Chapter 4

Dynamic Channel Relevance: A Filter-based Channel Selection Method

4.1. Introduction

This chapter explains the need for different information-theoretic concepts in the exploration of channel selection problems to minimize the complexity and improve the performance of modern BCI systems. Various experiments were performed to solve the optimal channel selection problem by correlating limb movements with respective cognitive patterns. In these experiments, prior information about neural correlation paved the way to select MI task-related channels from the motor cortex and its neighboring regions. This approach often yields a set of highly significant channels treated as candidate solutions that determine other relevant electrodes. Based on these assumptions, we introduce a filter-based dynamic channel selection method, namely, Dynamic Channel Relevance (DCR) using an improved Mutual-Information (MI) measure. Based on the DCR score, we sorted the channels and provided them with an index (DCRI). A channel with a higher DCR score was awarded with a higher index and similarly, a lower DCR score was given a lower index. In this manner, the channel with a lower DCRI representation is less significant/ relevant as compared to a higher DCRI channel. It has been discussed that the conventional mutual information metric ignores the relevance of new channels with respect to priorly known and candidate solutions in channel selection. Therefore, the proposed study implements a novel three-way interaction maximization strategy to maintain a good balance between relevancy and redundancy levels associated with the selected channel subset.

This method is motivated by the Dynamic Feature Importance (DFI) [75] approach that combines Gini Importance (GI) [76] and Mutual Information Coefficient (MIC) [77] to measure the importance and redundancy of a selected channel. The selected channels are used to extract spatial-temporal features using the Multivariate Empirical Mode Decomposition (MEMD) method [78]. The extracted features are used to discriminate different MI tasks (left hand, right hand, tongue, and foot) using the Support Vector Machine (SVM) [79]. The experiment is validated on three public EEG datasets (Dataset 1, Dataset 2, Dataset 3). The results show that the proposed method achieved a superior classification accuracy (85.4% on Dataset 1, 80.33% on Dataset 2, and 85.20% on Dataset 3) with a lesser number of channels compared to state-of-the-art methods. In addition, our method significantly reduced the computation time compared to other published results without compromising the classification accuracy. Topographical mapping between the selected channels and cognitive regions showed that the central, frontal, and parietal lobe contributes to the execution of various MI tasks during physical activities. The Python code used in this chapter can be accessed from Link 1 given in the footer. The block diagram of the proposed methodology is shown in Figure 4.1.

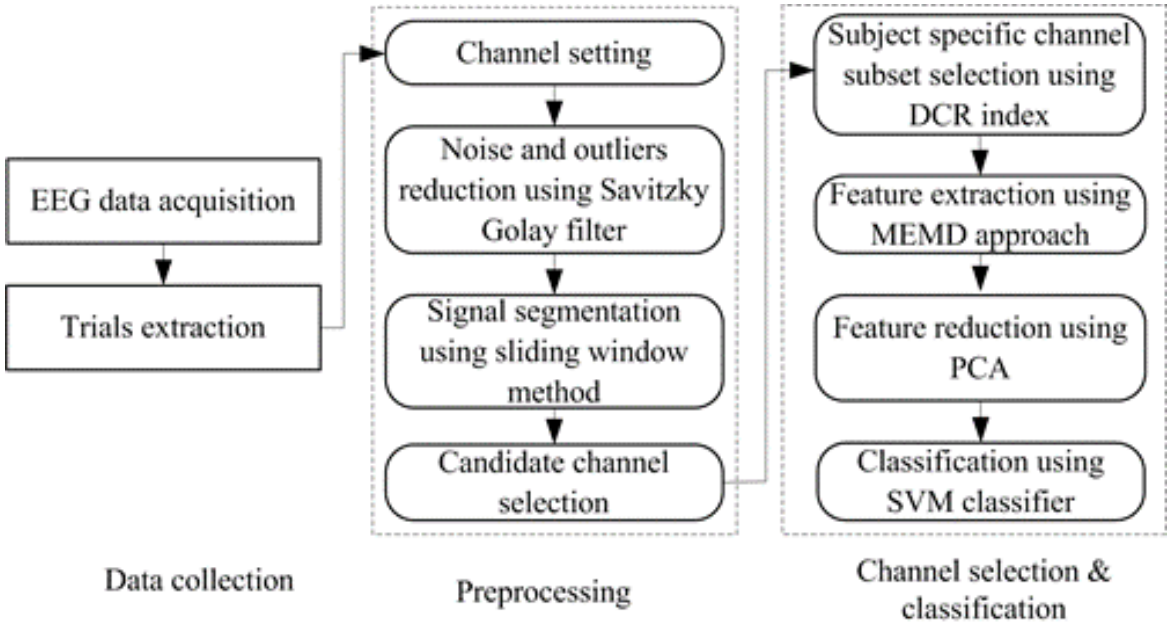

Figure 4.1. Block diagram of proposed EEG channel selection methodology using Dynamic Channel Relevance

[1] https://drive.google.com/drive/folders/122NnIK4MrvpZK5YGBLP3gUhW_Aic6j0-?usp=sharing

4.2. Limitations of the Existing Literature

It is known that filter methods provide popular search strategies to determine optimal solutions from a large search space. Moreover, these methods are fast. scalable, but have a risk of overfitting. In the channel selection problem, these methods suffer from the following limitations:

I. They often ignore the channel dependencies while designing the optimal channel subset which further increases the redundancy level in the final computed solution.

