

Chapter 6

Multi-objective X-shaped Binary Butterfly Algorithm: An Improved Wrapper-based Channel Selection Approach

6.1. Introduction

In this study, we extend the research work performed in the previous chapter in terms of solution quality improvement and making it suitable for large channel sets. We developed a Multiobjective X-shaped Binary Butterfly Optimization Algorithm (MX-BBOA) to select the most informative channels from the original set. Firstly, a fifth-order Butterworth bandpass filter [127] is used to collect relevant frequency responses, and then Independent Component Analysis (ICA) [119] is applied to remove artifacts from the filtered signals. The refined signals are further used to extract spatial-temporal features using the MEMD method. Our approach used an X-shaped transfer function to reduce continuous channel search space to binary search space. The extracted features are used to distinguish multiple MI task pairs such as left hand, right hand, tongue, and feet using the Support Vector Machine (SVM). The experiment is validated on three public EEG datasets (Dataset-1, Dataset-2, and Dataset-3). The results show that the proposed method achieved a superior classification accuracy (84.49% on Dataset 1, 79.74% on Dataset 2, and 84.55% on Dataset 3) with fewer channels than other state-of-the-art methods. In addition, the computation time compared to other published results was significantly reduced without compromising the classification accuracy. Topographical mapping between the selected channels and the cognitive regions showed that the central, frontal, and parietal lobes execute various MI tasks during physical activities. The detailed code of this chapter can be accessed from Link 1 given in the footer. The block diagram of the proposed methodology is shown in Figure 6.1.

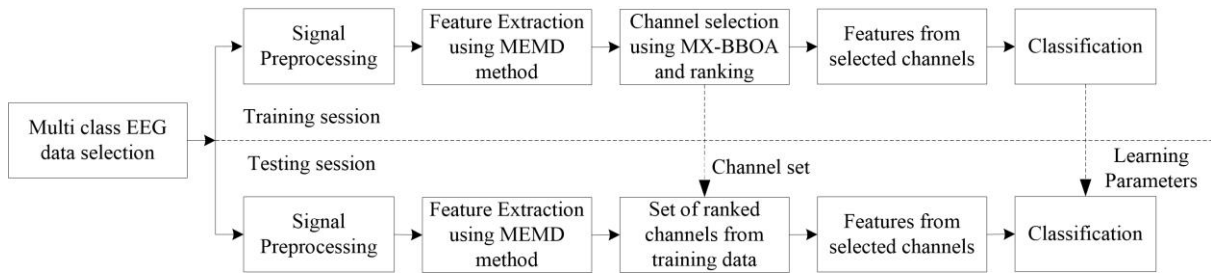


Figure 6.1. Block diagram of the proposed BCI classification framework

6.2. Limitations of the Existing Literature

In our previous work, we worked on the Firefly Algorithm (FA) to determine the optimal channel subset. Although FA based channel selection method realized good classification accuracy it suffers from the following limitations:

- I. Ignoring the global solution (local optima entrapment) because of the local search strategy.
- II. Low channel reduction rate when used with a large number of channels
- III. Premature convergence because of early-stage stability with weak solutions.

The above-mentioned limitations may degrade the quality of the computed solution and reduce the performance of the BCI classification model. Therefore, in this work, we attempt to resolve all the mentioned issues without compromising the classification accuracy.

6.3. Our Contribution

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

- I. In this study, we introduce a novel variant of the multi-objective Butterfly Optimization Algorithm (BOA) using the X-shaped transfer function to deal with the channel selection problem.
- II. To deal with the local optima problem, our X-shaped BOA computes the feasibility of two computed solutions in each iteration. If the feasibility of any of them is higher than the previous solution, then only the algorithm proceeds otherwise a mutation operator is employed to improve the quality of the solution.

- III. Similar to the FA-based channel selection approach, we implement a bi-objective function to evaluate the fitness score of the computed solution.
- IV. Based on the fitness score, we select the subject-specific optimal channel set and compute the respective classification accuracy. The global results reveal the superiority of the proposed methodology in terms of high classification accuracy, higher CRR score, and lower computational complexity over different baseline methods.
- V. The results also claim that the proposed MX-BBOA method performs better than FA when used with a large channel set.

6.4. Butterfly Optimization Algorithm: An Introduction

Butterfly Optimization Algorithm (BOA) [128] is a recently developed optimization technique that mimics the food-foraging behavior and the mating tendency of butterflies. In BOA, a butterfly generates fragrance with some intensity associated with its fitness function. When a butterfly moves from one position to another, its fragrance and respective fitness value change accordingly. In BOA, the fragrance of the i^{th} butterfly is defined as Equation 6.1.

$$f_i = cI^a \quad (6.1)$$

where f is perceived fragrance magnitude, c is sensor modality, I is the stimulus intensity, and a is the power exponent depending on modality, which shows the varying degree of absorption. For most research problems, a and c lie in the interval of $[0,1]$. There are two main steps in the BOA algorithm. i.e., the global search phase and local search phase. In the global search phase, the position vector of i^{th} the butterfly moves toward the best butterfly solution (g^*) according to Equation 6.2.

$$x_i^{t+1} = x_i + (r^2 \times g^* - x_i^t) \times f_i \quad (6.2)$$

where x_i is the position vector of the i^{th} butterfly in the iteration count ' t '. Here, g^* is the best solution among all the solutions in the current iteration, and r is a random number generated in the range of $[0,1]$. Similarly, in the local search phase, the position vector of a butterfly is changed according to Equation 6.3.

$$y_i^{t+1} = x_i + (r^2 \times x_j^t - x_k^t) \times f_i \quad (6.3)$$

where x_j^t and x_k^t are j^{th} and k^{th} butterflies from the solution search space and r is a random number generated in the range of $[0, 1]$. A switch probability p is employed to switch between global and local search spaces.

6.5. Proposed Methodology

In this study following steps are executed to determine the optimal channel subset.

6.5.1. Channel Setting

Similar to the previous experiment, we directly eliminated 3 redundant channels from DS1 because of their EOG characteristics. Therefore, only 22 channels were used in the experiment. Moreover, no changes are performed for DS2 and DS3.

6.5.2. Signal Filtering

Since the SNR of the EEG signals is relatively low and depends heavily on different muscle movement-based artifacts, it is essential to use various signal-filtering approaches to improve the SNR. Here, a fifth-order Butterworth filter is used to collect frequency responses to MI-specific neural oscillations in the range of 12 Hz to 30 Hz. This bandpass filtering discards enormous irrelevant information from the EEG spectra and reproduces clean data.

6.5.3. Artifacts Reduction

After bandpass filtering, EEG signals are analyzed to reduce different types of artifacts. Independent Component Analysis (ICA) based spatial filtering is a popular technique that procures the true statistically distinctive signals by linear decomposition of original oscillations. The crux of the ICA method is to decompose channel-wise EEG signals into an unmixing $n \times n$ square matrix W . Then, the artifacts corrected time series $Y(t)$ is computed by multiplying the matrix W with original EEG signals: $Y(t) = W^T \times X(t)$. Here, EEG data of all the channels are used for linear decomposition, and the construction of the matrix W , $X(t)$ is original time-series data sequences, and W^T represents the transpose of W . We installed the EEG lab software on MATLAB for implementing the ICA method.

6.5.4. Candidate Channel Selection

The optimal channel selection process revolves around the neighborhood of the MI-specific activation region. Therefore, it is required to consider one or multiple EEG channels as reference points to determine the ideal solution. With a reference point-guided method, it becomes simple to find an effective solution using a local search method [129]. MI-based BCI systems generally use three EEG electrodes $C3$, $C4$, and Cz as candidate channels, which record important properties of MI-specific electrical activity.

