our proposed method.

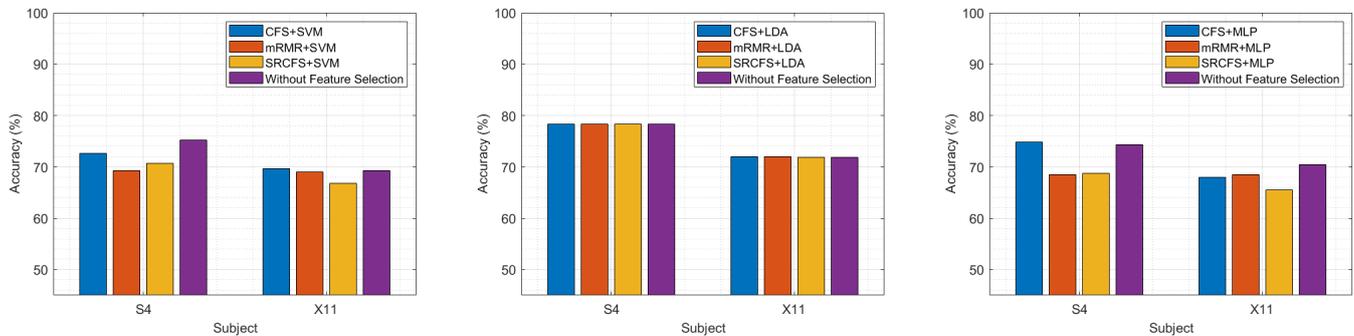


Figure 6. The motor imagery (MI) classification performance comparison among CFS, mRMR, SRCFCS feature selection methods and without feature selection. The left, middle, and right subplots represent the accuracies of different subjects for the BCI III-III B dataset, where SVM, LDA and MLP classifier has been used, respectively.

Figure 7 demonstrates that without feature selection and SRCFCS-based feature selection have similar accuracy for the dataset BCI III-III B. Due to the fewer number of channels, the dataset BCI-III B produced two pairs of spatial filters resulting in eight features. For low feature dimensions, SRCFCS with LDA can not overcome the accuracy without feature selection. However, selecting features reduces classification complexity.

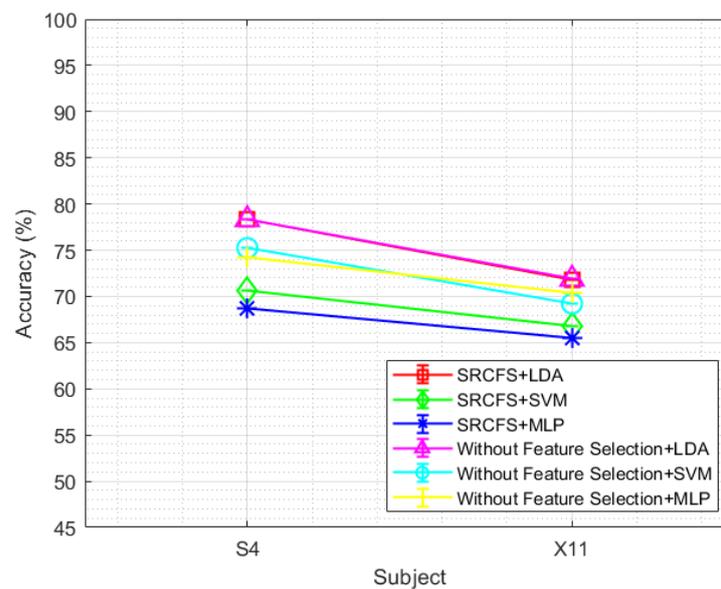


Figure 7. The motor imagery classification performance comparison between SVM, LDA, and MLP classifier using SRCFCS feature selection method and without feature selection. The figure represents the accuracies of different subjects for the BCI III-III B dataset.

Figure 8 compares the accuracy of the proposed method with different combinations of feature selections and classifiers as a function of the number of selected features. It has been found that utilizing four well-chosen features from dataset BCI-III B enables the classification of objects with the highest degree of accuracy.