II. Since channel subset selection depends on the neurophysiological basis of the motor-imagery region, the increased redundancy may overlook the significance of candidate solutions.

III. This problem results in poor classification accuracy with high computational complexity when used with a large number of channels.

4.3. Our Contribution

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

I. To study the variance among different cognitive patterns corresponding to four MI tasks.

II. We propose a channel relevance index (DCRI) to rank selected channels based on their importance and redundancy level (DCR score) associated with the candidate channels.

III. To develop a multiclass classification model to discriminate four MI tasks using the subject and the channel-specific MEMD features.

IV. To present the results in terms of Channel Reduction Rate (CRR) and classification accuracy (CA), Specificity, and Sensitivity when EEG chunks are discriminated against the extracted features.

V. To find the correlation between MI tasks and the different cognitive regions recognized by the topographical map of selected channels.

4.4. Redundancy and Relevancy in Channel Selection Problem

Relevance and redundancy are two important measures to determine the significance of an entity in an ongoing mathematical procedure. In the channel selection scheme, relevance indicates the effect of selected or rejected channels on realized classification accuracy. In the case of relevant channel selection, the model's classification accuracy always increases; otherwise, it is considered irrelevant and rejected. Moreover, redundant channels add duplicate information to the system, which increases the computational burden of the proposed method. Therefore, maintaining an optimal trade-off between relevancy and redundancy measures is required during channel selection. The following concepts are used to evaluate relevancy and redundancy measures in the applied channel selection approach.

4.4.1. Information-Theoretic Background

In information theory, the uncertainty or entropy [80,82] associated with a $n-$dimensional random variable $X = \{x_1, x_2, .. x_n\}$ is defined as follows:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{4.1}$$

where $p(x_i)$ shows the probability of x_i in X. If the Eq. 4.1 is extended to two variables X and $Y = \{y_1, y_2, .. y_n\}$, the joint and conditional entropy [81,82] will be defined by:

$$H(X,Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_i) \log p(x_i, y_i) \tag{4.2}$$

$$H(X|Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_i) \log p(x_i\ |y_i) \tag{4.3}$$

where $p(x_i, y_i)$ shows the joint probability of x_i and y_i, $p(x_i\,|y_i)$ indicates the conditional probability of x_i given y_i. Mutual Information (MI) [82] represented by $I(X;Y)$ shows the shared information between variables X and Y. It can be defined as:

$$I(X;Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_i)\{\log x_i - \log y_i\} \quad (4.4)$$

MI and entropy have the following relation [75,82]:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \quad (4.5)$$

$$= H(X) - H(X|Y) \quad (4.6)$$

$$= H(Y) - H(Y|X) \quad (4.7)$$

In the case of a third variable Z, conditional mutual information [82,83] is the mutual information between two other random variables X and Y, given by Eq. 4.8.

$$I(X;Y|Z) = H(X|Z) + H(X|Y,Z) \quad (4.8)$$

4.4.2. Channel Redundancy Measurement

Suppose $C = \{c_1, c_2, .., c_n\}$ is the global channel set, $C = \{C_1, C_2, ... C_k\}$ is the set of candidate channels, and S $(S \subseteq F)$ is the optimal channel subset. Considering the basic feature selection criteria defined by Wei et al. (2020) [75], channel redundancy and relevancy can be defined accordingly definitions (1-4).

Definition 1- Channel Redundancy. Let c_i and c_j $(i \neq j)$ are two channels in set C, the relationship between both is defined by channel redundancy. Specifically, the degree of redundancy between them is defined by Maximal Information Coefficient-MIC (c_i, c_j). The detailed approach to MIC estimation is given in the following steps.

1. Partition scatter plot of c_1 and c_2 over-ordered integer pair of (a, b).
2. To compute the maximum possible mutual information I^* (c_1, c_2, a, b), prepare a grid of size $(a \times b)$ applied to the scatter plot.
3. Normalize the largest mutual information values I^* (c_1, c_2, a, b) such that $M_{a,b}(c_1, c_2) \in [0,1]$ using Eq. 4.9.

$$M_{a,b}(c_1, c_2) = I^*(c_1, c_2, a, b) / \log_2 \min(a, b) \quad (4.9)$$

4. MIC is the maximum of $M_{a,b}(c_1, c_2)$ over the ordered pair of (a, b) and calculate using Eq. 4.10.

$$MIC(c_1, c_2) = \max_{a,b} M_{ab}(c_1, c_2) \quad (4.10)$$

Definition 2- Strong Redundancy. Channel c_i is more redundant to channel c_k than channel c_j, the relation defined in Eq. (4.11) holds.

$$MIC(c_i, c_k) > MIC(c_j, c_k) \quad (4.11)$$

where c_k is a candidate channel, and c_i, and c_j are selected channels.