6.5.5. X-shaped Binary Butterfly Optimization Algorithm

Although the generalized BOA approach effectively finds an optimal solution for different single and multiobjective optimization problems, its performance suffers from three disadvantages: local minima avoidance, an insufficient tradeoff between exploration and exploitation strategies, and poor correlation among local optimal butterfly positions. This subsection introduces a Multiobjective X-shaped Binary Butterfly Optimization Algorithm (MX-BBOA) to address the aforementioned problems. Further, we investigate the effectiveness of the MX-BBOA in finding the optimal channel set and improving the classification accuracy of a BCI system. In this study, we iteratively applied the BOA on the normalized EEG observations to update their instances. In each iteration, two sigmoid components [130] are applied to generate two different solutions using Equation 6.4 and Equation 6.6, where Equation 6.6 is the mirror image of Equation 6.4 along the X-axis. The produced solutions are often referred to as complements of each other. We computed the performance of each solution in terms of lower fitness value. Here, the lower fitness value refers to the respective solution with high classification accuracy with less number of channels. The best solution is selected and compared with the previous one defined in Equation 6.2 and Equation 6.3.

$$S_1(x) = \frac{1}{1+e^{-x}} \quad (6.4)$$

$$x_i^{t+1} = \begin{cases} 1 & \text{if } rnd1 < S_1(x_i^{t+1}) \\ 0 & \text{Otherwise} \end{cases} \quad (6.5)$$

$$S_2(x) = \frac{1}{1+e^x} \quad (6.6)$$

$$y_i^{t+1} = \begin{cases} 1 & \text{if } rnd2 > S_2(y_i^{t+1}) \\ 0 & \text{Otherwise} \end{cases} \quad (6.7)$$

where x_i and y_i are the binary versions of input butterfly positions generated by Equation 6.5 and Equation 6.7, respectively, and $rnd1$, $rnd2$ are two random numbers in the range of $[0,1]$. We get the best position of the butterfly in the given solution space by comparing the fitness score of both binary versions shown in Equation 6.8.

$$f_i^{t+1}(new) = \begin{cases} y_i & \text{if } fitness(y_i) > fitness(z_i) \\ x_i & \text{Otherwise} \end{cases} \quad (6.8)$$

Now, if the $fitness(f_i^{t+1}(new)) > fitness(f_i^t)$, then the $f_i^{t+1}(new)$ is updated as a new solution. Otherwise, a uniform crossover operation is applied between $f_i^{t+1}(new)$ and f_i^t to improve the quality of the solution. During crossover, a crossover point (channel 13) is selected, and binary vectors beyond this point are exchanged between parents to form new individuals. Here, the Crossover operation promoted the exploration of the solution space by combining beneficial traits from both parents, aiding in the evolutionary optimization process. This step results in two child solutions where the best one is selected as the next solution. Also, it increases solution diversity in terms of the best fitness score computed by parent butterfly and child positions that help avoid getting trapped in local optima. This step is iteratively implemented and solutions with minimum channels were selected. Finally, features from the selected channels were used to realize respective classification accuracy. This approach provided a specific interval for different classification accuracy levels computed during different iterations. In this interval, the lower bound indicated minimum classification accuracy and the upper bound showed the maximum one. A detailed description of the MX-BBOA is presented in Figure 6.2.

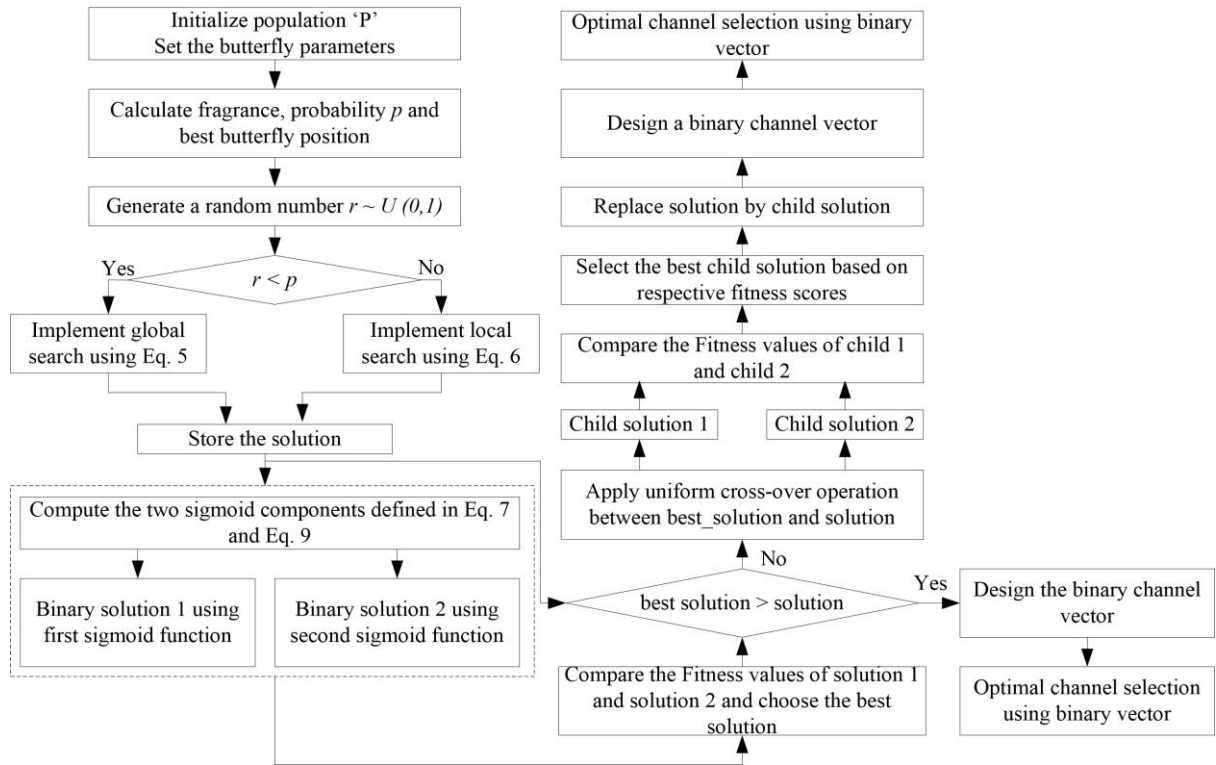


Figure 6.2. Working on the proposed X-shaped binary butterfly algorithm

6.5.6. Fitness Function

In the study, the MX-BBOA approach determines the best channels by optimizing a weighted fitness function with two conflicting objectives: obtaining maximum classification accuracy with the relative number of selected channels. These two objective functions are linearly combined and used as the fitness function $f(x)$, as defined in Equation 6.9.

$$f(x) = w_1 f_1(x) + w_2 f_2(x) \quad (6.9)$$

where $f_1(x)$ and $f_2(x)$ denote the classification error rate and relative number of selected channels, and the weights w_1 and w_2 are normalized in the range of $[0,1]$. The error rate is computed from a 5-fold cross-validation of training samples. The relative number of channels is derived by dividing the number of selected channels in the X-BBOA vector by the total number of available channels.

6.6. Feature Engineering

In this study, we use the MEMD feature extraction algorithm on the segmented chunks of Beta oscillations. If the size of the covariance matrix is given by (m, n) the total number of extracted features will be $m \times n$. To reduce the range of the computed features, columns of the MEMD features are normalized between the range of $[0, 1]$ by employing *min-max* normalization defined in [Equation 6.10](#).

$$\hat{x}_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (6.10)$$

6.7. Experimental Results

Three classifiers, namely, (1) Support Vector Machine, (2) Naïve Bayes, and (3) Decision Tree are used to discriminate subject-specific normalized MEMD features. The experimental results were implemented with Python 3.6 on a laptop with an Intel Core i5, 2.4GHz CPU, 8 GB memory, and 512 GB hard disk with a Windows 10 operating system. A detailed description of the results from each dataset is given in the subsequent subsections.

