Chapter 5

Wrapper-based Channel Selection Method Using Multi-objective Improved Firefly Algorithm

5.1. Introduction

Metaheuristic algorithms are computational intelligence paradigms especially used for solving sophisticated optimization problems. In this context, wrapper methods are another class of methods that seek the most feasible solution from a large search space using a metaheuristic-inspired iterative search strategy [107]. In this context, thousands of metaheuristic algorithms have been proposed to date. The performance of these algorithms is application-specific and depends upon various factors such as sample size, distribution of data, and number of classes. Therefore, it is highly challenging to select an effective metaheuristic algorithm from a large pool and realize better performance in the Channel Selection Problem. To resolve this problem, we fixed three main parameters: (1) high compatibility with hardware which is a major component of our BCI framework, (2) ability to work in a noisy environment (EEG signals are noisy), and (3) effective to handle multimodal data (current trends of BCI). Considering these objectives, we reviewed multiple research papers, articles, doctoral dissertations, and reports based on the behavior of multiple popular metaheuristic algorithms such as Genetic Algorithm (GA) [108], Particle Swarm Optimization (PSO) [109], Ant Colony Optimization (ACO) [110], and Firefly Algorithm (FA) [111]. Among these algorithms, two algorithms viz. FA and PSO have been implemented in various large-scale optimization problems to compute a feasible solution. In one study [112], authors performed several experiments to compare the performance of FA and PSO on the optimization of complex objective functions in a noisy environment. In these experiments, various levels of noise are added to each objective function and concluded that for higher noise levels, the FA wins out in terms of accuracy of the results **(ability to handle noisy data).** Here, the difference between speedups of different configurations (noise levels) shows that the FA runs faster on small to medium-scale problems compared to the PSO when implemented on different processors **(hardware compatibility)**. A similar performance was observed in a different research problem based on “Congestion management of deregulated power market”, where that FA was proved to be a better algorithm compared to PSO for minimizing the objective function [157]. This experiment was performed on the **IEEE 30-bus system** positively. In a different work [113], the performance of FA was superior in dealing with multimodal data compared to other metaheuristic algorithms **(Multimodal superiority)**. In addition, the convergence rate of FA is faster than, and parameters in FA can be tuned to control the randomness as iterations proceed, so convergence can also be sped up by tuning these parameters. Based on these studies, we used FA in our work to compute the optimal channel set for maximizing the classification accuracy of a multiclass BCI system.

In this work, Firstly, the preprocessed neural signals are used to extract spatial-temporal features using the Regularized Common Spatial Pattern with Aggregation (RCSPA) method [114]. Then, objective FA with two input variables (Spectral Entropy and Lyapunov exponent) is used to compute a weighted score for each channel in the neighborhood of a candidate solution. Finally, a novel Channel Set Relevance Index (CSRI) is developed to rank channels using their respective weighted score and Fisher information [115]. The RCSPA features of highly ranked channels are employed to discriminate different MI tasks using the Regularized Support Vector Machine (RSVM) classifier [116].

The proposed approach is cross-validated on three publicly available BCI competition datasets with varying numbers of channels. The validation results show that the proposed method achieved a superior classification accuracy (83.97% on Dataset 1, 80.85% on Dataset 2, and 84.19% on Dataset 3) with fewer channels than other baseline methods. In addition, our method significantly reduced the BCI preparation time, making it effective for conducting multiple experimental sessions for a large pool of subjects. The detailed code can be accessed from Link 1 given in the footer. The block diagram of the proposed methodology is shown in Figure 5.

[1] https://drive.google.com/drive/folders/1ETuYzlh7288KI8uBryHOLprNjj1dNam_?usp=sharing

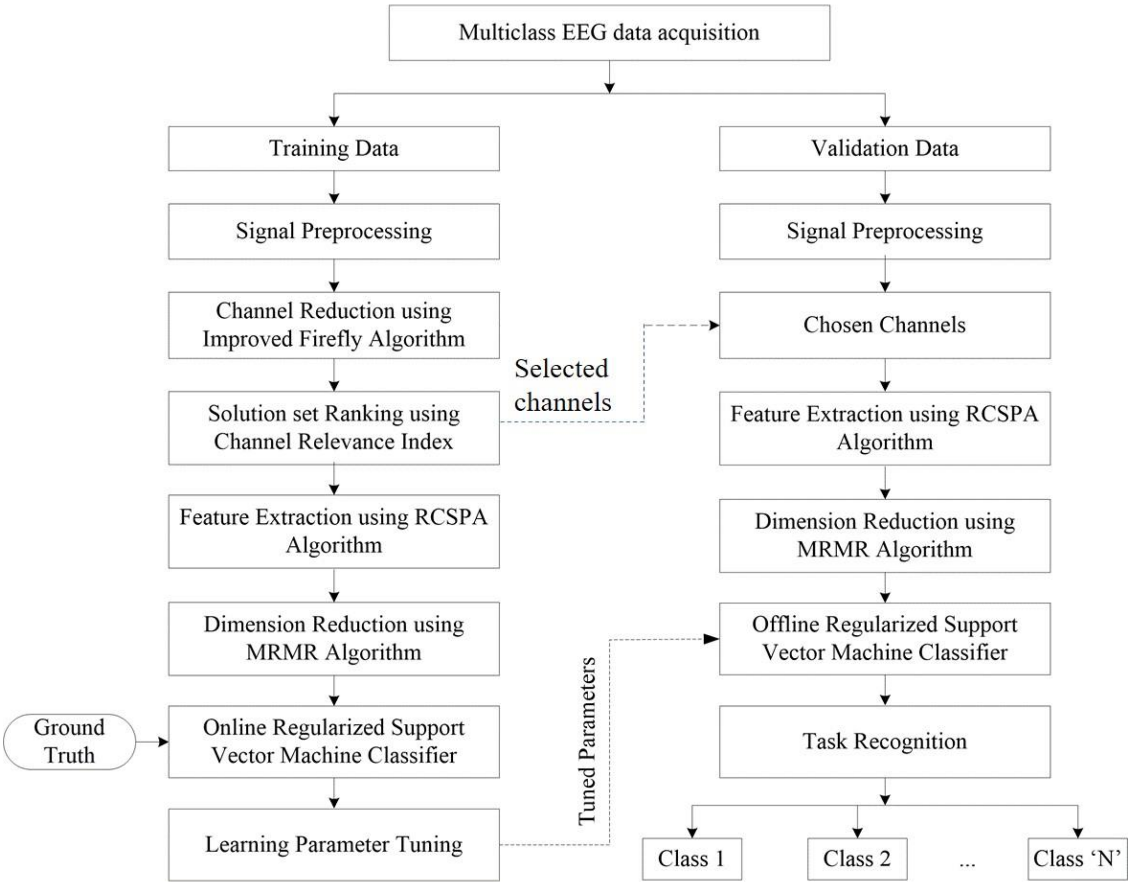


Figure. 5.1. Block diagram of proposed EEG channel selection methodology using improved Firefly Algorithm

5.2. Limitations of the Existing Literature

Wrapper methods are a popular class of optimization methods that compute better quality solutions than filter methods because of their generalization ability, robust interaction with classifiers, and rich support to advanced comprehensive search techniques such as neighborhood measure and dependency estimation. However, these methods have the following limitations:

- Poor compatibility with hardware when implemented for real-time data
- Inefficient while dealing with noisy data samples
- Incapable of working in a multimodel environment
- Longer execution and response time
- Computationally infeasible solutions when used with a large number of variables.

5.3. Our Contribution

The main contributions of the chapter can be summarised as follows:

I. To solve the above-mentioned issues, we developed a wrapper-based channel ranking method using a multi-objective improved variant of conventional FA and Fisher Score.

II. We ranked each channel by computing their relevance in the neighborhood of candidate solution using the below-mentioned two objectives: (1) Maximization spectral entropy and (2) Minimization of signal irregularity (Lyapunov exponent)

We showed that the proposed approach rapidly converges compared to state-of-the-art methods with a better trade-off between classification accuracy and the number of selected channels.