4.4.3. Channel Relevancy Measurement

Let C_s be any selected channel from the set C, the role of C_s to candidate channel C_c is defined by channel relevance. Here, we used an important information-theoretic paradigm termed ***Gini Index*** (***GI***) to calculate the relevance of C_s defined by $GI(C_s)$. The motivation behind the selection of *GI* over other significant measures such as Pearson correlation [84], ANOVA [85], Chi-square [86], LASSO [87], and Minimum redundancy and maximum relevance (mRMR) [88] is its ability to compute impurity or disorder within a dataset. The Pearson correlation coefficient is another important metric that is easy to interpret but fails to determine the non-linear relationship among features. This constraint applies to LASSO also because it is complex to implement and fails to determine non-linearity between two features. Moreover, the performance of the ANOVA test depends on several assumptions such as (1) Normal distribution of data, and (2) Assumption of independence (no relationship among features). Initially, we have to perform several preprocessing steps to determine the independence level of features and distribution of data which is again a time-consuming and complex procedure. Similarly, the Chi-square test also has two major considerations about data such as (1) Large sample size, and (2) assumption of independence. Therefore, we ignored this method because we did not consider the dependence level of features and sample size while estimating the relevancy of an attribute.

The mRMR is a popular algorithm that is effective in a variety of applications. However, it also has some limitations such as (1) Noise sensitivity, (2) Computationally expensive for large datasets, and (3) Biased towards correlated features [156]. Considering the above assumptions, we used the GI score to compute the relevancy score of individual features. The mathematical definition of the GI measure is discussed below.

Definition 3- The Gini Index (GI) is an important information-theoretic paradigm based on the Classification and Regression Tree (CART) [89]. In CART, the purity of dataset D at a given node can be estimated by using the *Gini Index* defined in Eq. 4.12.

$$Gini\ (p_1, p_2, \dots, p_k) = \sum_{j=1}^{k} p_j(1 - p_j) \tag{4.12}$$

where $p_1, p_2, \dots, p_k$ are the probabilities of class 1, 2..., and k, respectively. The criterion for CART is to maximize the Gini Decrease (GD). When a node is split according to the channel C_s, the GD of the channel C_s is calculated using Eq. 4.13.

$$GD(C_s) = Gini\ (D) - \sum_{v} \frac{|D^v|}{D}\ Gini(D^v) \tag{4.13}$$

where D^v is the dataset at the split child node, and CART is generally a binary node. The second term represents the weighted sum of the Gini index at the split child nodes, and its value is always lesser than the parent node.

Definition 4- High Channel Relevancy. Channel c_i is more irrelevant to the channel c_k than channel c_j, the relation defined in Eq. (4.14) holds.

$$GI\ (c_i) > GI\ (c_j) \tag{4.14}$$

where c_k is a candidate channel, and c_i, and c_j are selected channels.

4.5. Proposed Methodology: Dynamic Channel Relevance

As earlier discussed, two important factors: (1) channel relevance and (2) channel redundancy, are concentrated over as the objective functions for designing an optimal channel subset. This study developed the Dynamic Channel Relevance Index (DCRI) for objective function optimization. It dynamically selects channels according to their high importance and low redundancy score. Mathematically, the dynamic channel relevance of any channel C_i in the C is estimated as given in Eq. 4.15.

$$DCR\,(C_i) = GI(C_s) \times \prod_{C_x \in C} [1 - MIC\,(C_i, C_j)] \tag{4.15}$$

Initially, when the channel subset S is empty, the *DCR* of C_s is equal to GI. The pseudocode of the DCR algorithm is given in Algorithm 4.1.

Algorithm 4.1. Pseudocode of the DCR algorithm

Input: A training Sample D with channel set C= $\{C_1, C_2, \ldots, C_k\}$ and class c and two predefined thresholds *k* and *acc*.

Output: The selected channel subset *S*

S= ∅

K=0

a = 0

For *i= 1 to k* ***do***

Calculate the channel importance $GI(C_i)$

For *j= i+1 to k* ***do***

Calculate $MIC\ (C_i, C_j)$

End *For*

End *For*

While *k < K* **or** *a < acc* ***do***

Select a channel C_h with the largest $DCR\ (C_h)$

S= S ∪ $\{C_h\}$

F = F-$\{C_h\}$

Calculate classification accuracy obtained by features extracted from *S* and record it as *a*

K=k+1

End *While*

Update *S* with a decreasing order of DCR score and provide an index *i* [1,2,…*K*] to each channel based on their DCR score.

Return channel name, DCR score, DCR index (DCRI)

4.6. Feature Extraction Using Multivariate Empirical Mode Decomposition

The decomposed EEG segments of N observations are considered to extract segment-wise MEMD features. In Figure 4.2, the top nine MIMFs computed from the five channels of subject 1 of BCI competition IV-2008-IIA have been shown. If the size of the covariance matrix is given by (m, n) and the total number of channels is N, then the number of extracted features will be $N \times m \times n$. In dataset 1, we obtained 9724 (22*221*2) MEMD features from 22 channels. The same experiment was also performed for the remaining two datasets, and MEMD features were collected for further processing.