6.7.1. Parameter Tuning

In our work, a five-fold cross-validation approach is used on datasets 2 and 3 to test the effectiveness of the machine learning model and avoid overfitting issues. In the first iteration, 80% of feature vectors are used for training, and the remaining 20% are employed for testing purposes. In the next iteration, another 20% of feature vectors are used for testing, and the rest 80% are employed for training. This process is repeated until all the feature vectors are used for testing the proposed algorithm. All the data instances are normalized in the interval of 0 and 1. To quantify results statistically, each fold is repeated 30 times, and every experiment is continually performed 100 times, giving a total of 15,000 runs for each dataset. Next, the predictive classification model is developed on the training data and validated on the testing data, and the results are computed. Finally, the results are averaged over all the folds and compared with state-of-the-art methods. All parameter settings for the proposed algorithm are given in [Table 6.1](#). The proposed channel selection method's effectiveness is described in terms of the fitness function

defined in Equation 6.9. Here, w_1 and w_2 were used to balance the tradeoff between the classification error rate and the relative set of selected channels; therefore, it is important to determine the optimal pair of weights for the fitness function. In Figure 6.3, we have shown the effect of w_1 on the number of selected channels and classification accuracy on DS1. As it is clear, the maximum classification accuracy was computed by tuning $w_1 = 0.74$ and $w_2 = 0.26$. Therefore, we use this parameter pair for all three datasets in respective fitness function optimization. This parameter set shows that classification accuracy is more sensitive than the number of selected channels. It is also noticeable that all three classifiers achieve their highest fitness score at this parameter set. The tuned parameters were directly employed to train all three classifiers on the input data of all the participants, and maximum classification accuracy was recorded.

Algorithm	Switch probability (p)	sensory modality (c) [min-max]	power exponent (a)	SVM parameters
MX-BBOA	0.50	[0.01–0.50]	0.50	$C = 0.01, \gamma = 100$

Table 6.1. Values of the parameters used in the proposed MX-BBOA

6.7.2. Average Size of the Optimal Channel Set

In terms of optimal channel set size, maximum classification accuracy is achieved when the proposed channel selection scheme discards 60% of the original channels. As shown in Figure 6.4, the proposed algorithm could effectively determine the relevant channels by discarding many redundant and irrelevant channels. It can be noticed that the optimal size of the channel set is eight, which provides maximum classification accuracy. Interestingly, the Support Vector Machine (SVM) achieves maximum classification accuracy on validation data of A01, A02, A04, A05, A06, and A07 participants while Naïve Bayes and Decision Tree gain higher classification accuracy on A03, A08, and A09, respectively.

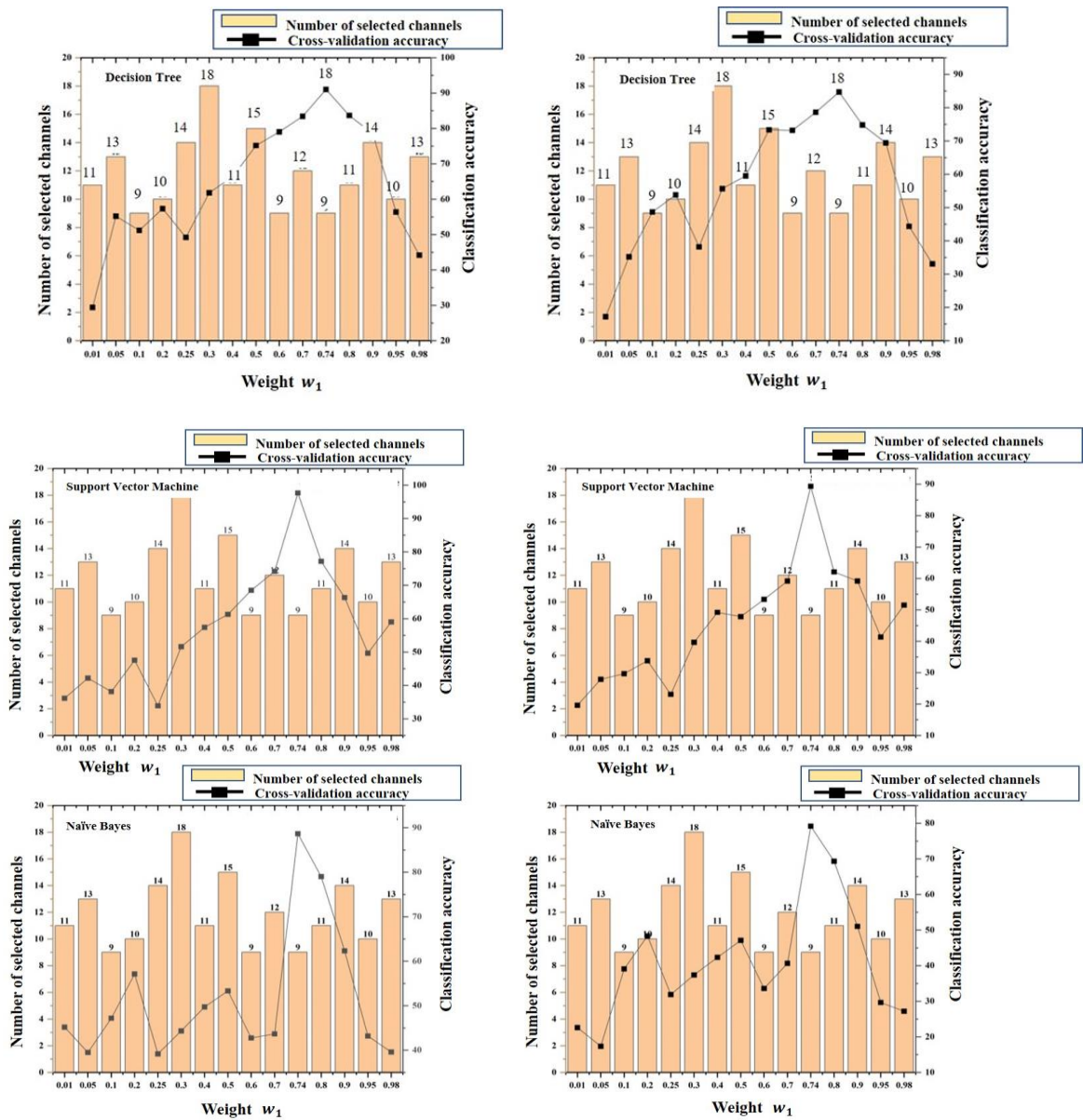


Figure 6.3. Effect of weight (w_1) on the selected channels and 5-fold cross-validation classification accuracy on dataset 1