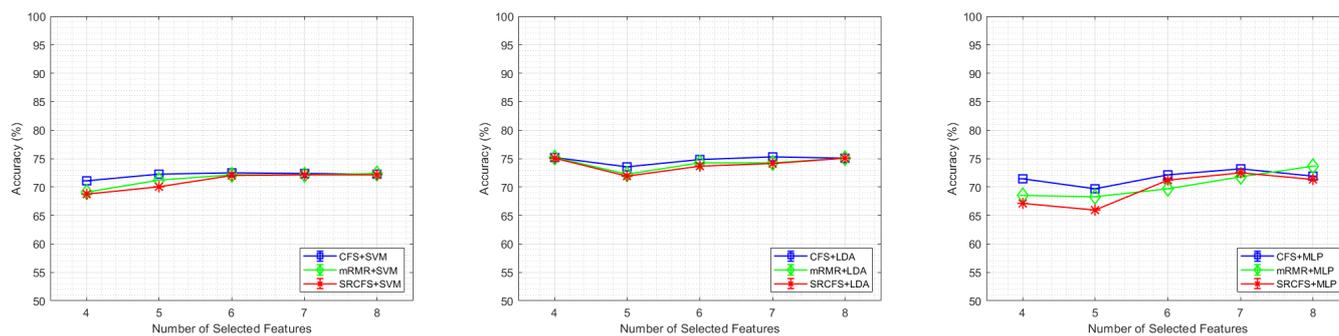


Figure 8. The motor imagery classification performance comparison using CFS, mRMR, and SRCFS feature selection methods with SVM, LDA, and MLP classifiers for different numbers of selected features. The left, middle, and right subplots represent the accuracies of for BCI III-III B dataset for different numbers of features (50% to 100%) selected by the feature selection algorithm where SVM, LDA, and MLP classifiers have been used.

Moreover, different statistical performance evaluation metrics have been calculated to validate the performance of our proposed method. Table 1 demonstrated the state-of-the-art comparison of the proposed model where our study achieved superiority over the competitive models. In addition, Tables 2–4 show the performance of the area under the ROC, F1 score, and computational time, respectively, of different subjects on BCI competition III dataset IVA. On the other hand, Table 5 shows the performance of AUROC, F1 score, and computational time, respectively, of different subjects on BCI competition III dataset IIIB. Here, the computational time is measured in seconds (s) and it represents the time required for training and classification of a single fold required by the classifier in a five-fold cross-validation technique. Moreover, some traditional feature selection methods like f test, random forests, and logistic lasso have also been studied. But, the performance of these methods is not further compared because of their high computational cost and low MI recognition rate. In addition, they are rarely used for MI task classification in BCI.

Furthermore, we have tested another feature selection technique named HSIC Lasso for 07 (seven) different kernels with LDA classifier using BCI competition III dataset IVA and IIIB. Since the LDA classifier performed best for our proposed method and other studied methods, we have considered this classifier for testing HSIC Lasso feature selection method in terms of AUROC, F1 score, computational time, and accuracy performance metrics. From our experimental results, it is shown that the performance of the HSIC Lasso with the best kernel ADMM is almost similar to mRMR for BCI Competition III dataset IVA, the accuracy of both HSIC Lasso and mRMR is 88.93 % and the performance of the mRMR is better than HSIC Lasso for BCI Competition III dataset IIIB, the accuracy of the mRMR and HSIC Lasso are 75.17% and 69.91%, respectively, in this case. Since the overall performance of HSIC Lasso is almost similar on BCI Competition III dataset IVA and slightly lower on BCI Competition III dataset IIIB compared to the proposed and other studied feature selection methods, the performance of this method is not further compared with others.

5.4. State of the Art Comparison with Previous Methods

Table 1 compares and contrasts the suggested method's classification accuracy results with those of recently developed algorithms. The proposed method's overall average classification accuracy is 90.05%. The performance of the proposed method is compared with the methods CSP-R-MF [57], R-MDRM [58], MKELM [59], and so on. It is observed that the average classification accuracy of the proposed method outperforms the other recently developed algorithm, as shown in Table 1. Table 1 demonstrated that for subjects aa, aw, and ay, the proposed method achieved the best performance.

Table 1. Performance comparison in terms of classification accuracy on BCI competition III dataset IVA of the proposed method with state-of-the-art works. The highest accuracy is marked in boldface.