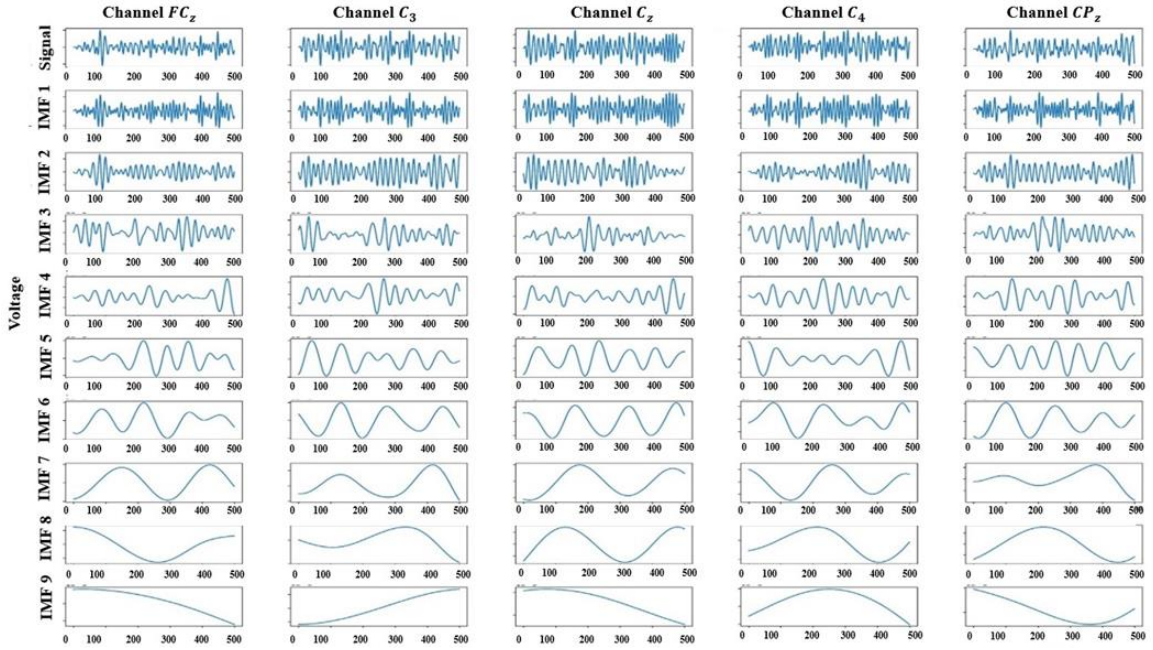

Figure 4.2. The EEG signals and their MIMFs corresponding to the channels FCz, C3, Cz, C4, and CPz for left hand MI task of A01

4.7. Feature Selection

MEMD results in a large feature space because of the numerous decompositions of the EEG signals extracted from the selected channels. To achieve better classification results, it is required to process only those significant features and effectively discriminate MI tasks. A popular feature reduction method called Principal Component Analysis (PCA) [90] is used to reduce the size of the large feature set using their covariance score. Previously, PCA has been successfully implemented in various research studies based on MI classification and performed well compared to baseline methods [91, 92]. In PCA, the feature matrix F is used to compute the orthogonal matrix $W_{K\times M}$ which is further used to produce the transformed matrix $Y_{K\times N}$ using Eq. 4.16:

$$Y_{K\times N} = W_{K\times M} \times F_{M\times N} \tag{4.16}$$

where M is the size of the feature set, N is the number of instances, and $K \leq M$ represents the dimension of the output feature set where K is selected based on the accumulation of the first few largest eigenvalues exceeding 98.6% of the total sum of all the eigenvalues computed from the feature covariance matrix. The first few smallest K values that satisfy Eq. 4.17 are selected as the model for enumerating the number of the principal components.

$$\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{C} \lambda_i} \times 100 \geq 98.6 \tag{4.17}$$

Finally, we selected the top P features from the original set to improve the average (avg.) classification accuracy compared to the absence of feature reduction. In Table 4.1, we compared the avg. Classification Accuracy (CA) computed on Left vs. Right-hand test data of dataset 1 in two cases: (1) without, (2) with PCA implementation. This comparative analysis indicates that the implantation of PCA on extracted MEMD features is vital in improving the average. classification accuracy.

Table 4.1. Comparison between with and without PCA implementation on average Classification accuracy computed from a test sample of dataset 1

Subject	Without PCA implementation classification accuracy (in %)	With PCA implementation classification accuracy (in %)
1	59.18	71.33
2	61.42	67.24
3	49.02	63.12
4	68.81	79.82
5	72.20	75.50
6	51.60	68.66
7	69.39	76.80
8	58.26	71.39
9	75.21	80.02
Average	62.78	72.65

4.8. Experimental Results

With the aforementioned three BCI competition datasets, extensive experiments are performed to compare the performance of the proposed method with various baseline methods. The performance of our method on each dataset is presented next.

4.8.1. Test Results on Dataset 1

BCI competition IV- 2008-IIA is a four-class MI dataset; therefore, a multiclass classification approach is required to obtain true class labels from the input feature set. However, the SVM model was meant to perform only binary classification because of the way it creates a hyperplane between two classes. Therefore, we extend it to resolve a multiclass classification problem by applying the One vs. All strategy [93] to the selected MEMD features. No cross-validation scheme is applied during the performance evaluation because training and testing samples are explicitly provided in this dataset. To investigate the proposed channel selection method's performance, we analyze the relationship between the optimum number of channels

and the corresponding classification accuracy. In the learning phase, six MI-task pairs: left vs. right (L vs. R), left vs. foot (L vs. F), left vs. tongue (L vs. T), tongue vs. feet (T vs. F), tongue vs. right (T vs. R), and right vs. feet (R vs. F) were created for each subject, and an SVM classifier is applied over each possible pair of MI tasks. Further, the trained classifier is used to classify all unknown test trials of the corresponding train datasets.