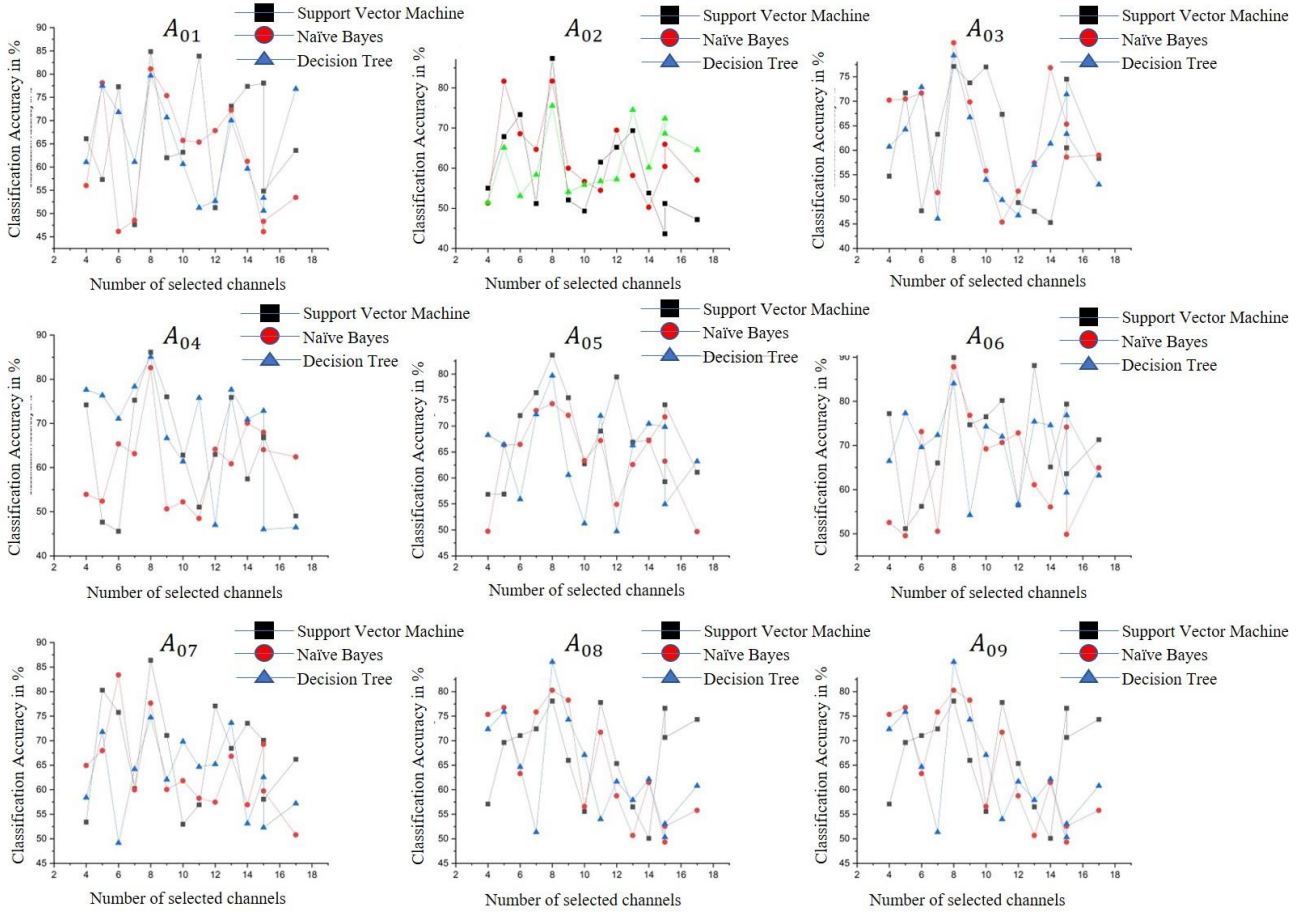


Figure 6.4. Graphs showing the effectiveness of the proposed approach for all nine participants (A01 – A09) of dataset 1.

6.7.3. Results Comparison with Baseline Methods

In this subsection, we compared the performance (average classification accuracy and the number of selected channels) of the proposed MX-BBOA algorithm with different state-of-the-art methods. Here, we have avoided the use of deep-learning models for the comparison of results because of their high computational complexity and dependency on a large amount of data to give better results than machine-learning techniques. Since, the size of DS2 and DS3 is small in terms of the number of participants (seven and five, respectively), therefore it was impractical to explore the popular deep learning models such as Convolution Neural Networks (CNNs), Long Short Term Memory (LSTM), and Recurrent Neural Networks (RNNs) on all three BCI datasets.

Similar to previous chapters, the computational complexity of the proposed methodology depends on four steps: (1) EEG preprocessing, (2) Channel selection using BOA, (3) Feature engineering, and (4) Classification. Here, the preprocessing involves Butterworth filtering which is a linear

filter and its overall computational complexity depends on the length of the input signal. If the input signal has a length of N , the complexity can be expressed as $O(kN)$, where k refers total number of samples in the input signal. In the channel selection, BOA has two “For” loops but they are not interlinked therefore, it is also a linear process with a complexity level of $O(k)$. In the feature engineering phase, the performance of MEMD is dominated by the calculation of the Hilbert transform which results in a time complexity of $O(n^m * d)$, where n is the number of data points, m is the number of IMFs, and d is the number of dimensions. In this experiment, we considered the top five IMFs in the classification, Finally, the classification step with the SVM technique will execute in $O(n^2)$ (LibSVM). Therefore, the overall complexity can be represented as an algebraic sum of all the complexity levels which is $O(kN + k + n^5 * d + n^2)$. This complexity level can be reviewed in terms of maximum complexity $O(n^5 * d)$.

6.7.3.1. Test Results on Dataset 1

In the comparison of the results, a pairwise binary classification approach is used to discriminate multiple MI activities. Since there are four MI classes, six possible pairs (4C2): left vs. right (L vs. R), left vs. foot (L vs. F), left vs. tongue (L vs. T), right vs. foot (R vs. F), right vs. tongue (R vs. T), and foot vs. tongue (F vs. T) are used in classification. [Table 6.2](#) and [Table 6.3](#) present the overall classification results in terms of selected channels and classification accuracy for all MI-task pairs on the validation data of dataset1 using different classifiers. For dataset 1(DS1), the classification results are compared with three-channel selection algorithms: (1) Sparse common spatial pattern (SCSP) based channel selections method, (2) Improved binary gravitation search algorithm (IBGSA), and (3) Nondominated sorting genetic algorithm II (NSGA). In the SCSP approach, two sparse filters were used to maximize weight sparsity between channels selected by the CSP projection matrix. Further, two selection criteria: (1) maximum channel irrelevance and classification accuracy, and (2) reducing the size of the selected channel set while maintaining classification accuracy were used for channel optimization. Finally, the SCSP algorithm with first and second selection criteria are, respectively, abbreviated as SCSP1 and SCSP2. The authors reported that the proposed channel selection approach achieved a mean classification accuracy of 81.63% with SCSP1 and 79.07% with SCSP2 approaches.

Table 6.2. Classification performance of different classifiers for three MI-task pairs (L vs. R, L vs. F, and L vs. T)

Subjects	Selected Channels	CRR	L vs. R			L vs. F			L vs. T		
			SVM	NB	DT	SVM	NB	DT	SVM	NB	DT
1	1, 3,4,5, 8,9,10, 12,14, 17,21	0.50	84.33	66.30	82.61	97.54	78.35	76.26	96.20	79.40	71.33
2	3,4,8, 9,10,12, 13, 19, 22	0.59	78.20	69.24	73.95	88.33	81.22	74.02	80.02	72.48	62.60
3	2,3,4,6,8,10,12,13,18	0.59	91.33	81.29	66.10	92.08	72.17	69.20	98.42	63.33	66.20
4	3,4,7,8,10,11,12,14,17	0.59	83.02	69.54	70.14	81.30	65.35	72.83	81.33	74.10	78.12
5	1,3,4,5, 8,10, 11,12,18,21	0.54	87.66	81.42	81.47	85.20	80.51	68.70	91.70	60.12	62.40
6	1,3,4,6, 8, 10,12,13, 17,21	0.54	80.04	65.98	74.45	72.80	82.89	72.92	87.42	66.20	77.42
7	1,3,4,8,10,12,15, 16, 20	0.59	82.30	84.06	72.41	69.10	72.98	73.13	68.33	81.33	70.33
8	3,4, 6, 7, 8,9,10,12,14, 17, 21	0.50	94.28	77.48	74.60	83.66	82.23	67.70	77.20	58.50	53.66
9	3,4, 6, 8,9,10,12, 13, 18, 19,21	0.50	79.33	78.06	81.89	77.60	65.07	75.53	85.69	70.02	67.12
Mean	3,4, 6,8 9, 10,12, 13, 17, 21 (Highly voted channels)	0.54	84.49	74.81	75.29	83.06	75.64	72.25	85.14	69.49	67.68