Studies	Methods	Subjects					Mean ± SD
		aa	al	av	aw	ay	
Belwafi et al. [32]	WOLA-CSP	66.07	96.07	52.14	71.43	50.00	67.29
Dai et al. [38]	TKCSP	68.10	93.88	68.47	88.40	74.93	79.17
She et al. [39]	H-ELM	63.39	98.39	64.08	85.67	85.16	79.33
Park et al.[60]	SSS-CSP	74.11	100	67.78	90.07	89.29	84.46
Jian et al. [57]	CSP-R-MF	81.43	92.41	70.00	83.57	85.00	82.48
Selim et al. [61]	AM-BA-SVM	86.61	100	66.84	90.63	80.95	85.00
Singh et al. [58]	SR-MDRM	79.46	100	73.46	89.28	88.49	86.13
Zhang et al. [59]	MKELM	83.30	98.50	71.40	91.30	93.30	87.50
Singh et al. [62]	R-MDRM	81.25	100	76.53	87.05	91.26	87.21
Proposed Method	SRCFS + LDA	88.03	97.98	74.17	94.76	95.31	90.05 ± 9.60

Table 2. Performance of different studied methods in terms of area under the receiver operating characteristic curve (AUROC) on BCI competition III dataset IVA for each of the five subjects, the best result is marked in boldface.

Feature Selection Methods and Classifiers	AUROC				
	aa	al	av	aw	ay
CFS + SVM	0.9306	0.9922	0.8297	0.9836	0.9826
mRMR + SVM	0.9205	0.9936	0.7916	0.9921	0.9796
SRCFS + SVM	0.9242	0.9881	0.7513	0.9717	0.9823
CFS + LDA	0.9030	0.9911	0.7743	0.9914	0.9821
mRMR + LDA	0.9363	0.9968	0.7530	0.9929	0.9838
SRCFS + LDA	0.9356	0.9918	0.8072	0.9905	0.9861
CFS + MLP	0.9135	0.9944	0.8115	0.9802	0.9731
mRMR + MLP	0.9192	0.9892	0.7664	0.9795	0.9621
SRCFS + MLP	0.9263	0.9972	0.8137	0.9864	0.9844

Table 3. Performance of different studied methods in terms of F1 score on BCI competition III dataset IVA for each of the five subjects, the best result is marked in boldface.

Feature Selection Methods and Classifiers	F1 Score				
	aa	al	av	aw	ay
CFS + SVM	0.8593	0.9638	0.7287	0.9534	0.9534
mRMR + SVM	0.8364	0.9712	0.7015	0.9606	0.9568
SRCFS + SVM	0.8571	0.9562	0.6512	0.9187	0.9391
CFS + LDA	0.8470	0.9825	0.7000	0.9568	0.9373
mRMR + LDA	0.8582	0.9788	0.7254	0.9677	0.9373
SRCFS + LDA	0.8633	0.9789	0.7317	0.9496	0.9489
CFS + MLP	0.8443	0.9753	0.7092	0.9386	0.9353
mRMR + MLP	0.8520	0.9788	0.6886	0.9603	0.9304
SRCFS + MLP	0.8592	0.9787	0.7285	0.9458	0.9500

From Tables 1–4, it is clearly depicted that the SRCFS and LDA-based MI tasks classification system is robust and effective in terms of the performance metrics: accuracy, AUROC, F1 score, and computational time for BCI competition III dataset IVA. On the other hand, Table 5 shows that the computational time of the SRCFS and LDA-based system is low compared to others for the BCI competition III dataset IIIB dataset. It is also observed that the MLP classifier is more computationally costly than the others. From the above discussion, we can conclude that the SRCFS feature selection method with LDA classifier is undoubtedly a robust and effective system for MI tasks classification using EEG signal.

Table 4. Performance of different studied methods in terms of computational time on BCI competition III dataset IVA for each of the five subjects, the best result is marked in boldface.

Feature Selection Methods and Classifiers	Computational Time (s)				
	aa	al	av	aw	ay
CFS + SVM	0.1804	0.0133	0.0085	0.0082	0.0090
mRMR + SVM	0.1582	0.0137	0.0091	0.0087	0.0089
SRCFS + SVM	0.1573	0.0154	0.0087	0.0088	0.0087
CFS + LDA	0.1857	0.0133	0.0074	0.0101	0.0096
mRMR + LDA	0.1683	0.0127	0.0080	0.0082	0.0077
SRCFS + LDA	0.1642	0.0134	0.0079	0.0075	0.0073
CFS + MLP	0.7557	0.1970	0.1480	0.1629	0.2056
mRMR + MLP	0.8339	0.3850	0.1699	0.1936	0.2653
SRCFS + MLP	0.7235	0.2497	0.1937	0.2041	0.2471

Table 5. Performance of different studied methods in terms of AUROC, F1 score, and computational time (Com. Time) on BCI competition III dataset IIIB for each of the two subjects, the best result is marked in boldface.