We execute our algorithm on all nine subject-specific evaluation samples and calculate the mean classification accuracy for all ${}^{4}C_{2}$ MI task pairs. Figure 4.3 shows the feature distribution or covariance of the top five left-hand and right-hand MI tasks based on MEMD features. It depicts the boxplot for the MEMD features for the C3 candidate channel estimated by the Kruskal-Wallis test [94]. These 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 [95]. Then, all the features are arranged based on their rank in the descending order of class separability. A similar approach is also applied for the tongue and foot classes with relative *p*-values of 0.0421, 0.0032, 0.0330, 0.0582, and 0.0326. The set of selected normalized features with significant p-values is further used to classify all the MI tasks using the SVM classifier.

Table 4.2 provides the subject-wise performance of the proposed channel selection scheme in terms of average classification accuracy and channel reduction rate. It can be noticed that maximum classification accuracy is achieved when the proposed channel selection scheme discards approximately 50% of the irrelevant channels. It shows that the proposed channel selection scheme effectively balances the trade-off between the redundancy and relevance level associated with the raw EEG channels. It can be noticed that the average classification performance achieved by the proposed approach is maximum for the R vs. F MI task pair. In Figure 4.4, the effectiveness of the optimal channel subset is computed in terms of the average GI score and size of the optimal channel subset for individual subjects. Initially, the Gini information of the channel subset depends on candidate channels which varies when a new one is added to the subset.

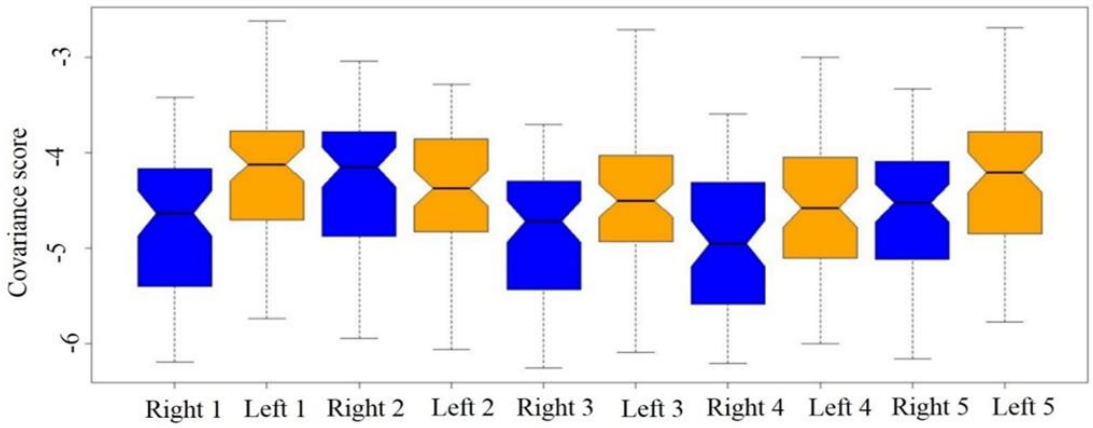

Figure 4.3. Kruskal -Wallis test on top five features obtained after applied MEMD approach on Dataset 1(left hand shown in orange color and right hand shown in blue color). The similarity between the second feature of the left and the right hand can be interpreted as a resemblance between the corresponding cognitive control commands. The remaining pairwise (1-1, 3-3, 4-4, 5-5) left-hand and right-hand movement features show a high variance gain.