5.4. Firefly Algorithm: Mathematical Background

The Firefly Algorithm was developed by Yang (2009) [111] for continuous optimization, which was subsequently applied to structural optimization [117] and image processing [118]. In essence, FA uses the following three idealized rules: (1) fireflies are unisex so that one firefly will be attracted to other fireflies regardless of their sex; (2) the attractiveness of a firefly is proportional to its brightness, and they both decrease with distance. Thus for any two flashing fireflies, the less bright one will move towards the brighter one. If there is no brighter one than a particular firefly, it will move randomly; (3) the brightness of a firefly is determined by the landscape of the objective function.

In the simplest case for maximum optimization problems, the brightness I of a firefly at a particular location x can be chosen as $I(x) \epsilon f(x)$. However, the attractiveness β is relative, it should be seen in the eyes of the beholder or judged by the other fireflies. Thus, it will vary with the distance r_{ij} between firefly i and firefly j. Therefore, we can now define the attractiveness β of a firefly by

$$\beta = \beta_o e^{-\gamma r_{i,j}^2} \tag{5.1}$$

The distance between any two fireflies i and j at x_i and x_j, respectively, is the Cartesian distance $r_{ij} = \|x_i - x_j\|$. For any given two fireflies x_i and x_j, the movement of firefly i is attracted to another more attractive (brighter) firefly j is determined by

$$x_i^{t+1} = x_i^t + \beta_o e^{-\gamma r_{ij}^2}\left(x_j^t - x_i^t\right) + \alpha_t \varepsilon_i^t \tag{5.2}$$

where the second term is due to the attraction. The third term is randomization with α_t being the randomization parameter, and ε_i^t is a vector of random numbers drawn from a Gaussian distribution or uniform distribution. The location of fireflies can be updated sequentially, by comparing and updating each pair of them in every iteration cycle.

5.5. Proposed Methodology

A variety of pre-processing steps are applied to improve the signal quality. Further, the pre-processed signals are used in the channel selection and discrimination of MI-specific cognitive state patterns. The detailed work of the proposed approach is discussed in the subsequent subsections.

5.5.1. Signal Preprocessing

The raw data undergoes multiple data transformation steps. They concentrate on improving SNR by reducing the dispersion level of signals and artifacts. The following steps are used in the preprocessing phase.

5.5.1.1. Noise Filtering

Since the SNR of the EEG signals is relatively low and depends heavily on different muscle movement-based artifacts, they must be minimized before application. Hence, we evaluated the signal spectra's dispersion as the absolute deviation from the mean signal value. In our previous work, we have shown that the Savitzky- Golay function performed well in smoothing the raw EEG signals without altering their functional properties. Therefore, in this experiment, we use the same refined neural signals of all three datasets (Dataset 1, Dataset 2, and Dataset 3).

5.5.1.2. Rhythm Isolation

We separate central beta frequency from EEG signal spectra. Two second-order zero-phase IIR (Infinite Impulse Response) filters are linearly cascaded using EEGLAB software [119]. In the first IIR filter, we specified a 10 Hz to 30 Hz band as the cutoff to extract nearly localized frequency components. Further, the extracted frequency region is input in the second IIR filter, and frequency components between 12 Hz to16 Hz are filtered.

5.5.2. Channel Selection

A distance-based Firefly Algorithm (FA) is used in the proposed study to find an optimal EEG channel set with minimum redundant information. We employed FA as a two-feature-based mapping approach where a weighted objective function takes in spectral entropy [120] and Lyapunov exponent [121] of a given channel as an input and provides a score to rank it. This approach starts by choosing three channels (C_3, C_4, C_z) as a candidate solution from a set of N channels. Next, it constructs a population set of the remaining (N-3) channels except for the earlier candidates. Further, the channel instances are updated using the FA (Eq. 5.1 and 5.2), and the fitness value is computed for each iteration. The algorithm computes the relative closeness (($\|\Delta_{ij}\|$)) between the mean of the fitness value of candidate channels and i^{th} channel of the set using Equation 5.3. If this proximity is less than the defined threshold (≤0.5), the channel is selected for further processing; else eliminated.

$$\Delta_{ij} = f(C_j) - f(R_i) \tag{5.3}$$

where $f(C_j)$ and $f(R_i)$ represent the mean objective function score of candidate channels and i^{th} channel respectively. Here, the main goal of finding relative closeness between the candidates and remaining channels is to provide an ordering based on the firefly movement function (Equation 5.2). To estimate the weightage of the selected channel, we computed the weighted linear bi-objective function score (WLBOFS) for the respective channel (Equation 5.4). The channel with the lowest WLBOFS score is considered the highest-ranked and used for deciding the rank remains.

$$\text{WLBOFS}\,(i,j) = \frac{\|\Delta_{i,j}\|}{\sum_{j=2}^{j=N} \|\Delta_{i,j}\|} \tag{5.4}$$

where 'i' and 'j' are indices of selected and candidate channel, respectively. WLBOFS (i, j) represents a weighted score of channel 'j' with respect to candidate channel (i), and 'N' is the total number of channels. In the case of the lower score, a higher rank is achieved, and vice versa. The pseudocode of the proposed algorithm is given in Algorithm 5.1.

Algorithm 5.1: Algorithm for proposed channel ranking approach

```
Input:  Channels: X = $\{c_1, c_2, \dots c_N\}$ and a weighted bi-objective function $f(x_i) = w_1 * (\rho_i) + w_2 * (\varphi_i)$ with input features spectral entropy $(\rho)$ and Lyapunov exponent$(\varphi)$ where $i \in (1, \dots, N)$
Output: The sequence of channels is obtained based on the computed WLBOFS score

Start:
Step 1: Define candidate channels set from X as a candidate $(c_k)$ and update the population by removing the $c_k$, thereby updating the population: $X' = \{c_1, c_2, \dots c_N\} - \{c_k\}$
Step 2: Compute input features for the candidate and estimate $f(c_k)$ for them. Initialize weights $[w_1, w_2]$ with [0.99,0.01] for the first iteration and update it with [0.99-$\epsilon$, 0.01+ $\epsilon$] in further iterations where $\epsilon$ is the step size that decides the number of iterations. In our case, $\epsilon$ = 0.01, t=0, Max_iteration = 100, $w_1 + w_2 = 1, w_1 \neq 0, w_2 \neq 0$
Step 3: while (t < Max_iteration) do
            for (i =1 to N) do
                Evaluate the next updated value of $f(c_k)$ for each $c_i \in X$
                Calculate proximity $(\Delta_{ki})$ between the candidate channel and each $c_i \in X$
                Arrange them in increasing sequence using WLBOFS(k, i) score such that the channel at the $j^{th}$ position has a lower score as compared to the one at the $(j + 1)^{th}$ position
                Calculate mean WLBOFS (k, i) for 100 iterations
                    if (mean WLBOFS (k, i) < threshold)                threshold =0.50
                        Select $i^{th}$ channel for further processing; else, eliminate it
                    end if
            end for
            t= t+1;
        end while
End
```

5.5.3. Channel Role Evaluation Using Fisher Information Criterion

In machine learning, channel role evaluation refers to a task that measures the relevance of individual input channels in the performance of the classification model. Channel set hierarchy allows the model to avoid redundancy that reduces classification accuracy. In other words, it maximizes the discrimination ability of the applied classifier by listing significant features in terms of minimum redundancy and maximum relevance. The main aim of the channel ranking process is threefold:

I. Reduction of computational complexity by selecting the relevant channels,
II. To minimize the amount of overfitting that may arise due to the utilization of unnecessary channels
III. Minimizing the setup time in some applications.

It provides ordering of the channels using a scoring function that measures channel set importance. This study designs a channel set relevance index (CSRI) using Fisher information, which helps arrange all channels according to their mutual importance. Each test case's maximum likelihood score is evaluated using mean absolute Fisher information. Statistically, Fisher information is used for measuring the amount of information carried by a random candidate channel P about an unknown parameter θ that represents the entropy enclosed with the associated set C_k $(1 \leq k \leq N)$.