Feature Selection Methods and Classifiers	Evaluation Metrics					
	AUROC		F1 Score		Com. Time (s)	
	S4	X11	S4	X11	S4	X11
CFS + SVM	0.8379	0.7567	0.7430	0.6640	0.2093	0.0149
mRMR + SVM	0.7670	0.7488	0.6863	0.6402	0.1819	0.0143
SRCFS + SVM	0.7916	0.7236	0.7188	0.6439	0.1866	0.0147
CFS + LDA	0.7811	0.7638	0.6965	0.6992	0.2040	0.0115
mRMR + LDA	0.7384	0.7481	0.6704	0.6732	0.1860	0.0115
SRCFS + LDA	0.8054	0.7552	0.7431	0.6922	0.1675	0.0108
CFS + MLP	0.8216	0.7482	0.7395	0.6614	0.8545	0.1718
mRMR + MLP	0.7276	0.7298	0.6756	0.6654	0.7804	0.1800
SRCFS + MLP	0.7509	0.7119	0.6922	0.6504	0.8262	0.1709

6. Conclusions

Supervised and Unsupervised feature selection methods are investigated in this paper to classify motor imagery-based EEG signals. The experiment is evaluated using two publicly available BCI Competition III Dataset IVA and BCI Competition III Dataset IIIB. The multichannel EEG signal is decomposed into four subbands. Features are extracted from each subband. Then the extracted features are combined to make a high-dimensional feature vector. Not all features are important for classification. The irrelevant feature may degrade the performance of the system. The performance of the classification is

improved by properly removing redundant and irrelevant characteristics from the feature vector, which increases the feature vector's discriminative power. With the given class label, the unsupervised feature selection outperforms the supervised feature selection, as demonstrated in Table 1. The key benefit of using an unsupervised feature selection method is that each sample of a feature vector does not need to have its labels provided. It chooses features by taking the relationship between feature dimensions into account. It is clear that when the feature selection method has been applied, the accuracy is increased. The combination of features also plays a vital role. As shown in Table 1, the proposed combination of full band and subband signals and the use of the feature selection strategy improve the MI classification accuracy. It can be expanded to include multiclass MI classification issues in the BCI paradigm and we will study more feature selection methods and classifiers in future work.

Author Contributions: Conceptualization, M.H.K., S.M. and A.A.S.; methodology, M.H.K. and S.M.; software, M.H.K. and S.M.; validation, M.H.K. and S.M.; formal analysis, M.H.K., A.S.M.M., J.S. and M.K.I.M.; investigation, M.H.K. and S.M.; data curation, M.H.K., S.M. and A.S.M.M.; writing—original draft preparation, M.H.K., S.M. and A.A.S.; writing—review and editing, M.H.K.; A.S.M.M., J.S. and M.K.I.M.; visualization, M.H.K. and S.M.; supervision, M.K.I.M. and J.S.; funding acquisition, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by the Competitive Research Fund of The University of Aizu, Japan.

Data Availability Statement: The proposed model is evaluated with two benchmark datasets, namely BCI Competition III dataset IVA and dataset IIIB, which are publicly available. The dataset links are provided below, https://www.bbci.de/competition/iii/#data_set_iva; https://www.bbci.de/competition/iii/#data_set_iiib.

Acknowledgments: This paper is a part of a project supported by Bangamata Sheikh Foijilatunnesa Mujib Science & Technology University, Jamalpur 2012, Bangladesh.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

BCI	Brain-Computer Interface
EEG	Electroencephalography
MEG	Magnetoencephalogram
fMRI	Functional Magnetic Response Imaging
MI	Motor Imagery
SSVEP	Steady-State Visual-Evoked Potential
SVM	support vector machines
MLP	Multi-layer Perceptron
LDA	Linear Discriminant Analysis
AAR	Adaptive Autoregressive
CSP	Common Spatial Pattern
NCA	Neighbourhood Component Analysis
PCA	Principal Component Analysis
CSP	Common Spatial Pattern
CFS	Correlation-Based Feature Selection
mRMR	Minimum Redundancy and Maximum Relevance
SRCFS	Multi-Subspace Randomization and Collaboration-Based Unsupervised Feature Selection
GCN	Graph Convolutional Network
GRACES	Graph Convolutional Network Feature Selector