Table 6.3. Classification performance of different classifiers for three MI-task pairs (R vs. F, R vs. T, and F vs. T)

Subjects	Selected Channels	CRR	R vs. F			R vs. T			F vs. T		
			SVM	NB	DT	SVM	NB	DT	SVM	NB	DT
1	1, 3,4,5, 8,9,10, 12,14, 17,21	0.50	82.40	70.68	72.68	74.62	78.20	66.30	83.20	63.30	78.33
2	3,4,8, 9,10,12, 13, 19, 22	0.59	62.40	76.83	70.50	84.68	79.42	69.24	89.66	78.12	81.33
3	2,3,4,6,8,10,12,13,18	0.59	79.33	71.89	69.04	97.20	81.50	81.29	80.40	59.33	77.48
4	3,4,7,8,10,11,12,14,17	0.59	93.69	73.84	74.66	92.40	87.68	69.54	85.20	81.48	72.80
5	1,3,4,5, 8,10, 11,12,18,21	0.54	84.42	80.64	68.18	87.50	61.88	81.42	96.33	86.12	80.50
6	1,3,4,6, 8, 10,12,13, 17,21	0.54	91.30	78.18	76.22	67.54	73.33	65.98	93.18	80.14	68.70
7	1,3,4,8,10,12,15, 16, 20	0.59	85.66	81.35	66.94	82.70	59.12	84.06	71.50	77.60	78.66
8	3,4, 6, 7, 8,9,10,12,14, 17, 21	0.50	88.33	80.71	73.46	90.33	72.30	77.48	68.91	70.33	77.42
9	3,4, 6, 8,9,10,12, 13, 18, 19,21	0.50	78.42	69.33	78.27	75.38	83.66	78.06	82.20	74.42	81.47
Mean	3,4, 6,8 9, 10,12, 13, 17, 21 (Highly voted channels)	0.54	82.88	75.93	72.21	83.59	75.23	74.81	83.39	74.53	77.41

In the second work, IBGSA automatically detected more informative channels in the left or right-hand classification. The IBGSA algorithm is a binary version of the generalized Gravitational Search Algorithm (GSA) that explores gravity and the Newtonian law of motion. Here, time and wavelet features were extracted to discriminate four MI tasks using an SVM classifier. The classification results confirm that the based channel selection method achieved a maximum of 80% and an average of 76.24% classification accuracy with an average number of 8.6 EEG channels. Another work based on NSGA, a popular variant of the conventional Genetic algorithm (GA), is applied with the CSP features and a maximum of 82.60% classification accuracy with an average number of 11 channels. A detailed comparison of all the baseline approaches and proposed methodology is given in [Table 6.4](#).

Table 6.4. Comparison of classification accuracy (in %) for L Vs. R task of the proposed channel selection method with state-of-the-art-methods on dataset 1

Subject	(MX-BBOA)		(NSGA II)		(SCSP1)		(SCSP2)		(IBGSA)	
	channels	L vs. R	channels	SVM	channels	SVM	channels	SVM	channels	SVM
1	11	84.33	9	81.52	13	91.66	13	91.66	9	76.66
2	9	78.20	15	71.20	9	67.36	4	60.41	6	76.66
3	9	91.33	8	94.18	14	97.91	12	97.14	7	73.33
4	9	83.02	13	77.40	14	72.22	11	70.83	8	73.33
5	10	87.66	10	82.33	11	65.27	9	63.19	11	80
6	10	80.04	9	70.54	14	66.67	10	61.11	12	73.33
7	9	82.30	13	83.66	19	84.72	15	78.47	15	76.66
8	11	94.28	15	89.02	15	97.22	5	95.13	11	80
9	11	79.33	7	93.55	10	91.66	5	93.75	-	-
Mean	9.88	84.49	11	82.60	13.22	81.63	8.55	79.07	8.6	76.24

6.7.3.2. Test Results on Dataset 2

[Table 6.5](#) presents the comparative performance (number of channels and achieved classification accuracy) between the proposed MX-BBOA algorithm and three-channel selection methods on test data of all subjects. The last row of the table shows the p -value computed from the paired t -test of the results of all channels and state-of-the-art methods ([refer to chapters 4 and 5](#)). With the proposed MX- BBOA approach, the average classification accuracy was improved by 28.61% compared to a method where all channels were used in the classification approach. We compared our results with the Improved SFFS method in the second case and obtained better classification accuracy with a 2.2% improvement across all nine subjects. Our method showed a slight improvement of 1.8% in

classification accuracy compared to the SFFS method in the third case. Here, we obtained a substantial gain of 22.33% in the channel reduction rate, which showed that our approach effectively reduces irrelevant electrodes from the original channel set. Our proposed method accomplished superior classification accuracy with a 9.98% improvement rate over the SVM-RFE method. However, our method employed more channels to obtain this performance gain. Interestingly, our classification results are statistically significant ($p = 0.0015$) compared to all state-of-the-art methods.

Table 6.5. Performance comparison between different channel selection methods on dataset 2

Subject	All channels Acc.(%)	Proposed method		Improved SFFS		SFFS		SVM-RFE	
		Acc.(%)	channels	Acc.(%)	channels	Acc.(%)	channels	Acc.(%)	channels
A	43	73.10	11	69	6	60	9	57	4
B	42	67.50	9	63	15	66	19	54	2
C	62	83.60	22	87	26	91	21	84	32
D	81	91.66	26	94	29	94	34	88	26
E	91	97.20	21	96	19	96	21	95	10
F	49	61.80	9	65	8	58	19	58	4
G	66	83.33	14	72	22	83	21	71	18
Mean	62	79.74	16	78.0	17.9	78.3	20.6	72.5	13.7
STd	18.9	11.89	6.41	14.0	8.7	16.5	7.3	16.6	11.8
<i>p</i> -value		0.0015	--	0.0025	--	0.002	--	0.0045	--

6.7.3.3. Test Results on Dataset 3

The results reported in [Tables 6.6](#) and [6.7](#) showed the performance of the proposed MX-BBOA algorithm with seven-channel selection algorithms on the evaluation data of all five subjects. Unlike datasets 1 and 2, these results are computed from a dense electrode setup of 118 channels. The average classification accuracy of our method was 84.55% and obtained an improvement of 6.3% by considering only 24 out of 118 channels. In the second comparison, the proposed approach achieved a slight gain of 1.5% in classification accuracy but offered a better channel reduction rate ($\approx 80\%$) than the [Improved SFFS](#) method. Although the average classification accuracy of the SFFS method is similar to the proposed MX-BBOA, it requires a relatively large number of channels. Compared with the SVM-RFE, our method showed a low channel reduction rate and obtained superior classification accuracy. However, the *p*-value computed from the classification results of the [SVM-RFE method](#) is low. It indicates the poor performance of the SVM-RFE method compared to the generalized method where all channels were used in the

classification process (Table 6.7-column 2). The 3C-CSP method [131] employs the CSP features from only three candidate channels in the MI recognition. Since this method depends on only candidates and avoids the relevance of other channels, it achieves inferior classification accuracy compared to all the channel selection algorithms. The only advantage of the 3C-CSP method is fast BCI preparation time because of the fewer channels. The CSP rank is another channel selection method that uses sorted CSP filters to separate different MI tasks. In this case, our method improved by 3.36% in classification accuracy by using only 24 channels with a 3.38% better channel reduction rate.