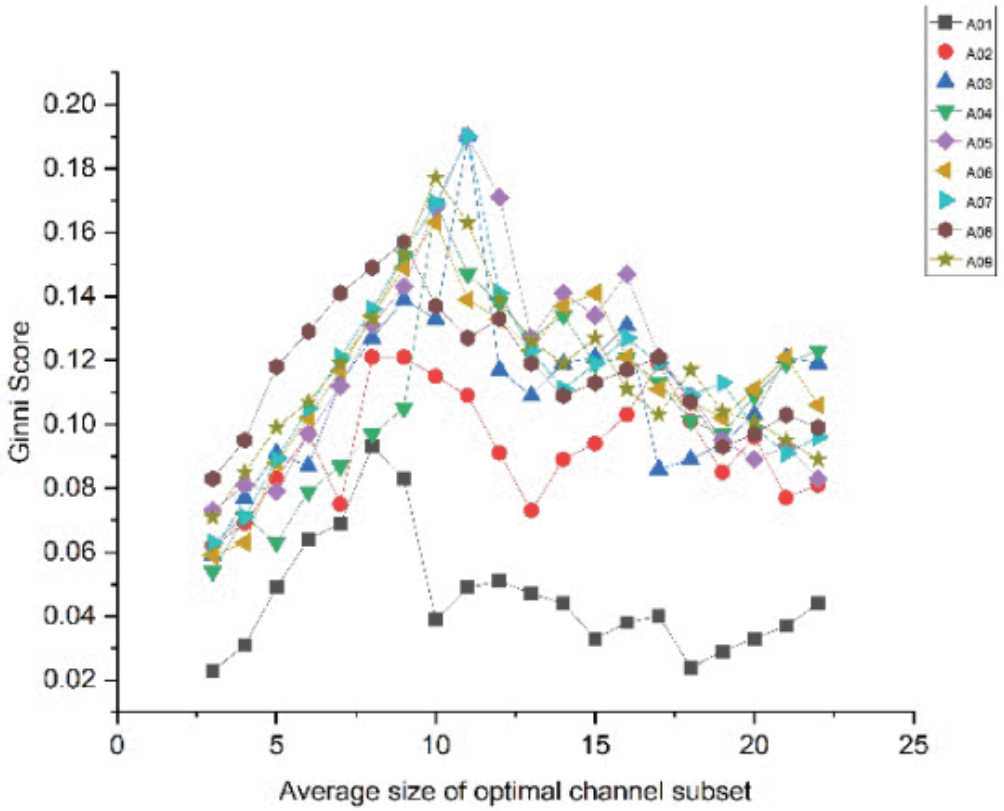

Figure 4.4. A relationship visualization between optimal channel subset and Gini score

It can be observed that the dynamic relevance of a newly selected channel (Ci) mainly depends on the maximum information that it shares with priory selected channels (Cj). Moreover, if there is a strong redundancy between the Cj and Ci , the DCR (Ci) will decrease significantly; otherwise, it will increase. Deleting a redundant channel from the global channel set (Gcs) will reduce the sharable information among Cj and deleted channels and therefore no change will occur in the GI score of Gcs . On the other hand, the selection of redundant channels will increase the sharable information among all the channels and provide a non-zero value of $MIC(Ci, Cj)$, reducing the GI value. Similarly, the selection of non-redundant channels will reduce the sharable information among the channels, maximizing the global channel's GI score. The maximum Gini information of 0.19 is achieved during the experiment for A03, A05, and A07, with a channel subset size of 11 channels. Since EEG signals are subject-specific, computing a global DCR threshold for the total population is impractical.

4.8.1.1. Results Comparison with Baseline Methods

In this subsection, BCI classification results on dataset 1 are compared with two types of published results: (1) publications without and (2) publications with channel optimization. All 22 channels are used in classification in the former, while in the latter, a specific search criterion filters a relevant set. A detailed description of both classes of comparison is given below.

I. Publications without Channel Optimization

The reported results in Table 4.2 confirm that the proposed channel selection scheme helps to achieve a maximum classification accuracy of 97.12% (R vs. T) for the eighth participant with an average classification accuracy of 87.72%. Similarly, the other five cases also achieved superior classification accuracy compared to earlier published results where no channel optimization was used. Table 4.3 shows the comparison of the proposed channel selection scheme with other state-of-the-art methods described in Gaur et al. (2018) [96]. Method 1 demonstrates the classification results obtained by first executing time window-based bandpass filtering in the range of 8-30 Hz, and then extracted spatiotemporal features were classified using the Riemannian classifier. The proposed method improved classification results by 9.19%, and five out of nine participants showed improved classification accuracy.

Method 2 shows the classification accuracy in the training session using the Subject-Specific Multivariate Empirical Mode Decomposition (SS-MEMD) with CSP features. In this case, we obtained a 9.71% improvement in the classification accuracy across all nine subjects. Method 3 used the CSP features on bandpass-filtered EEG in the range of 8 to 30 Hz. Then, the log-variance of extracted features was computed, and transformed features were classified by Linear Discriminant Analysis (LDA) [97]. Methods 4 and 5 used CSP features and identified the covariate shift followed by adaptive and transductive learning, respectively. In method 6, a feature covariance matrix is constructed using the SS-MEMD method, and the Riemannian classifier was used to discriminate four MI tasks. The proposed method improved classification by 7.39% compared to method 5.

II. Publications with Channel Optimization

In Table 4.4, the obtained classification results are compared with two published results: (1) Sparse Common Spatial Pattern (SCSP) based channel selections method and (2) Channel set optimization using Improved Binary Gravitation Search Algorithm (IBGSA). In the SCSP approach [98], two sparse filters were used to maximize weight sparsity among the channels selected by the CSP projection matrix. Further, two selection criteria: (1) minimum noise and channel irrelevance with utmost classification accuracy, and (2) selected channel minimization while maximizing classification accuracy were used for the optimization.

Finally, the SCSP algorithm with the first and the second selection criteria are, respectively, abbreviated as SCSP1 and SCSP2. The authors reported that the proposed channel selection approach achieved average classification accuracy of 81.63% with SCSP1 and 79.07% with SCSP2.