Let a dataset of size $[N, D]$ where D denotes channels and N refers to total instances in the dataset. The log-likelihood function I' for the channel, X is calculated using Eq. 5.5.

$$I'(X \vee \theta) = log f(X \vee \theta) \tag{5.5}$$

where θ represents entropy associated with a given EEG sample of channel X, the Fisher information (concerning entropy variation) is defined in Eq. 5.6.

$$\mathrm{I}(\theta) = E_\theta\{[I'(X \vee \theta)^2]\} = \int I'(X \vee \theta)^2 f(X \vee \theta) dx \tag{5.6}$$

Therefore, the mean Fisher information (MFI_N) across all N channels is calculated using Eq. 5.7.

$$MFI_N = E\,[\textstyle\sum_{k=1}^{N} I_k(\theta)] = \frac{\sum_{k=1}^{N} I_k(\theta)}{N} \tag{5.7}$$

The channel set relevance index (CSRI) is:

$$\text{CSRI } (\%) = \left(\left|\frac{MFI_N - I_k(\theta)}{MFI_N}\right|\right) \times 100 \tag{5.8}$$

The CSRI can rank all the channels according to their relative closeness with mean Fisher information. Note that this index shows the normalized percentage difference between the mean Fisher information and Fisher information of k^{th} channel. An example of the histogram count representing the CSRI score for all channels is shown in Figure. 5.2, where the x axis represents the channel index according to the specification defined in the dataset DS1, and y axis indicates the normalized CSRI (%) score of each channel.

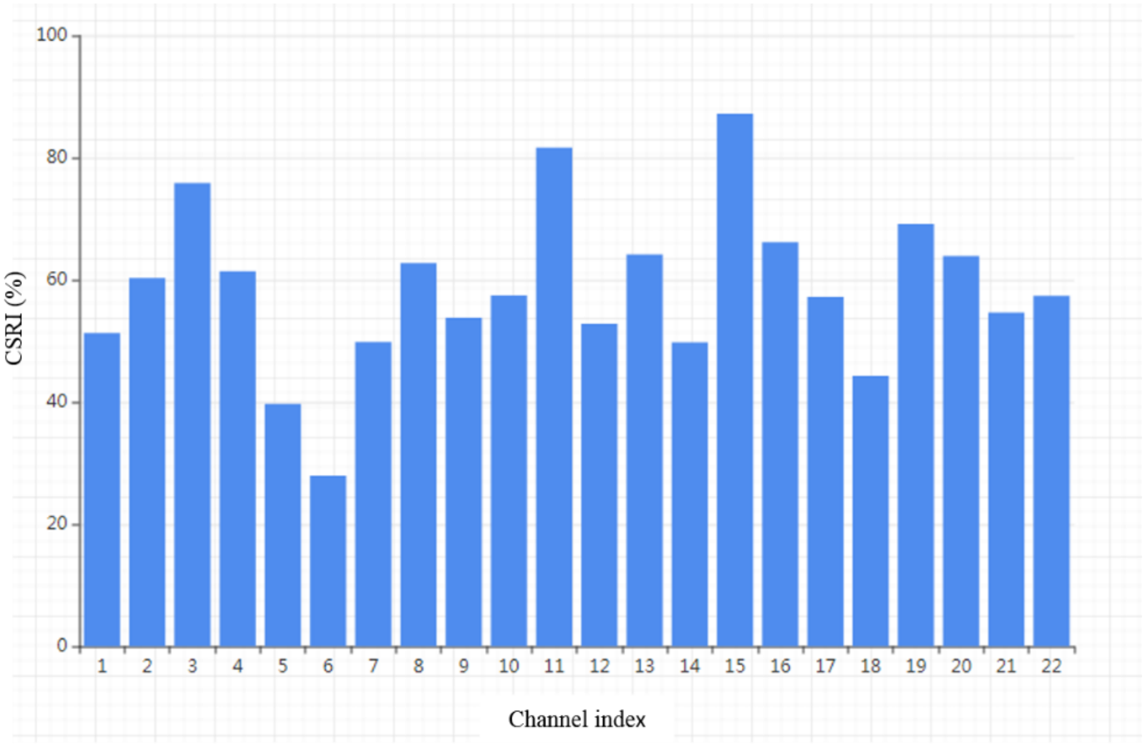


Figure 5.2. A histogram count representation of CSRI score for all channels of dataset 1

5.6. Feature Extraction Using Regularized Common Spatial Pattern with Aggregation

The common spatial pattern (CSP) method is an effective feature extraction algorithm that discriminates motor-based EEG signals by obtaining suitable spatial filters. By applying spatial filters to the data, the variance is maximized for one class of data while minimized for the other.

Therefore, the two classes have the more distinctive features to be used in the classification stage. Despite its advantages of wide versatility and high efficiency, CSP is shown to be noise-sensitive and not suitable for dealing with limited training samples [65]. Therefore, in this study, we employ an improved version of the CSP algorithm, namely, Regularized Common Spatial Pattern with Aggregation (RCSPA), to compute spatiotemporal features of MI tasks. Unlike CSP features, RCSPA features are robust and effective in dealing with a small sample problem. The RCSPA method controls the bias-variance tradeoff among MI tasks enclosed within small samples using two regularization parameters. As defined in [65], a regularized CSP computes the average spatial covariance matrix for each MI class in a specific subject as follows:

$$\hat{\Sigma}_c((\beta,\gamma)) = (1-\gamma) * (\beta) + \frac{\gamma}{N} tr[(\beta)] * I \qquad (5.9)$$

where β, γ are regularization parameters in the range of [0,1], and I is an identity matrix of size $N \times N$ where N is the size of each segment. $\Omega_c(\beta)$ comprises covariance matrices for the trials of a class of a specific subject to the trials of the same class of other subjects. It is defined as:

$$\hat{\Omega}_c(\beta) = \frac{(1-\beta)*S_c+\beta*\hat{S}_c}{(1-\beta)*M+\beta*\hat{M}} \qquad (5.10)$$

where S_c indicates the sum of the sample covariance matrices for all M training observations in class c

$$S_c = \sum_{m=1}^{M} S(c,m) \qquad (5.11)$$

and $\hat{S}_c$ is the sum of the sample covariance matrices for $\hat{M}$ training observations from other subjects in class c

$$\hat{S}_c = \sum_{\hat{m}=1}^{\hat{M}} S(c,\hat{m}) \qquad (5.12)$$

For testing the model for each class, we chose those values from the covariance matrices $\hat{\Sigma}_c((\beta,\gamma))$ (Equation 5.9) that is higher than our chosen threshold (0.5) and used as features for classification. These features are arranged equally in a group of N columns where each column represents row-wise entries of covariance matrices $\hat{\Sigma}_c((\beta,\gamma))$.

The RCSPA results in a larger feature space because of the data decomposition of all the selected channels. To obtain better classification accuracy, it is required to process only those features that significantly contribute to the classification process and MI-tasks relevant. A popular feature selection approach called minimum redundancy maximum relevancy (mRMR) is used in classification problems based on non-invasive recordings [88]. This approach considers the mutual information between the class labels C_i and individual feature F_i and ranks them accordingly. Some features may depend on each other in a given feature space, and their combinations could be redundant. Hence, it is necessary to have only important features that are significant to improve classification accuracy.

In our study, the selected five features are statistically significant with a p-value < 0.05 in the training session for the left-hand and right-hand MI tasks. The feature ranking is performed on the extracted features with the Wilcoxon test method for the left-hand and right-hand MI tasks. Then, all the features are arranged based on their rank in the descending order of class separability. A similar approach is also applied to tongue and foot classes with relative p-values of 0.0191, 0.0207, 0.0429, 0.0167, and 0.0091. The extracted features are further normalized using the following Eq. 5.13.

$$x' = \frac{\mathrm{x} - \min(\mathrm{x})}{\max(\mathrm{x}) - \min(\mathrm{x})} \tag{5.13}$$

where x is a real signal instance at a time, 't' and x' is the scaled value of the same instance at a time 't'.