The Rayleigh Coefficient-based Genetic Algorithm RC-GA method [132] is a hybrid method that uses the Rayleigh coefficient maximization-based Genetic algorithm to filter the best channels from the original channel set. In this case, our method showed better performance in terms of both classification accuracy and channel reduction rate. Correlation coefficient-based channel selection (CCS) [131] is a recently published approach that selects highly correlated electrodes to improve classification accuracy. This method achieved slightly better classification accuracy than our method but used four more EEG channels, therefore, showing a poor channel reduction rate. However, this method has obtained a maximum p -value because of the subject-wise classification accuracy rate variability.

Table 6.6. Comparison of the proposed MX-BBOA method with state-of-the-art methods on dataset 3

Subject	All channels Acc. (%)	Proposed method		Improved SFFS		SFFS		SVM-RFE	
		Acc. (%)	channels	Acc. (%)	channels	Acc. (%)	channels	Acc. (%)	channels
aa	75.7	78.12	23	76.4	27	78.3	26	68.6	7
al	85.7	92.40	33	94.3	47	93.6	43	92.9	8
av	62.1	71.88	11	65.0	18	68.6	18	62.9	14
aw	87.1	92.33	24	89.5	27	87.9	20	80.0	35
ay	87.1	88.02	29	91.4	35	92.7	35	88.6	18
Mean	79.5	84.55	24	83.3	30.8	84.2	28.4	78.6	16.4
STd	10.9	8.19	7.42	12.3	10.9	10.6	10.5	12.8	11.3
p -value		0.032		0.047		0.023		0.75	

Table 6.7. Comparison of the proposed method with remaining state-of-the-art methods on dataset 3

Subject	Proposed method		3C-CSP		CSP-rank		RC-GA		CCS-CSP	
	Acc. (%)	channels	Acc. (%)	channels	Acc. (%)	channels	Acc. (%)	channels	Acc. (%)	channels
aa	78.12	23	63.6	3	80.0	40	78.6	39	80.7	22
al	92.40	33	85.4	3	94.6	24	91.4	18	96.8	34
av	71.88	11	59.6	3	53.6	10	55.4	47	70.4	13
aw	92.33	24	80.0	3	88.6	25	89.6	26	92.9	14
ay	88.02	29	91.1	3	92.1	43	89.6	16	92.1	59
Mean	84.55	24	75.9	3	81.8	28	80.9	29	85.6	28
STd	8.19	7.42		0		11.97		12.02		17.04
<i>p</i> -value	0.032		0.264		0.487		0.548		0.003	

Dataset 3 is one of the largest BCI repositories in the number of channels. Since we have compared the performance of the proposed method with seven existing channel selection algorithms; therefore, it is obvious to investigate the time complexity of all the applied approaches over the entire solution space. In [Figure 6.5](#), we compared the execution speed of all state-of-the-art models with our MX-BBOA scheme when the number of channels increases. All algorithms are compared in terms of execution time (T) and the number of selected channels (N). In [Figure 6.5. A](#), for a few channels ($N < 10$), the speed of Improved SFFS and the SFFS were better than the proposed method, but their performance deteriorates with the number of channels. In addition, our method is almost two times faster than the SVM-RFE method and 1.5 times more efficient than the rest of the methods. Finally, the execution order of all the algorithms on dataset 3 was MX-BBOA > Improved SFFS > SFFS > SVM-RFE. In [Figure 6.5.B](#), the speed of the MX-BBOA is compared with the rest of the three-channel selection methods. Here, the execution rate of CCS-CSP was the worst among the rest of the channel selection methods. Since the CCS-CSP algorithm computes the correlations between every pair of channels which requires $O(N^2)$ comparisons, the speed of CCS-CSP alleviated with the number of channels. The CSP rank and the RC-GA execution rate were very competitive because both algorithms use similar statistical filters to rank the EEG channels. However, the execution sequence of all the applied algorithms was MX-BBOA > RC-GA > CSP-rank > CCS-CSP.

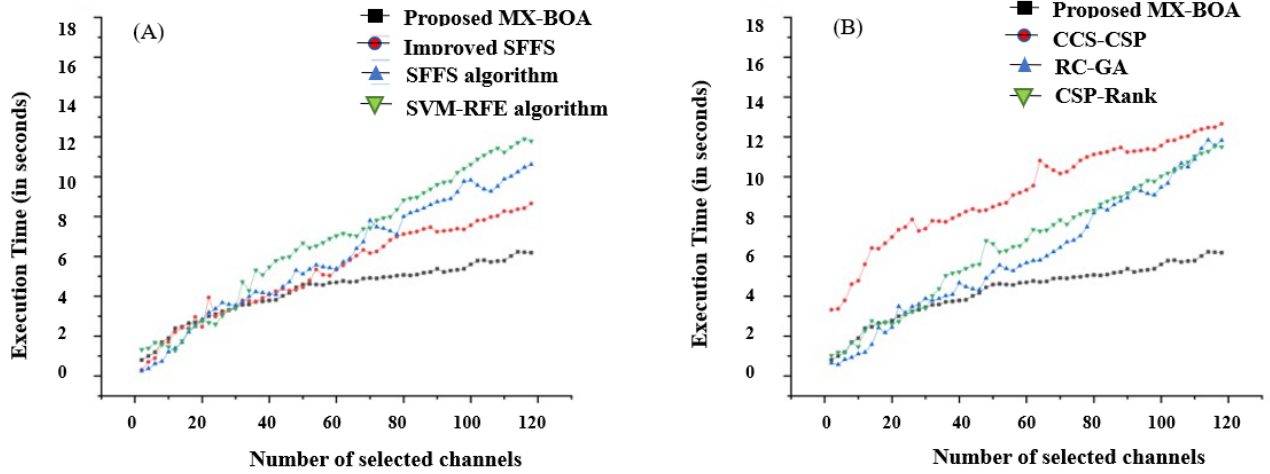


Fig. 6.5. Execution speed comparison between existing methods and the proposed approach for dataset 3

6.8. Discussion

6.8.1. Topographical Distribution of Selected Channels

We have shown the topographical distribution of the selected channels on DS1, DS2, and DS3 in Figure 6.6, Figure 6.7, and Figure 6.8. In both datasets, the selected channels are marked with different colors. For all three datasets, $C3$, $C4$, and Cz channels are selected from the central or motor cortex and used as a candidate solution. These channels are implicitly present in every optimal subset. Mathematically, other channels are selected in the neighborhood of the candidate channel using the proposed channel selection method. We can observe that selected channels are symmetrically distributed in the vicinity of the central horizontal axis where the candidate channels were located. In DS1, for four out of nine participants (A04, A07, A08, and A09), the selected channels (shown in the blue color) are equally distributed in the frontal and parietal lobe of the human brain, while for another four participants (A01, A03, A05, and A06), the selected channels are mainly distributed in the frontal lobe of the cerebrum.