ERD	Event-Related Desynchronization
ERS	Event-Related Synchronization
GED	Graph Eigen Decomposition
HSIC	Hilbert-Schmidt Independence Criterion
Lasso	Least Absolute Shrinkage and Selection Operator
BAHSIC	Backward Elimination Hilbert-Schmidt Independence Criterion
CCM	Conditional Covariance Minimization
MML	Meta Machine Learning
VR	Virtual Reality
HDLSS	High-Dimensional and Low-Sample Size

References

- Molla, M.K.I.; Saha, S.K.; Yasmin, S.; Islam, M.R.; Shin, J. Trial regeneration with subband signals for motor imagery classification in BCI paradigm. *IEEE Access* **2021**, *9*, 7632–7642. [[CrossRef](#)]
- Yang, L.; Song, Y.; Ma, K.; Xie, L. Motor imagery EEG decoding method based on a discriminative feature learning strategy. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 368–379. [[CrossRef](#)] [[PubMed](#)]
- Stegman, P.; Crawford, C.S.; Andujar, M.; Nijholt, A.; Gilbert, J.E. Brain-Computer Interface Software: A Review and Discussion. *IEEE Trans. Hum.-Mach. Syst.* **2020**, *50*, 101–115. [[CrossRef](#)]
- Miah, A.S.M.; Shin, J.; Islam, M.M.; Molla, M.K.I.; Abdullah. Natural Human Emotion Recognition Based on Various Mixed Reality (MR) Games and Electroencephalography (EEG) Signals. In Proceedings of the 2022 IEEE 5th Eurasian Conference on Educational Innovation (ECEI) IEEE, Taipei, Taiwan, 10–12 February 2022; pp. 408–411.
- Miah, A.S.M.; Shin, J.; Hasan, M.A.M.; Molla, M.K.I.; Okuyama, Y.; Tomioka, Y. Movie Oriented Positive Negative Emotion Classification from EEG Signal using Wavelet transformation and Machine learning Approaches. In Proceedings of the 2022 IEEE 15th International Symposium on Embedded Multicore/Many-Core Systems-on-Chip (MCSoc) IEEE, Penang, Malaysia, 19–22 December 2022; pp. 26–31.
- Miah, A.S.M.; Mouly, M.A.; Debnath, C.; Shin, J.; Bari, S.S. Event-Related Potential Classification based on EEG data using xDWT with MDM and KNN. In Proceedings of the Computing Science, Communication and Security: Second International Conference, COMS2 2021, Gujarat, India, 6–7 February 2021; Revised Selected Papers; Springer: Berlin/Heidelberg, Germany, 2021; pp. 112–126.
- Zobaed, T.; Ahmed, S.R.A.; Miah, A.S.M.; Binta, S.M.; Ahmed, M.R.A.; Rashid, M. Real time sleep onset detection from single channel EEG signal using block sample entropy. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *928*, 032021. [[CrossRef](#)]
- Wang, Y.; Nakanishi, M.; Zhang, D. EEG-based brain-computer interfaces. In *Neural Interface: Frontiers and Applications*; Springer: Singapore, 2019; pp. 41–65.
- Sun, B.; Zhang, H.; Wu, Z.; Zhang, Y.; Li, T. Adaptive spatiotemporal graph convolutional networks for motor imagery classification. *IEEE Signal Process. Lett.* **2021**, *28*, 219–223. [[CrossRef](#)]
- Georgiadis, K.; Adamos, D.A.; Nikolopoulos, S.; Laskaris, N.; Kompatsiaris, I. A graph-theoretic sensor-selection scheme for covariance-based Motor Imagery (MI) decoding. In Proceedings of the 2020 28th European Signal Processing Conference (EUSIPCO) IEEE, Amsterdam, The Netherlands, 18–21 January 2021; pp. 1234–1238.
- Akter, M.S.; Islam, M.R.; Tanaka, T.; Iimura, Y.; Mitsuhashi, T.; Sugano, H.; Wang, D.; Molla, M.K.I. Statistical features in high-frequency bands of interictal iEEG work efficiently in identifying the seizure onset zone in patients with focal epilepsy. *Entropy* **2020**, *22*, 1415. [[CrossRef](#)] [[PubMed](#)]
- Nuyujukian, P.; Fan, J.M.; Kao, J.C.; Ryu, S.I.; Shenoy, K.V. A high-performance keyboard neural prosthesis enabled by task optimization. *IEEE Trans. Biomed. Eng.* **2014**, *62*, 21–29. [[CrossRef](#)] [[PubMed](#)]
- Lotte, F.; Bougrain, L.; Cichocki, A.; Clerc, M.; Congedo, M.; Rakotomamonjy, A.; Yger, F. A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update. *J. Neural Eng.* **2018**, *15*, 031005. [[CrossRef](#)]
- Miah, A.S.M.; Islam, M.R.; Molla, M.K.I. EEG classification for MI-BCI using CSP with averaging covariance matrices: An experimental study. In Proceedings of the 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2) IEEE, Rajshahi, Bangladesh, 11–12 July 2019; pp. 1–5.
- Higashi, H.; Tanaka, T. Common spatio-time-frequency patterns for motor imagery-based brain machine interfaces. *Comput. Intell. Neurosci.* **2013**, *2013*, 8. [[CrossRef](#)]
- McFarland, D.J.; Miner, L.A.; Vaughan, T.M.; Wolpaw, J.R. Mu and beta rhythm topographies during motor imagery and actual movements. *Brain Topogr.* **2000**, *12*, 177–186. [[CrossRef](#)]
- Dornhege, G.; Blankertz, B.; Curio, G.; Muller, K.R. Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 993–1002. [[CrossRef](#)] [[PubMed](#)]
- Ramoser, H.; Muller-Gerking, J.; Pfurtscheller, G. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.* **2000**, *8*, 441–446. [[CrossRef](#)] [[PubMed](#)]
- Luo, J.; Wang, J.; Xu, R.; Xu, K. Class discrepancy-guided sub-band filter-based common spatial pattern for motor imagery classification. *J. Neurosci. Methods* **2019**, *323*, 98–107. [[CrossRef](#)] [[PubMed](#)]