5.7. Results & Discussion

A regularized version of the Support Vector Machine (RSVM) is implemented to improve accuracy, and stability and minimize the over-fitting problem in the multiclass MI tasks classification. A primary motivation behind the application of RSVM is inspired by its ability to deal with non-linear data observations successfully. To reduce the complexity of multiclass classification, a One-vs.-One classification strategy is applied to all the extracted features. Henceforth, K training sessions (${}^{K}C_2$) are created for all three datasets, where K refers to the number of classes in the respective dataset. For example, six sessions (Left vs. Right, Left vs. Tongue, Left vs. Feet, Right vs. Tongue, Right vs. Feet, and Tongue vs. Feet) are created in the training session of DS1. The BCI competition community separately provides training and evaluation data samples for DS1; therefore, an additional application of any cross-validation scheme is not required. However, a fivefold cross-validation technique decomposes DS2 and DS3 in the training and validation set. Finally, RSVM is trained on all MI-tasks pairs and directly used to classify all unknown test trials of the corresponding evaluation datasets for each subject and respective pair of classes. In the subsequent subsections, we described the performance of the proposed channel selection method.

5.7.1. Test Results on Dataset 1

To investigate the proposed channel selection method's performance, we analyze the relationship between the optimum number of channels and corresponding classification accuracy. We execute our algorithm on all nine subject-specific evaluation samples and calculate the mean classification accuracy for all ${}^4C_2$ pairs of MI tasks. The three best sets of optimal channels (C_3, C_4, C_z) are considered for further comparative analysis between the proposed method and state-of-the-art algorithms.

The obtained classification results over all nine subjects and ${}^4C_2$ pairs are given in Table 5.1. Our proposed methodology realized an average classification accuracy of 80.28 % in the discrimination of Left vs. Right-hand classes with a 1.2% improvement over the Improved Binary Gravitation Search Algorithm (IBGSA) based channel selection method.

Table 5.1. Classification accuracy of the proposed approach for all nine subjects with six training sessions using a one-to-one classification method

Subject	Classification accuracy for C_3 candidate channel with proposed methodology (in %)					
	Left vs Right	Left vs Tongue	Left vs Feet	Right vs Tongue	Right vs Feet	Feet vs Tongue
A01	89.04	91.21	94.59	97.14	96.41	81.05
A02	64.80	81.43	77.18	87.18	71.58	71.04
A03	83.68	93.30	91.04	94.04	91.57	70.01
A04	84.01	85.67	82.38	89.50	84.29	78.34
A05	62.50	72.51	89.60	71.32	86.18	83.14
A06	74.31	77.91	67.29	79.83	91.03	71.03
A07	82.30	83.15	83.18	88.16	89.51	79.64
A08	89.71	90.05	85.29	91.52	73.18	77.20
A09	92.18	87.30	93.17	71.29	97.36	72.79
Average	80.28	84.72	84.85	85.55	86.79	76.02

In a comparative analysis, we execute both optimized FA-based channel selection and the IBGSA method for 100 independent runs subject-wise. In a similar comparison, achieved classification results are compared with Glow Swarm Optimization (GSO) [122] and the Robust and Subject-Specific Sequential Forward Search Method (RSS-SFSM) [123]. Here, the classification accuracy results are 73.91 ± 3.18 for GSO and 67.19 ± 6.04 for the RSS-SFSM approach, which shows the superior classification performance of the optimized FA-based channel selection approach. In summary, the proposed method's classification accuracy is compared with the above-mentioned methods for all the ${}^4C_2$ cases shown in Figures 5.3, 5.4, 5.5, 5.6, 5.7, and 5.8.

From Figs. 5.3 and 5.4, it is clear that the proposed method realizes maximum classification accuracy in the Left vs. Right-hand and Left vs. Tongue classification problems for six out of nine participants. The GSO algorithm achieves maximum classification accuracy for two subjects in both cases, while the IBGSA method outperforms one participant. Similarly, for Left vs. Feet (Figure 5.5) and Right vs. Tongue (Figure 5.6), our method obtains the best classification accuracy for seven participants while GSO achieves superior classification accuracy for the remaining two participants. For the Right vs. Feet classification problem (Figure 5.7), our approach outperforms baseline channel selection algorithms for seven participants while the GSO and IBGSA achieve superior classification accuracy for the remaining two participants. Finally, in the Feet vs. Tongue discrimination (Fig. 5.8), the proposed method computes the maximum accuracy for four participants while the GSO and IBGSA methods achieve maximum classification performance for three and two participants, respectively.

In Table 5.2, we summarize the overall performance of all discussed channel selection algorithms in terms of three parameters: (1) Number of selected channels, (2) Channel reduction rate, (3) and Classification accuracy. Here, it can be seen that the proposed channel selection scheme achieves better mean classification accuracy for all pairwise MI tasks as compared to IBGSA, GSO, and RSS-SFSM algorithms. It enhances classification results by 10.41% as compared to the IBGSA approach. Compared with GSO and RSS-SFSM methods, we found that our proposed algorithm improved classification results by 15.93% and 5.15%, respectively. In addition, Our approach achieves a better channel reduction rate than the remaining three algorithms with the least execution time. One of the main reasons for showing a better execution rate is its attractive function inspired by firefly behavior. Therefore, we can conclude that the optimized FA scheme deals with non-linear and noisy data more optimally than IBGSA, GSO, and RSS-SFSM-based methods.

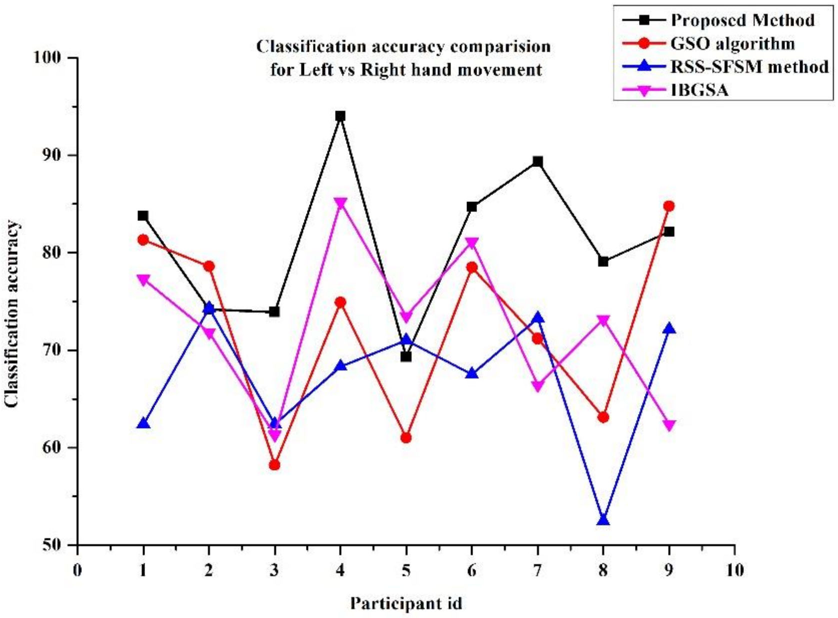


Figure 5.3. Classification accuracy of the proposed method and other methods for Left Hand Vs. Right-hand MI activities

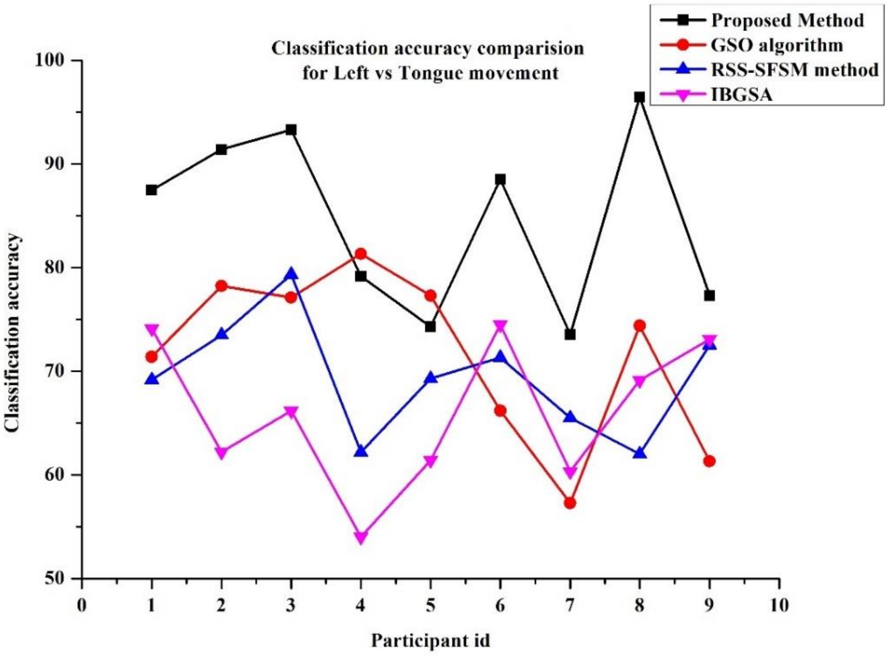


Figure 5.4. Classification accuracy of the proposed method and other methods for Left Hand Vs. Tongue MI activities