In recent work, the Improved Binary Gravitation Search Algorithm (IBGSA) [21] automatically detected more informative channels in left or right-hand classification. Here, time and wavelet features were extracted to discriminate four MI tasks using an SVM classifier. The results confirm that the IBGSA-based channel selection method achieved a maximum of 80% and an average of 76.24% classification accuracy. The results reported in Table 4.4 show that the proposed channel selection scheme is more effective than both state-of-the-art methods. It improved classification accuracy by 5.1% and 8.5% when compared with SCSP1 and SCSP2

approaches, respectively. Compared with the IBGSA-based channel selection approach, the classification results improved by 12.6%, and six out of nine participants showed enhanced classification results without attenuating the average. classification accuracy. We also compared the effect of channel reduction rate on subject-wise classification accuracy, as shown in Figure 4.5. To do so, a new performance variable, Channel Importance Factor (*CIF*) is calculated using Equation 4.18.

$$CIF_i = \frac{Classification\ accuracy\ of\ i^{th}\ participant}{total\ number\ of\ selected\ channels} \tag{4.18}$$

Here, the CIF_i variable measures the role of each selected channel to obtain classification accuracy. For example, our method achieved 89.79% classification accuracy using eight channels for subject 1; therefore, the Channel Importance Factor for subject 1 (CIF_i) will be 11.22 (89.79/8). It can be observed that a channel selection approach will exhibit higher CRR if it has a maximum CIF_i value for the i^{th} subject.

Figure 4.5 shows that the proposed DCR methodology consists of a maximum *CIF* for subjects 1,4,6, and 7 compared to the other three algorithms. Similarly, the IBGSA and SCSP2 achieved the highest *CIF* for subjects 3, 5, and 2, 8, respectively. The SCSP1 obtained the best *CIF* for subject nine compared to other approaches. We can conclude that the proposed channel selection approach effectively removes redundant and irrelevant information from original EEG channels without attenuating classification accuracy. To evaluate the correctness of classification results, two performance metrics: (1) Sensitivity (SE), and (2) Specificity (SP), were computed using Equations 4.19 and 4.20.

$$Sensitivity = \frac{TP}{TP+FN} \tag{4.19}$$

$$Specificity = \frac{TN}{TP+FN} \tag{4.20}$$

where TP, TN, FP, and FN are True positive, True negative, False positive, and False-negative instances, respectively.

Table 4.2. Performance of the proposed channel selection approach, when applied in the classification of six MI tasks, pair on test data

Subject	Classification accuracy of all MI pairs						Selected channels using DCR measure. Three candidate channels + (Selected channels)	Number of channels	CRR score
	Lvs. R	Lvs. T	L vs. F	R vs. T	R vs. F	F vs. T			
A01	89.79	87.48	84.79	87.84	91.31	71.08	C_3 ,C_4, C_z, 7, 11, 14, 19, 21	8	0.63
A02	94.18	91.40	94.12	77.88	79.68	87.16	C_3 ,C_4, C_z, 1, 5, 11, 13, 16, 20	9	0.59
A03	78.92	93.30	89.44	91.52	88.44	78.10	C_3 ,C_4, C_z, 2, 4, 7, 11, 15, 16, 19, 22	11	0.50
A04	94.01	79.17	72.68	86.28	87.42	71.44	C_3 ,C_4, C_z, 4, 5, 9, 11, 16, 17, 19	10	0.55
A05	71.32	74.31	83.29	94.42	96.28	93.64	C_3 ,C_4, C_z, 2, 4, 7, 9, 11, 15, 16, 18, 20	12	0.45
A06	86.71	88.51	72.49	87.29	78.33	81.91	C_3 ,C_4, C_z, 1, 4, 6, 9, 14, 16, 18, 19	11	0.50
A07	89.36	73.55	93.18	84.72	93.71	89.24	C_3 ,C_4, C_z, 2, 4, 5, 9, 13, 15, 17, 20	11	0.50
A08	82.11	96.45	85.19	97.12	87.58	82.28	C_3 ,C_4, C_z, 3, 5, 9, 11, 14, 17, 18, 22	11	0.50
A09	86.18	77.30	87.40	81.09	91.76	77.59	C_3 ,C_4, C_z, 2, 6, 9, 11, 15, 16, 17	10	0.55
Avg.	85.84	84.60	84.73	87.72	88.27	81.38	Highly voted channels: 11, 9, 16, 4, 2, 5 15, 17, 19	11	0.53

Table 4.3. Comparison of classification accuracy (in %) for L Vs. R task with baseline results (without any optimization). All results of baseline methods are taken from Gaur et al., (2018) [96]

Participants	Proposed method (L vs. R)	Method-1	Method-2	Method-3	Method-4	Method-5	Method-6
1	89.79	91.49	92.91	88.89	90.28	90.28	91.49
2	94.18	61.27	59.86	51.39	54.17	57.64	60.56
3	78.92	94.89	95.62	96.53	93.75	95.14	94.16
4	94.01	75.86	66.38	70.14	64.58	65.97	76.72
5	71.32	60.00	62.96	54.86	57.64	61.11	58.52
6	86.71	70.37	68.52	71.53	65.28	65.28	68.52
7	89.36	65.00	68.57	81.25	62.5	61.11	78.57
8	82.11	96.27	96.27	93.75	90.27	91.67	97.01
9	86.18	92.31	93.08	93.75	85.42	86.11	93.85
Avg.	85.84	78.61	78.24	78.01	73.84	74.92	79.93
p-value	0.2478	-	0.7831	0.8147	0.0025	0.0116	0.4241

Table 4.4. Comparison of classification accuracy (in %) for L Vs. R task with baseline results (with optimization)