Only in the case of A02, the parietal region was more active than the frontal lobe in the execution of the MI tasks. Finally, the set of frequently selected channels was almost equally distributed along with the motor cortical of the brain. In this

case, the temporal lobe was the least active during the imagination of motor activities. Although DS2 and DS3 are relatively larger datasets, the channel distribution is similar to DS1. It shows that the proposed method effectively correlates relevant EEG channels with priorly selected electrodes that chiefly distinguish cognitive zones related to motor imagery tasks. In addition, our approach consistently recognizes the active regions of the brain independent of the size of the original channels set. Thus, we can conclude that the performance of the proposed MX-BBOA is neurophysiological meaningful and can be effectively used in MI-based BCI systems.

6.9. Limitations of the Study

Despite obtaining superior performance of the MX-BBOA algorithm on three varied datasets, we deduced the following limitations:

- I. The performance of Metaheuristic algorithms is always application-dependent. One may compute slightly better than the other in a specific optimization problem and worse in another type of problem. Therefore, a strong comparison is required before applying it to a particular problem. This procedure requires a lot of study about the behavior of different algorithms, which is time-consuming and tedious.
- II. Metaheuristic algorithms are non-deterministic and iterative; therefore, the solution obtained after each run may not be similar. So, it is unfair to compare the performance of the MX-BBOA with other algorithms after the solutions are evaluated from the first few runs.
- III. The performance of our method depends on the prior selection of candidate channels. Therefore, researchers must know the neurophysiology of the human brain. It implies that in the absence of true candidate channels, the performance of the MXBBOA may degrade.
- IV. Our classification model is validated only on motor imagery datasets. Therefore, its scope is limited only to voluntary limb movements. Extended research is

required by our approach in classifying other correlated problems such as visual skills, emotion detection, and language interpretation.

- V. Although we have validated our approach on three BCI datasets varying in the EEG channels, the number of participants and dataset size is very limited, resulting in a lack of sufficient persuasiveness. Since neural oscillations are subject-specific and non-deterministic, the proposed method may not provide a global solution.