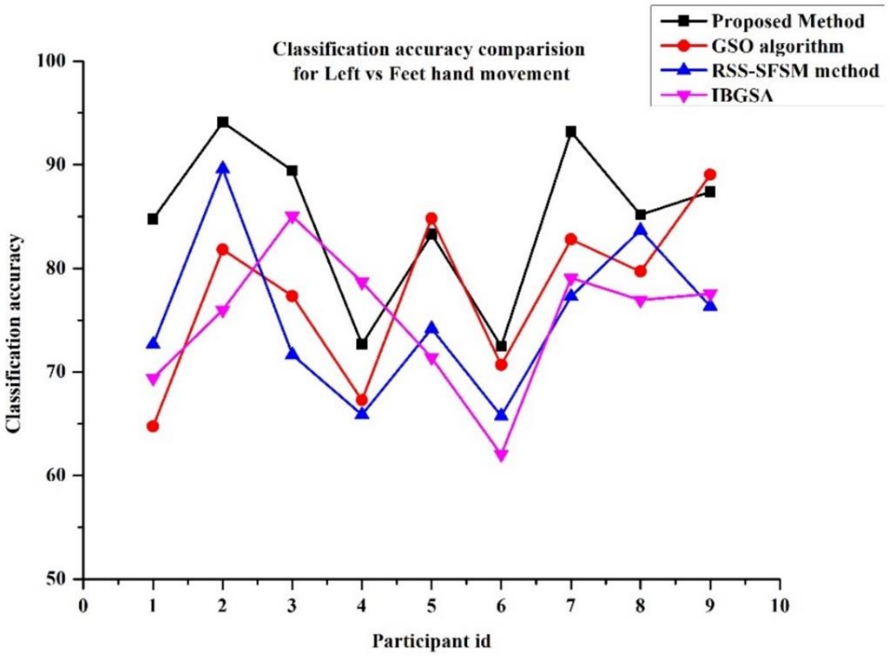


Figure 5.5. Classification accuracy of the proposed method and other methods for Left Hand Vs. Feet MI activities

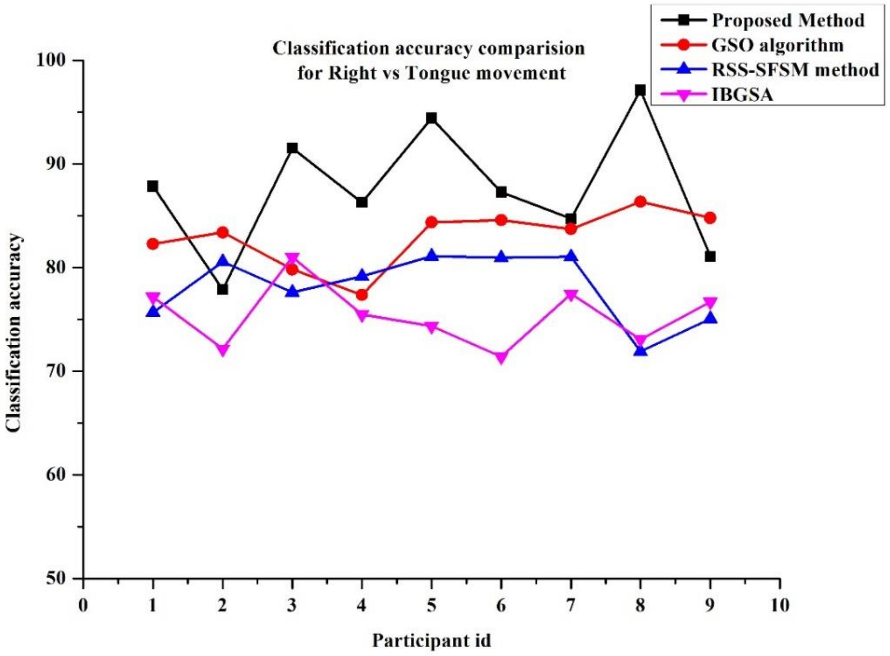


Figure 5.6. Classification accuracy of the proposed method and other methods for Right Hand Vs. Tongue MI activities

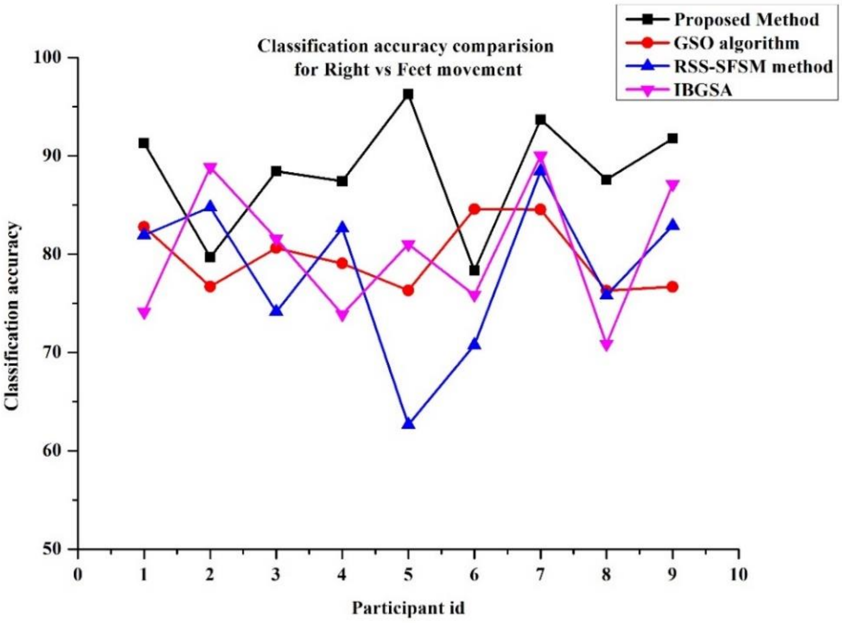


Figure 5.7. Classification accuracy of the proposed method and other methods for Right Hand Vs. Feet MI activities

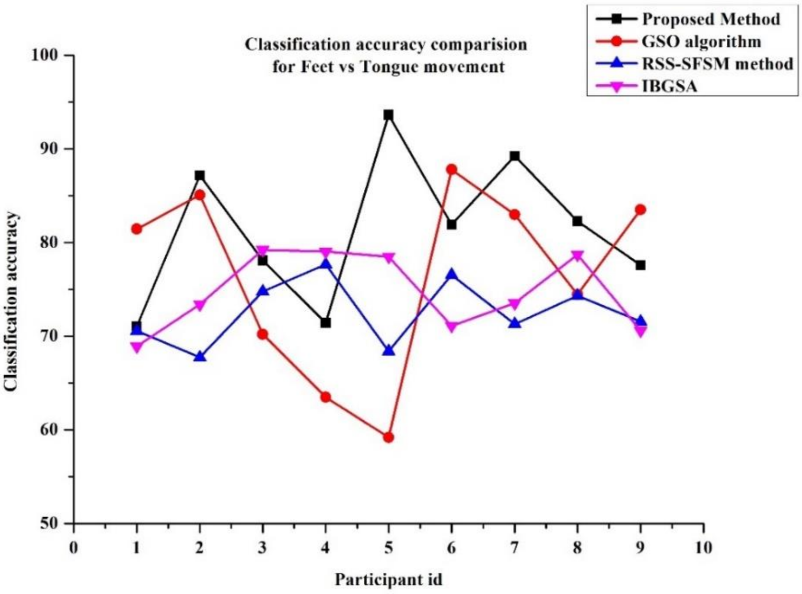


Figure 5.8. Classification accuracy of the proposed method and other methods for Feet Vs. Tongue MI activities.