20. Udhaya Kumar, S.; Hannah Inbarani, H. PSO-based feature selection and neighborhood rough set-based classification for BCI multiclass motor imagery task. *Neural Comput. Appl.* **2017**, *28*, 3239–3258. [[CrossRef](#)]
21. Dy, J.G.; Brodley, C.E. Feature selection for unsupervised learning. *J. Mach. Learn. Res.* **2004**, *5*, 845–889.
22. Song, L.; Smola, A.; Gretton, A.; Borgwardt, K.M.; Bedo, J. Supervised feature selection via dependence estimation. In Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007; pp. 823–830.
23. Goldberger, J.; Hinton, G.E.; Roweis, S.; Salakhutdinov, R.R. Neighbourhood components analysis. In *Advances in Neural Information Processing Systems 17*; NeurIPS: San Diego, CA, USA, 2004.
24. Zifkin, B.G.; Avanzini, G. Clinical neurophysiology with special reference to the electroencephalogram. *Epilepsia* **2009**, *50*, 30–38. [[CrossRef](#)]
25. Mahmood, S.; Shin, J.; Farhana, I.; Islam, M.R.; Molla, M.K.I. Frequency Recognition of Short-Time SSVEP Signal Using CORRCA-Based Spatio-Spectral Feature Fusion Framework. *IEEE Access* **2021**, *9*, 167744–167755. [[CrossRef](#)]
26. Wolpaw, J.R.; Birbaumer, N.; Heetderks, W.J.; McFarland, D.J.; Peckham, P.H.; Schalk, G.; Donchin, E.; Quatrano, L.A.; Robinson, C.J.; Vaughan, T.M.; et al. Brain-computer interface technology: A review of the first international meeting. *IEEE Trans. Rehabil. Eng.* **2000**, *8*, 164–173. [[CrossRef](#)]
27. Pfurtscheller, G.; Neuper, C.; Schlogl, A.; Lugger, K. Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Trans. Rehabil. Eng.* **1998**, *6*, 316–325. [[CrossRef](#)]
28. Joy, M.M.H.; Hasan, M.; Miah, A.S.M.; Ahmed, A.; Tohfa, S.A.; Bhuiyan, M.F.I.; Zannat, A.; Rashid, M.M. Multiclass MI-Task Classification Using Logistic Regression and Filter Bank Common Spatial Patterns. In Proceedings of the Computing Science, Communication and Security, Gujarat, India, 26–27 March 2020; pp. 160–170.
29. Pfurtscheller, G.; Da Silva, F.L. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* **1999**, *110*, 1842–1857. [[CrossRef](#)]
30. Pfurtscheller, G.; Pregenzer, M.; Neuper, C. Visualization of sensorimotor areas involved in preparation for hand movement based on classification of μ and central β rhythms in single EEG trials in man. *Neurosci. Lett.* **1994**, *181*, 43–46. [[CrossRef](#)] [[PubMed](#)]
31. Ang, K.K.; Chin, Z.Y.; Zhang, H.; Guan, C. Filter bank common spatial pattern (FBCSP) in brain-computer interface. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence) IEEE, Padua, Italy, 18–23 July 2008; pp. 2390–2397.
32. Belwafi, K.; Romain, O.; Gannouni, S.; Ghaffari, F.; Djemal, R.; Ouni, B. An embedded implementation based on adaptive filter bank for brain-computer interface systems. *J. Neurosci. Methods* **2018**, *305*, 1–16. [[CrossRef](#)] [[PubMed](#)]
33. Molla, M.K.I.; Hassan, K.M.; Islam, M.R.; Tanaka, T. Graph eigen decomposition-based feature-selection method for epileptic seizure detection using electroencephalography. *Sensors* **2020**, *20*, 4639. [[CrossRef](#)] [[PubMed](#)]