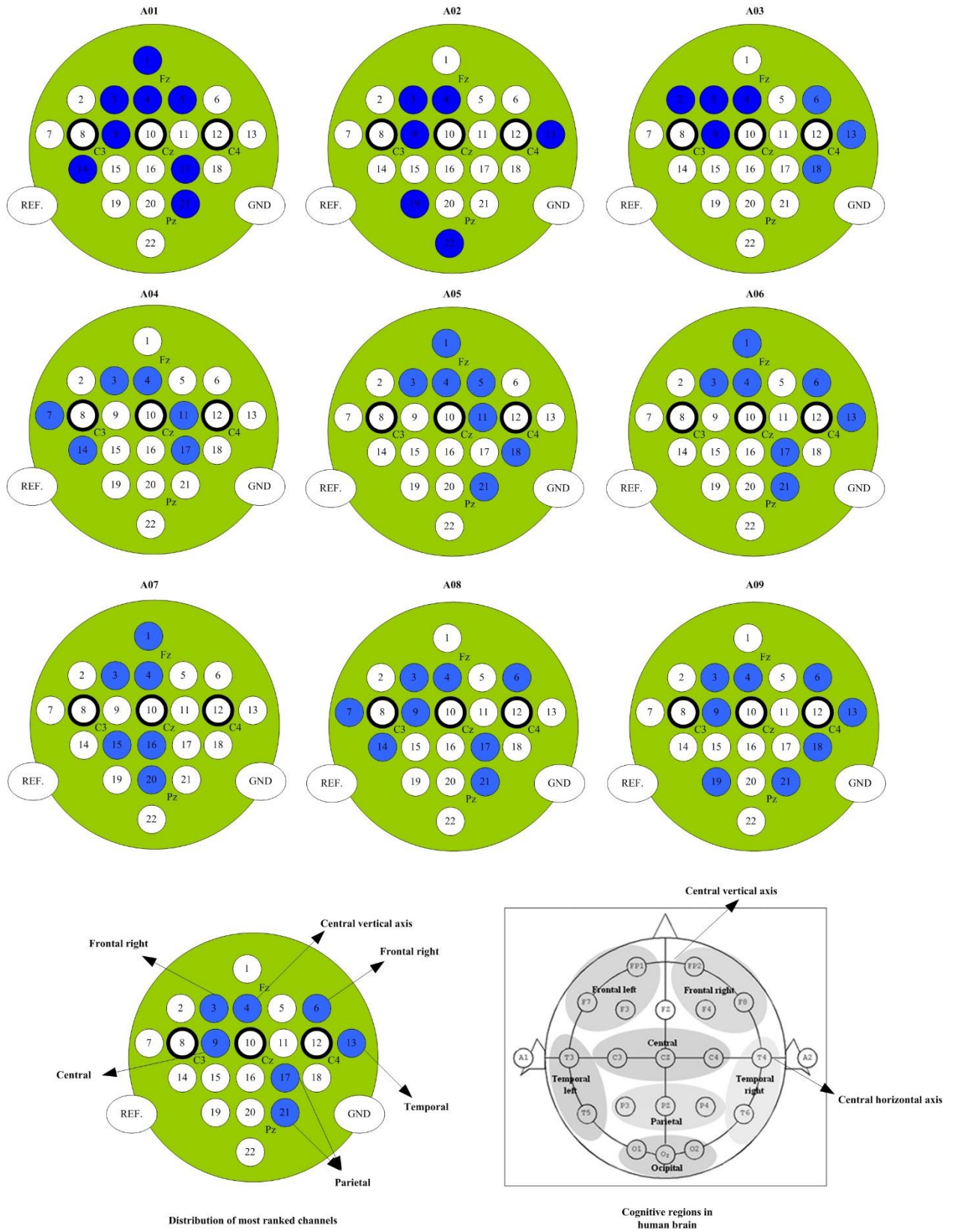


Figure 6.6. Topographical distribution of selected channels in Dataset 1

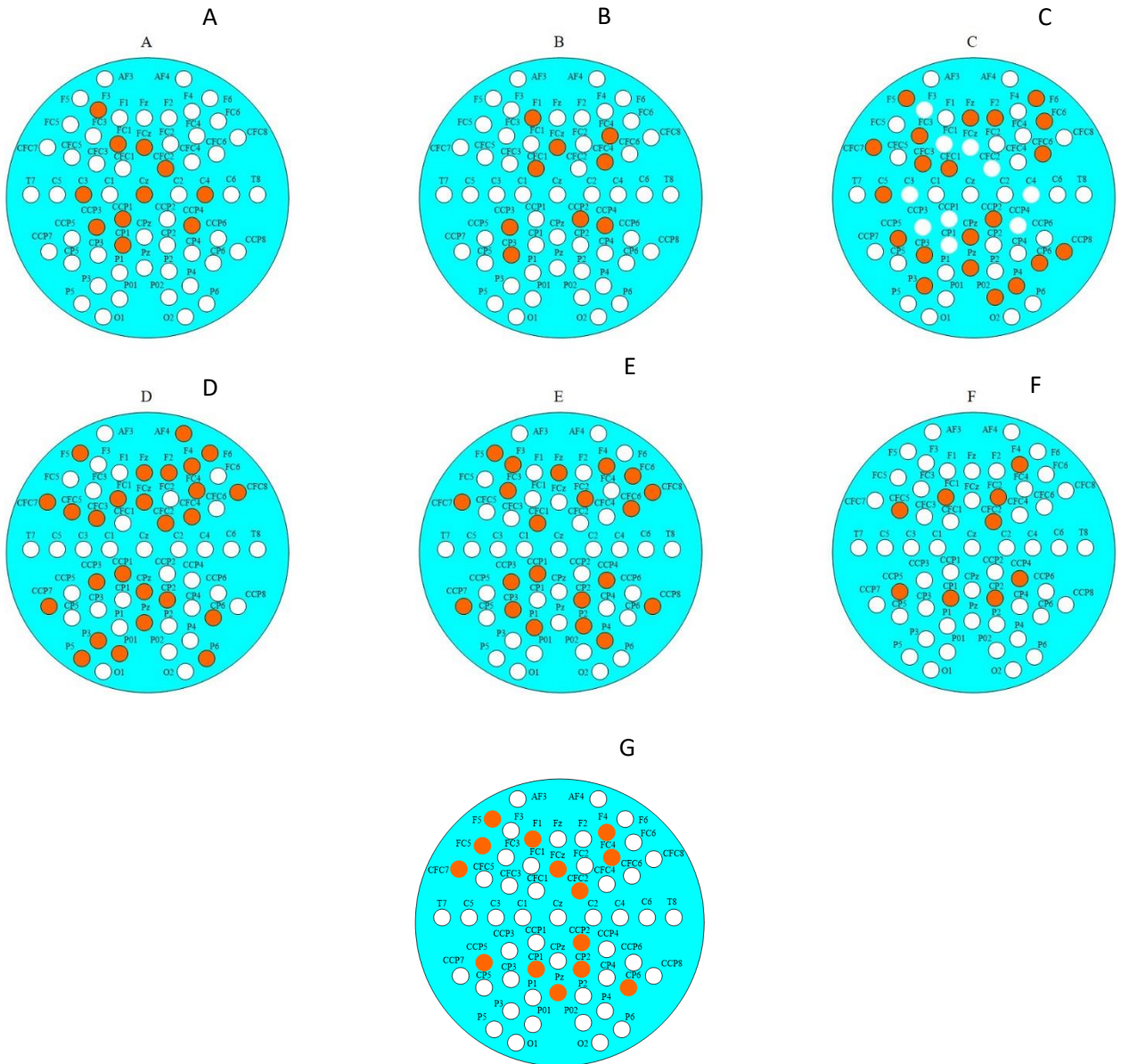


Figure 6.7. Topographical space visualization of selected channels in dataset 2

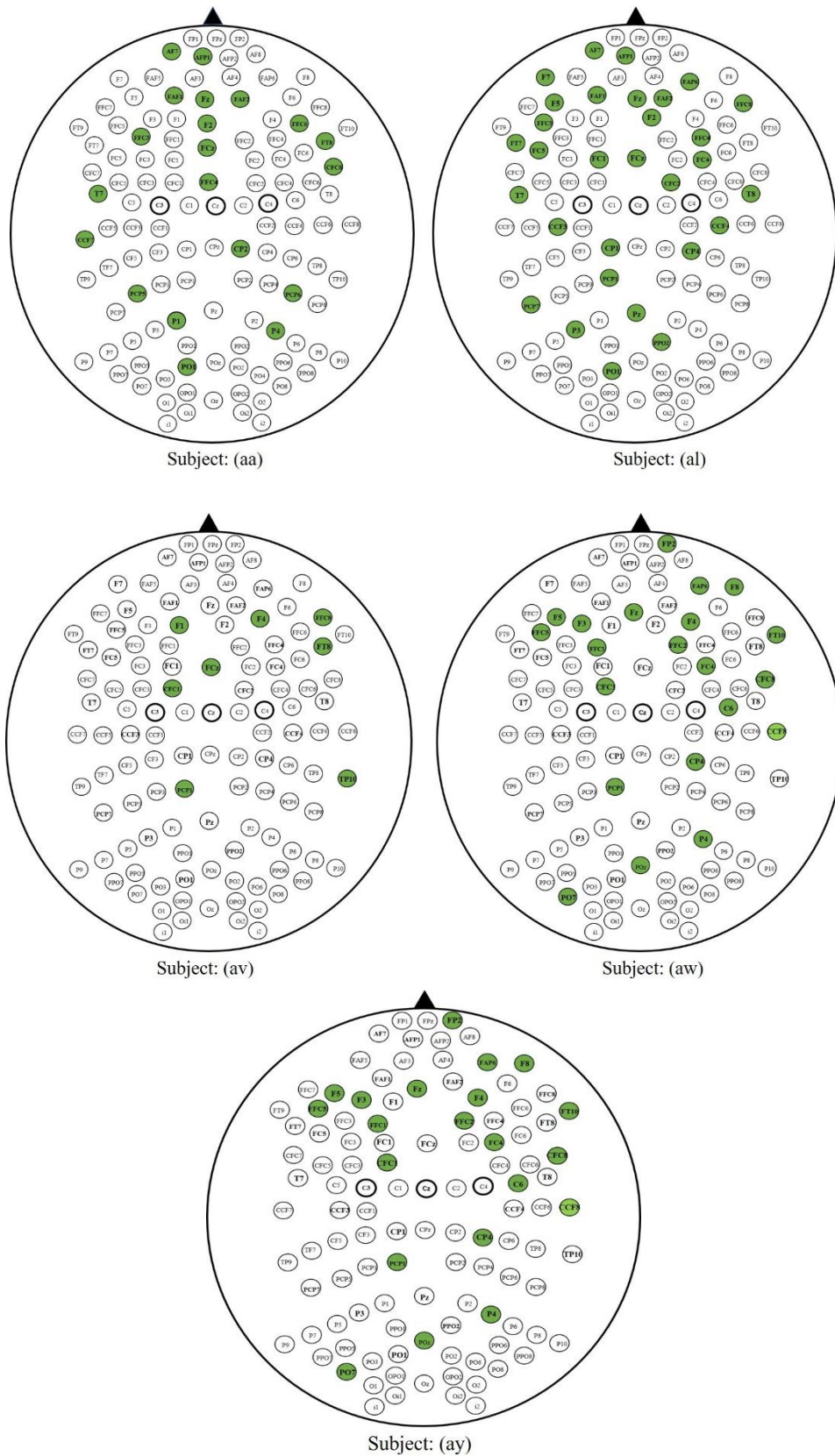


Figure 6.8. Topographical space visualization of selected channels in dataset 3

6.10. Conclusions & Future Scope

This study presents a novel multiobjective X-shaped Binary Butterfly optimization algorithm (MX-BBOA) to select the best set of EEG channels. The MX-BBOA method aims to maintain a good balance between the number of channels and classification accuracy. This approach explores the natural behavior of butterflies with dual sigmoid functions to solve the channel selection problem. Here, a sigmoid function was employed to keep the solution in the binary search space, while its dual function was used to maximize the searchability of butterflies. To achieve a good balance, our algorithm designs a bi-objective function with the number of channels and classification accuracy as input parameters. The performance of the proposed channel selection approach is validated on three EEG-based BCI datasets (BCI competition dataset IV-2a, BCI competition IV datasets 1, and BCI competition III datasets IV a) with a different number of EEG channels. The results indicated that the proposed channel selection approach significantly reduces irrelevant and redundant channels from the original set without compromising classification accuracy. In addition, the MX-BBOA has a better execution speed compared to different state-of-the-art algorithms. The topographical plots of the selected channels reveal that the selected channels are symmetrically distributed in the vicinity of the horizontal plane (Fig. 9) of the human brain where the candidate channels were located. In the future, recently introduced metaheuristic algorithms such as jellyfish search [133], coral reef [134], and water wave optimization [135] methods and their variants can be used to improve the performance of BCI systems. These algorithms have the excellent property of clustering the object in a specific manner that provides a better output than other feasible solutions. For example, in jellyfish search optimization, jellyfish gather together and form a bloom-like structure which is most dangerous and has an effective food-searching ability. Similarly, coral reef and water wave optimization approaches have been used in numerous clustering problems, such as secure routing [136], and combinatorial problems [136].