5.7.2. Test results for Dataset 2 & Dataset 3

We will discuss both jointly in this subsection since the BCI community has provided learning and validation data for DS1 contrary to DS2 and DS3. Therefore, a fivefold cross-validation approach is used to split into learning and validation sets. We compared the performance of our method on both the datasets with four state-of-the-art algorithms (1) all channels method, (2) Support Vector Machine Recursive Feature Elimination (SVM-RFE), (3) Sequential Floating Forward Selection (SFFS), and (4) Improved Sequential Floating Forward Selection (Improved SFFS) in Table 5.3 and 5.4. The details of these algorithms have been discussed in chapter 5. The results corresponding to all four algorithms are directly taken from Qiu et al., 2016 [29]. The last row of the tables indicates the p-values computed from the paired t-test [124] of the results of "all channels" and the channels selected by baseline methods. Tables 5.3 and 5.4 summarize the performance of our method with different channel selection approaches mentioned above on DS2 and DS3. It is already discussed that DS2 and DS3 are the BCI competition datasets with a moderate (59 channels) and large (118) number of channels compared to DS1 (only 22 channels). From Table 5.3, it can be noticed that our method realizes maximum classification accuracy for five out of seven participants. Moreover, the proposed algorithm achieved a mean classification accuracy of 80.85%, significantly higher than the other four methods. Compared to the SVM-RFE (p-value = 0.0045), our method yielded an 11.51% improvement in mean classification accuracy but used two more channels. However, the current method maintained a proper balance between the selected channels and achieved mean classification accuracy compared to the SFFS (p-value = 0.002) and Improved SFFS (p-value = 0.0025) methods. Here, it should be noticed that SVM-RFE eliminated a few relevant and non-redundant channels without considering their role in improving global classification accuracy. Two other methods, namely SFFS and Improved SFFS, have shown almost equal performance with marginal differences in respective CRR values. Moreover, our method obtained superior performance in classification accuracy and CRR.

The performance of the proposed method on DS3 is compared in Table 5.4. This dataset represents the highly dense distribution of electrodes on the human scalp. Although the performance of the proposed method (p-value = 0.058) is almost equal to the SFFS algorithm (p-value = 0.023), our results are relatively less significant than the SFFS method. However, the channel reduction rate of our method is better than the SFFS algorithm. In addition, our method realized the mean classification accuracy of 84.19%, which yields an improvement of 7.11%, 1.00%, and 5.89% compared to the SVM-RFE, Improved SFFS, and all channels, respectively, using only 27 channels. It indicates that the application of the proposed method should be preferred over other methods on large-size datasets.

Similar to previous chapters, the overall asymptotic complexity of the proposed approach depends on four steps: (1) EEG preprocessing, (2) Channel selection using FA, (3) Feature engineering, and (4) Classification. It has been discussed in Chapter 3 that preprocessing using the SG filter requires $O(n^2)$ comparison where n is the size of the filtering window. The Firefly algorithm has two inner loops when going through the population n, and one outer loop for iterations K. So the complexity in the extreme case is $O(n^2K)$. In feature engineering, the complexity of the RCSP algorithm mainly arises from the eigenvalue decomposition or singular value decomposition (SVD) of the covariance matrix of the input EEG signals. The SVD or eigenvalue decomposition has a computational complexity of $O(L^3)$, where L is the number of EEG channels. In feature reduction, the MRMR algorithm typically involves computing pairwise relevance and redundancy scores for each feature. If there are N features in the input data, the complexity can be proportional to $O(N^2)$, or higher, depending on the set of calculations involved. Finally, the classification step with the SVM technique will execute in $O(n^2)$ (LibSVM).

Table 5.2. Performance comparison of the proposed channel selection approach with IBGSA, GSO, RSS-SFSM.

Subject	MI tasks	IBGSA	GSO	RSS-SFSM	Proposed approach
A01	Mean accuracy	71.30	68.31	73.91	91.57
	No. of selected channels	7	11	9	5
	Channel reduction rate	0.68	0.5	0.59	0.77
A02	Mean accuracy	66.04	56.91	70.08	75.53
	No. of selected channels	10	13	6	5
	Channel reduction rate	0.54	0.40	0.72	0.77
A03	Mean accuracy	81.54	79.68	85.02	87.27
	No. of selected channels	8	13	5	5
	Channel reduction rate	0.63	0.40	0.77	0.77
A04	Mean accuracy	81.91	69.43	77.29	84.03
	No. of selected channels	9	14	12	7
	Channel reduction rate	0.59	0.36	0.45	0.68
A05	Mean accuracy	76.66	73.33	80.00	77.54
	No. of selected channels	13	9	11	7
	Channel reduction rate	0.40	0.59	0.50	0.68
A06	Mean accuracy	66.66	83.52	79.52	85.32
	No. of selected channels	13	9	6	5
	Channel reduction rate	0.40	0.59	0.72	0.77
A07	Mean accuracy	73.57	66.33	82.45	84.32
	No. of selected channels	11	14	9	7
	Channel reduction rate	0.50	0.36	0.59	0.68
A08	Mean accuracy	79.32	63.10	83.11	84.49
	No. of selected channels	7	5	8	11
	Channel reduction rate	0.68	0.77	0.63	0.50
A09	Mean accuracy	87.52	91.33	89.28	85.68
	No. of selected channels	12	7	13	6
	Channel reduction rate	0.45	0.68	0.40	0.72
Average	Mean accuracy	76.05	72.43	79.85	83.97
	No. of selected channels	10	11	8.77	6.44
	Channel reduction rate	0.54	0.51	0.59	0.70

Table 5.3. Performance comparison of the proposed method with baseline approaches on dataset 2 with 59 channels

Subject	All channels (%)	Improved SFFF (%)		SFFS (%)		SVM-RFE (%)		Proposed method (%)	
A	43.0	69.0	6.0	60.0	9.0	57.0	4.0	72.80	8
B	42.0	63.0	15.0	66.0	19.0	54.0	2.0	66.18	12
C	62.0	87.0	26.0	91.0	21.0	84.0	32.0	83.72	26
D	81.0	94.0	29.0	94.0	34.0	88.0	26.0	96.10	13
E	91.0	96.0	19.0	96.0	21.0	95.0	10.0	83.50	15
F	49.0	65.0	8.0	58.0	19.0	58.0	4.0	76.33	14
G	66.0	72.0	22.0	83.0	21.0	71.0	18.0	87.33	17
Mean	62.0	78.0	17.9	78.3	20.6	72.5	13.7	80.85	15
STd	18.9	14.0	8.7	16.5	7.3	16.6	11.8	9.18	5.18
p-value	-	0.0025		0.002		0.0045		0.0073	

Table 5.4. Performance comparison of the proposed method with baseline approaches on dataset 3 with 118 channels

Subject	All channels (%)	Improved SFFF (%)		SFFS (%)		SVM-RFE (%)		Proposed method (%)	
		ACC.	Channels	ACC.	Channels	ACC.	Channels	ACC.	Channels
Aa	75.7	76.4	27	78.3	26	68.6	7	81.33	23.0
Al	85.7	94.3	47	93.6	43	92.9	8	87.92	38.0
Av	62.1	65	18	68.6	18	62.9	14	71.20	25.0
Aw	87.1	89.5	27	87.9	20	80.0	35	86.30	22.0
Ay	87.1	91.4	35	92.7	35	88.6	18	94.20	27.0
Mean	79.5	83.3	30.8	84.2	28.4	78.6	16.4	84.19	27.0
STd	10.9	12.3	10.9	10.6	10.5	12.8	11.3	7.68	5.76
p-value		0.047		0.023		0.75		0.058	

5.7.3. Comparison of Channel Reduction Rate

We compared the *CRR* scores of all the channel selection methods to investigate the effect of eliminated channels on mean classification accuracies. Figure 5.9 shows the comparative performance for DS1, DS2, and DS3, respectively. The bar graph in Figure 5.9. (A) represents the relationship between mean classification accuracy and the average number of eliminated channels for all methods on DS1. Here, it can be observed that mean classification accuracy is achieved when the proposed channel selection method discards 70% of the entire set. The remaining three-channel selection approaches, namely, IBGSA, RSS-SFSM, and GSO algorithms obtain their best accuracy when eliminating 54%, 59%, and 51% channels, respectively.