34. Siuly; Li, Y.; Wen, P. Identification of motor imagery tasks through CC-LR algorithm in brain computer interface. *Int. J. Bioinform. Res. Appl.* **2013**, *9*, 156–172. [[CrossRef](#)]
35. Ali, S.; Ferdous, J.; Hamid, E.; Molla, K.I. A novel features selection approach with common spatial pattern for EEG based brain-computer interface implementation. *IETE J. Res.* **2022**, *68*, 1757–1771. [[CrossRef](#)]
36. Kevric, J.; Subasi, A. Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system. *Biomed. Signal Process. Control* **2017**, *31*, 398–406. [[CrossRef](#)]
37. Chaudhary, S.; Taran, S.; Bajaj, V.; Siuly, S. A flexible analytic wavelet transform based approach for motor-imagery tasks classification in BCI applications. *Comput. Methods Programs Biomed.* **2020**, *187*, 105325. [[CrossRef](#)]
38. Dai, M.; Zheng, D.; Liu, S.; Zhang, P. Transfer kernel common spatial patterns for motor imagery brain-computer interface classification. *Comput. Math. Methods Med.* **2018**, *2018*, 9871603. [[CrossRef](#)]
39. She, Q.; Chen, K.; Ma, Y.; Nguyen, T.; Zhang, Y. Sparse representation-based extreme learning machine for motor imagery EEG classification. *Comput. Intell. Neurosci.* **2018**, *2018*, 9593682. [[CrossRef](#)]
40. Chen, J.; Stern, M.; Wainwright, M.J.; Jordan, M.I. Kernel feature selection via conditional covariance minimization. In *Advances in Neural Information Processing Systems 30*; NeurIPS: San Diego, CA, USA, 2017.
41. Constantinopoulos, C.; Titsias, M.K.; Likas, A. Bayesian feature and model selection for Gaussian mixture models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1013–1018. [[CrossRef](#)]
42. Chen, C.; Weiss, S.T.; Liu, Y.Y. Graph Convolutional Network-based Feature Selection for High-dimensional and Low-sample Size Data. *arXiv* **2022**, arXiv:2211.14144.
43. Molla, M.K.I.; Al Shiam, A.; Islam, M.R.; Tanaka, T. Discriminative feature selection-based motor imagery classification using EEG signal. *IEEE Access* **2020**, *8*, 98255–98265. [[CrossRef](#)]
44. Blankertz, B.; Muller, K.R.; Curio, G.; Vaughan, T.M.; Schalk, G.; Wolpaw, J.R.; Schlogl, A.; Neuper, C.; Pfurtscheller, G.; Hinterberger, T.; Schroder, M. The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1044–1051. [[CrossRef](#)] [[PubMed](#)]
45. Galán, F.; Oliva, F.; Guàrdia, J., III. BCI Competition III. Dataset V: Algorithm Description. 2005. Available online: http://www.bbci.de/competition/iii/results/martigny/FerranGalan_desc.pdf (accessed on 13 October 2018).
46. Miah, A.S.M.; Ahmed, S.R.A.; Ahmed, M.R.; Bayat, O.; Duru, A.D.; Molla, M.K.I. Motor-Imagery BCI task classification using riemannian geometry and averaging with mean absolute deviation. In Proceedings of the 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT) IEEE, Istanbul, Turkey, 24–26 April 2019; pp. 1–7.