It means that our method could effectively identify relevant channels by correctly eliminating less important channels. Similarly, on DS2 (Fig. 5.9. (B)) and DS3 (Fig. 5.9. (C)), our algorithm achieved maximum mean classification accuracy compared to other methods by dumping 74% and 77% channels, respectively. Unlike other methods, especially SVM-RFE, our method effectively balances the tradeoff between classification accuracy and eliminated channels. The current approach shows a better channel reduction rate than other methods. In addition, the classification accuracy achieved by our method is superior to the rest of the three baseline algorithms. Overall, the current approach shows a better channel reduction rate than other existing methods. In addition, the classification accuracy achieved by our method is superior to the rest of the three baseline algorithms.

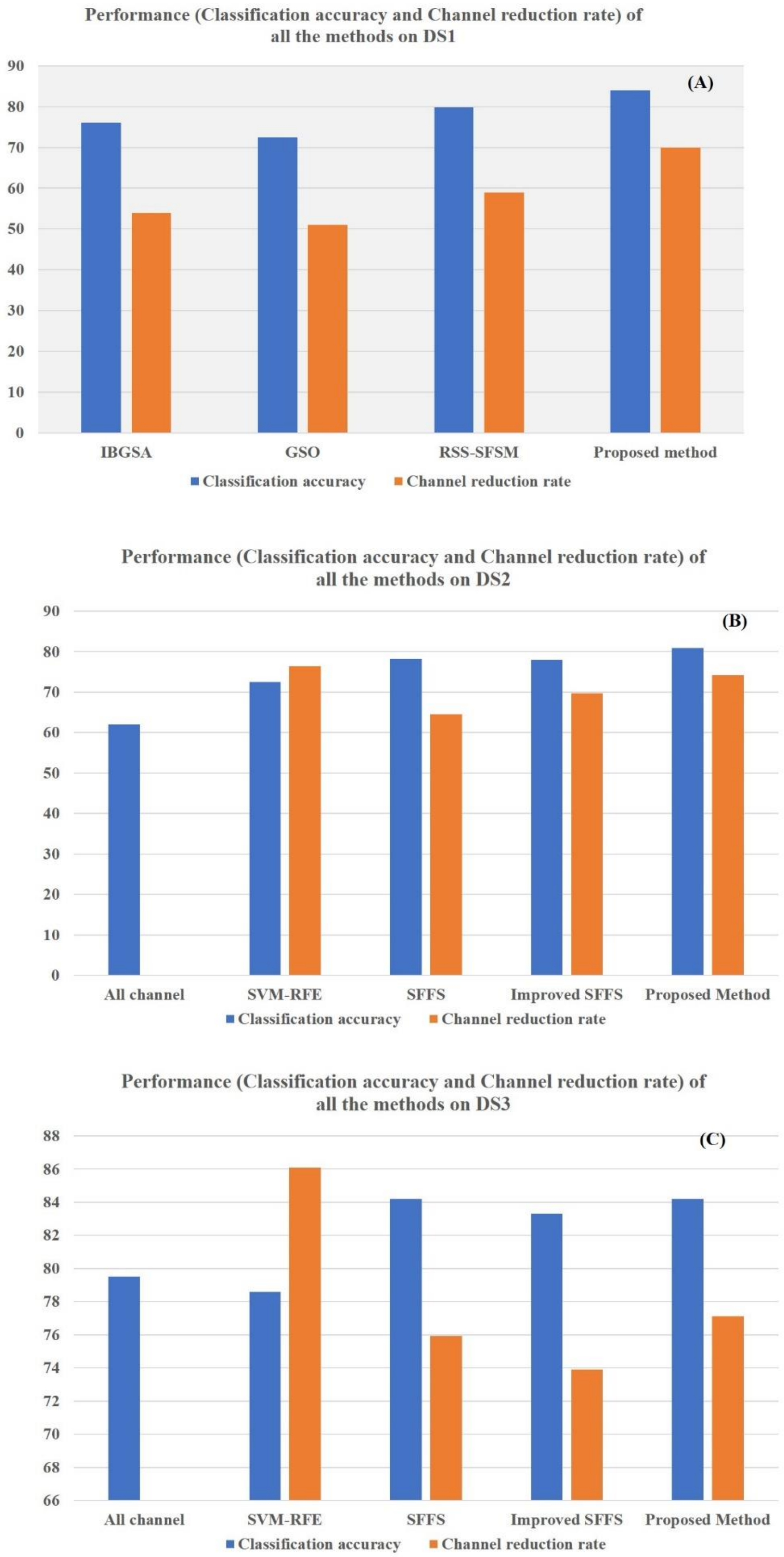


Figure 5.9. Performance of all baseline methods (classification accuracy and channel reduction rate) on Dataset 1, Dataset 2, and Dataset 3

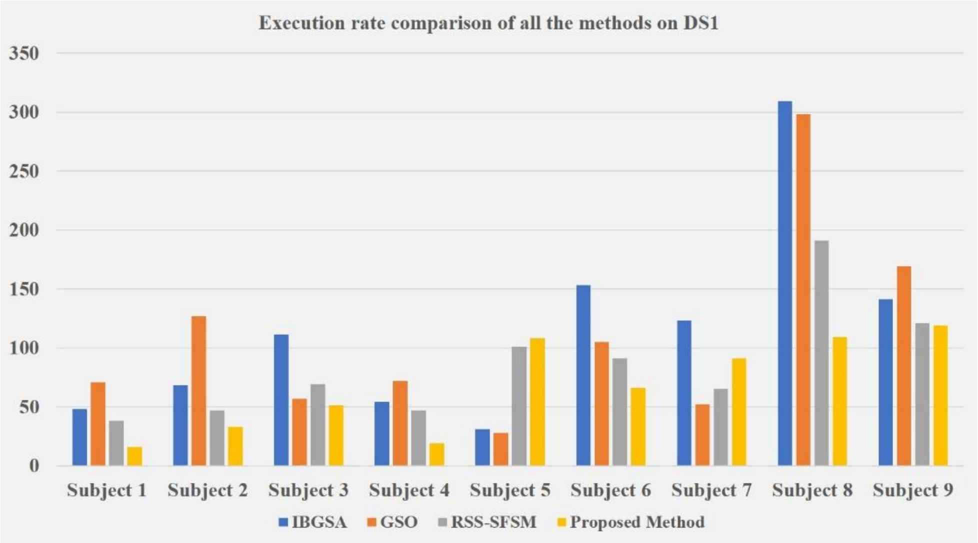
5.7.4. Time Complexity Analysis

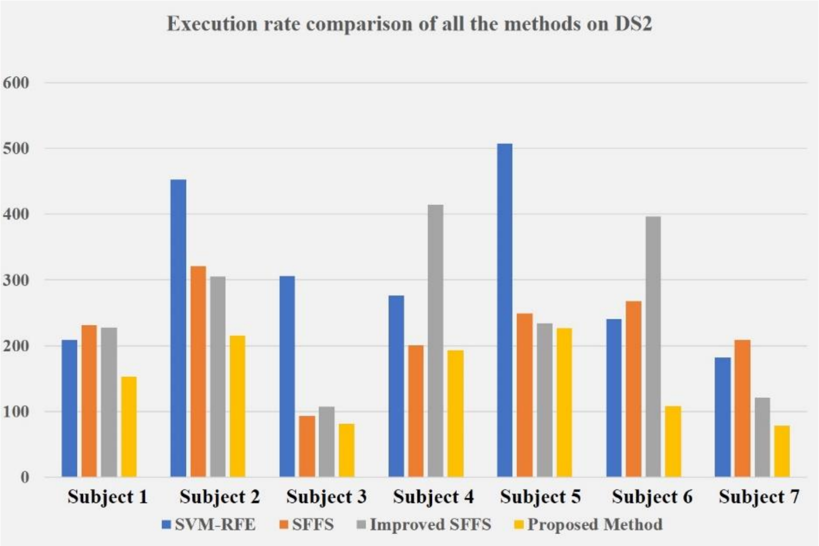
In addition to classification accuracy and channel reduction rate, execution time is another important performance indicator of algorithms. The total computation time of the proposed study mainly depends on three steps: (1) channel selection, (2) feature engineering, and (3) classification. Here, we compare all the algorithms' overall execution speeds in Figure 5.10 on all three BCI datasets. On DS1, our method's execution time is minimal compared to the state-of-the-art methods except for two participants (subjects 5 and 7). In both cases, the GSO method outperformed other methods, including the proposed one. However, our method reduced the mean computation time for all nine participants by approximately 30% (8.58 seconds) compared to the GSO method. For the other two algorithms, our approach is 43.84% and 44.22% faster than the GSO and IBGSA methods, respectively. On DS2, the current approach realized the least execution time for all seven participants compared to other competitive algorithms. Here, it reduced the average computation time by 40%, 19%, and 15%, compared to the SVM-RFE, SFFS, and Improved SFFS methods.

The comparative study of execution speeds of all the algorithms on DS3 is relatively tedious compared to DS1 and DS2. Our method showed a slower execution speed for three out of five subjects than other algorithms. Here, the Improved-SFFS algorithm achieved the least execution time among the others. For subjects four and five, our method acquired the best execution speed and hence the least running time compared to the remaining algorithms. Therefore, it can be concluded that the proposed channel selection algorithm may limit the running speed of the BCI system where a large number of channels are used for signal acquisition.

5.7.5. Results Implications

The proposed channel selection approach aims to enhance the classification accuracy of MI-based BCI systems using the optimal number of selected channels. This minimizes the BCI setup preparation time and reduces the computational complexity due to eliminating irrelevant channels from the original set. The realized results suggest that the Firefly algorithm can effectively rank the competitive entities when the candidate solution is given. We know that the main purpose of BCI systems is to detect and quantify the properties of cognitive signals based on the user's intentions and to translate them into computer-based control commands that accomplish the user's intent.





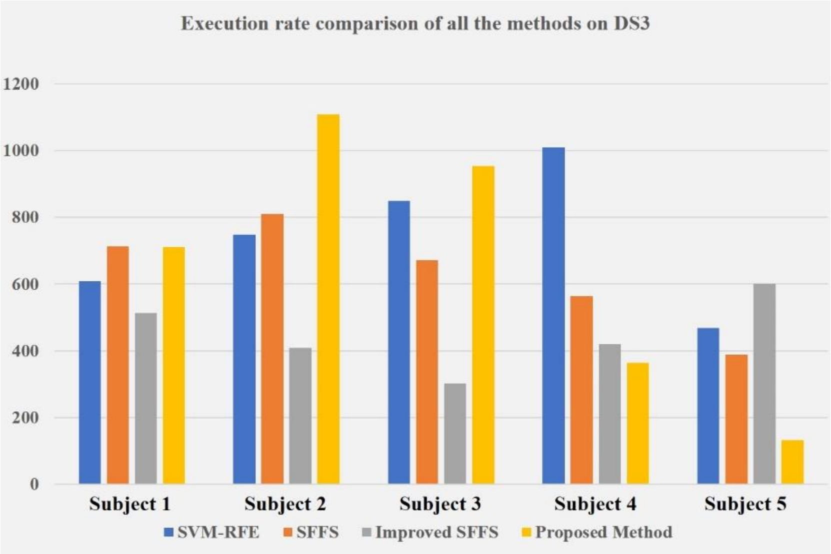


Figure 5.10. Execution time comparison for all state-of-the-art methods on respective BCI datasets

In this case, our results may help in recognizing different types of cognitive decline and motor control impairments simultaneously by comparing the properties of healthy and infected brains. Since the current method effectively ranks the channels according to their similarities with candidate channels, it can be used as a feature selection method where the selected features can be mapped with the target class or known candidate features set.

As we know that cognitive signals are high-dimensional and nonlinear by nature, our proposed approach can be useful for measuring similarity among large-scale protein sequences, which is also an important research area in machine learning applications in proteomics/bioinformatics. Some other possible applications of the proposed channel ranking method can be related to information retrieval problems such as document retrieval, collaborative filtering, sentiment analysis, and online advertising, where queries and documents are matched with each other with a relevant degree of each match. Conventionally, it is not feasible to check the relevance of all the documents with the given set of queries, and so typically, an approach called pooling is used to determine the relationship between queries and documents. Here, our method can be successfully used to find the relevance of documents corresponding to the query by transforming the query into an encoded sequence. And, the proposed firefly-based similarity measure can be used to find the relevance or similarity score between the query and the document.

5.8. Limitations of the Study

The proposed channel ranking method yielded better classification accuracy with fewer channels. Despite providing a high channel reduction rate and higher execution rate, our method has the following limitations:

I. The performance of the proposed method is primarily based on a neurophysiological basis of implanted electrodes and respective cognitive regions. In the case of missing information from these regions, the performance of the proposed method might be reduced.

II. Metaheuristic algorithms are often non-deterministic by nature. They may provide good results for one problem but may fail in similar problems. Since the conventional Firefly Algorithm is a popular metaheuristic approach, its application in other correlated research domains is limited and requires multiple application-specific improvements.

III. Cognitive data is highly dynamic and subject-dependent; therefore, the performance of the proposed method depends on the number of participants and channels. Henceforth, it would be interesting to demonstrate the performance of our approach on other datasets with a larger number of participants and channels.

IV. The performance of the proposed model eventually depends only on the local search strategy of FA that causes local optima entrapment problems with the generated solutions. It reduces the quality of the produced solutions especially when large solution space is available. In this case, the proposed approach may not be suitable for real-time systems where a large number of sensors are used in signal acquisition (for example, DS3).

V. Like existing metaheuristic channel reduction methods, our approach also seeks a single optimal solution from the set of outcomes, often ignoring the relevance of other equivalent solutions. As a disadvantage, a single solution may be accepted for a predictive model, but it may not be enough for knowledge discovery when the channel reduction approach is applied. Thus, the proposed model may be misleading in exploring the underlying mechanism of associated brain regions for which the channels are eliminated. Hence, the proposed channel reduction procedure must be carefully implemented after investigating the role of all possible solutions in search space reduction.

5.9. Conclusion & Future Scope

The proposed firefly algorithm-based channel reduction method is based on the relative closeness between the candidate channel and the remaining EEG channels. A channel with a maximum closeness score (WLBOFS score) with the candidate channel is selected as a priority as compared to other channels with minimum closeness. In our work, Spectral entropy and Lyapunov exponent features are used to design weighted linear objective functions in a weighted manner. In the preprocessing phase, signal optimization is performed using the Savitzky-Golay function, which provides a better signal approximation in terms of random spike minimization and outlier reduction. The optimized signals are used to filter the effective range of higher alpha and lower beta waves using an infinite impulse response (IIR) filter.

Further, multiple optimal channel sets are formed using the WLBOFS metric computed by the Firefly algorithm and used as different test cases for demonstration. To reduce the number of test cases, we use Fisher's information as an elimination criterion and filter those having maximum scores. Finally, RCSPA-based covariant features are extracted from selected test cases and used for classification using a fine-tuned regularized support vector machine classifier. We measure the proposed model's performance in terms of classification accuracy, number of selected channels, channel reduction rate, and execution time. Moreover, we obtain significant performance results compared to some recently published results that show the novelty of the proposed work. In the future, BCI developers can use the proposed channel reduction method with different likelihood criteria, such as Akaike's information criterion (AIC) [125] or Bayesian information criterion (BIC) [126], to minimize redundancy levels. BCI developers can attempt to demonstrate the utility of the proposed approach in the online classification problem of real-time BCI systems using EEG and MEG recording techniques.