47. Miah, A.S.M.; Islam, M.R.; Molla, M.K.I. Motor imagery classification using subband tangent space mapping. In Proceedings of the 2017 20th International Conference of Computer and Information Technology (ICIT) IEEE, Dhaka, Bangladesh, 22–14 December 2017; pp. 1–5.
48. Gaur, P.; Gupta, H.; Chowdhury, A.; McCreddie, K.; Pachori, R.B. A Sliding Window Common Spatial Pattern for Enhancing Motor Imagery Classification in EEG-BCI. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–9. [[CrossRef](#)]
49. Saha, S.K.; Sarker, P.K.; Abdullah Al Shiam, M.; Rahoman, M. Motor Imagery EEG Signal Classification Using MWT-CSP for Online BCI Implementation. *Int. J. Comput. Sci. Inf. Secur. (IJCSIS)* **2020**, *18*, 124–130.
50. Yamada, M.; Jitkrittum, W.; Sigal, L.; Xing, E.P.; Sugiyama, M. High-dimensional feature selection by feature-wise kernelized lasso. *Neural Comput.* **2014**, *26*, 185–207. [[CrossRef](#)]
51. Zhao, Z.; Anand, R.; Wang, M. Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform. In Proceedings of the 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA) IEEE, Washington, DC, USA, 5–8 October 2019; pp. 442–452.
52. Liu, T.; Jiang, H.; Chen, Q. Input features and parameters optimization improved the prediction accuracy of support vector regression models based on colorimetric sensor data for detection of aflatoxin B1 in corn. *Microchem. J.* **2022**, *178*, 107407. [[CrossRef](#)]
53. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28. [[CrossRef](#)]
54. Izenman, A. Linear Discriminant Analysis. In *Modern Multivariate Statistical Techniques*; Springer: New York, NY, USA, 2013; pp. 237–280.
55. Sánchez-Reolid, R.; García, A.S.; Vicente-Querol, M.A.; Fernández-Aguilar, L.; López, M.T.; Fernández-Caballero, A.; González, P. Artificial neural networks to assess emotional states from brain-computer interface. *Electronics* **2018**, *7*, 384. [[CrossRef](#)]
56. He, Y.; Lu, Z.; Wang, J.; Ying, S.; Shi, J. A Self-Supervised Learning Based Channel Attention MLP-Mixer Network for Motor Imagery Decoding. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *30*, 2406–2417. [[CrossRef](#)]
57. Feng, J.K.; Jin, J.; Daly, I.; Zhou, J.; Niu, Y.; Wang, X.; Cichocki, A. An optimized channel selection method based on multifrequency CSP-rank for motor imagery-based BCI system. *Comput. Intell. Neurosci.* **2019**, *2019*, 8068357. [[CrossRef](#)] [[PubMed](#)]
58. Singh, A.; Lal, S.; Guesgen, H.W. Reduce calibration time in motor imagery using spatially regularized symmetric positive-definite matrices based classification. *Sensors* **2019**, *19*, 379. [[CrossRef](#)] [[PubMed](#)]
59. Zhang, Y.; Wang, Y.; Zhou, G.; Jin, J.; Wang, B.; Wang, X.; Cichocki, A. Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces. *Expert Syst. Appl.* **2018**, *96*, 302–310. [[CrossRef](#)]
60. Park, Y.; Chung, W. BCI classification using locally generated CSP features. In Proceedings of the 2018 6th International Conference on Brain-Computer Interface (BCI) IEEE, Gangwon, Republic of Korea, 15–17 January 2018; pp. 1–4.
61. Selim, S.; Tantawi, M.M.; Shedeed, H.A.; Badr, A. A csp\am-ba-svm approach for motor imagery bci system. *IEEE Access* **2018**, *6*, 49192–49208. [[CrossRef](#)]
62. Singh, A.; Lal, S.; Guesgen, H.W. Small sample motor imagery classification using regularized Riemannian features. *IEEE Access* **2019**, *7*, 46858–46869. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.