Subjects	Proposed method		Arvaneh et al. 2011 (SCSP1) [98]		Arvaneh et al. 2011 (SCSP2) [98]		Ghaemi et al. 2017 [21]	
	Selected channels	L vs. R	Selected channels	SVM	Selected channels	SVM	Selected channels	SVM
1	8	89.79	13	91.66	13	91.66	9	76.66
2	9	**94.18**	9	67.36	4	60.41	6	76.66
3	11	78.92	14	97.91	12	97.14	7	73.33
4	9	94.01	14	72.22	11	70.83	8	73.33
5	11	71.32	11	65.27	9	63.19	11	80
6	10	86.71	14	66.67	10	61.11	12	73.33
7	11	89.36	19	84.72	15	78.47	15	76.66
8	9	82.11	15	97.22	5	95.13	11	80
9	10	86.18	10	91.66	5	93.75	-	-
Avg.	10	85.84	13.22	81.63	8.55	79.07	8.6	76.24
p-value		0.4598		-		0.0368		

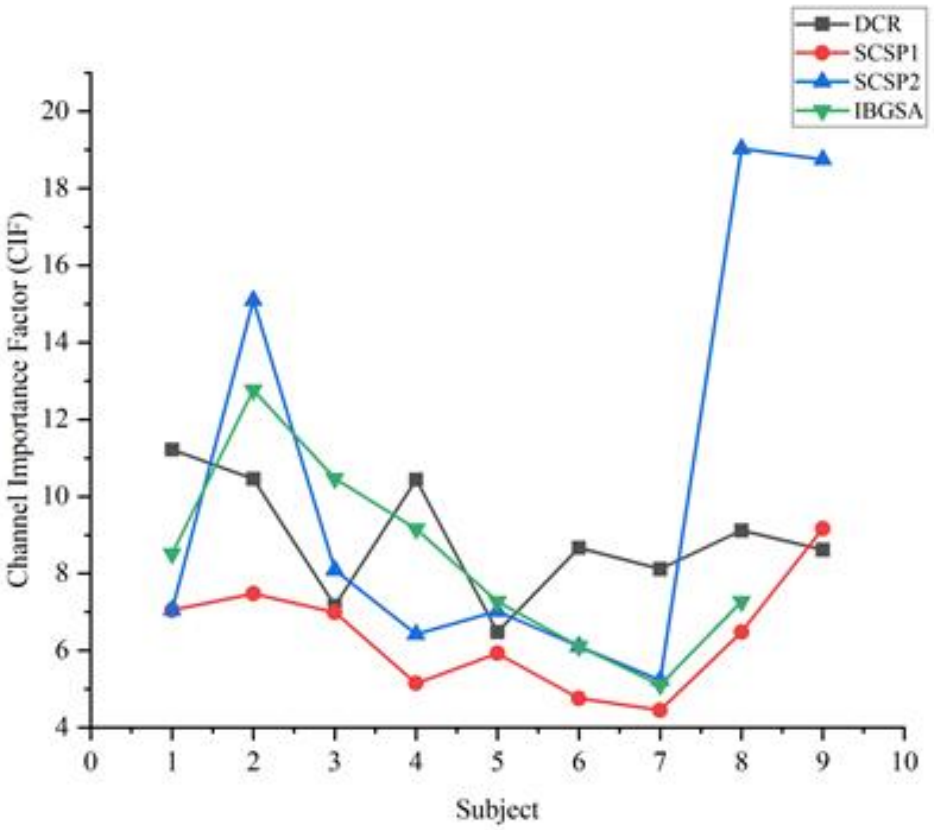

Fig. 4.5. Subject-wise comparative analysis of Channel Importance Factor on dataset 1

Based on the selected optimal channels, global test results are shown in Figure 4.6 as a radar graph [99]. For the given dataset, the average classification accuracy, SE, and SP over all nine subjects was 85.4%, 82.76%, and 81.34%, respectively. In Figure 4.7, we have shown the relationship between the position of selected channels and different cognitive regions using subject-wise topographical plots. It explains that recognition of demonstrated MI tasks involves a different combination of channels. Based on selection frequency (>4), a total of nine channels were considered highly voted channels. It is clear that the selected channels are mostly distributed in the human brain's central and parietal regions. Since the parietal region is just below the central lobe (location of candidate channels), it can be concluded that there may be a functional similarity between both cortical regions in terms of motor control and stimuli responses.

4.8.2. Test Results on Datasets 2 and 3

As discussed in Chapter 2, Datasets 2 and 3 consist of more channels compared to Datasets 1. Therefore, we compare our results with only those methods where any optimization process was applied to determine only significant channels in terms of improving classification accuracy. The classification results are compared in the following subsection:

4.8.2.1. Results Comparison with Baseline Methods

We will discuss both jointly in this subsection since the BCI community has provided training and testing data for dataset 1 contrary to both datasets 2 and 3. We compared the performance of our method on datasets 2 and 3 with three baseline algorithms: (1) Support Vector Machine Recursive Feature Elimination [100], (2) Sequential Floating Forward Selection [101], and (3) Improved Sequential Floating Forward Selection [101] in Tables 4.5 and 4.6. The last row of the tables indicates the p-values computed Wilcoxon signed-rank test of the results of "all channels" and the channels selected by baseline methods.

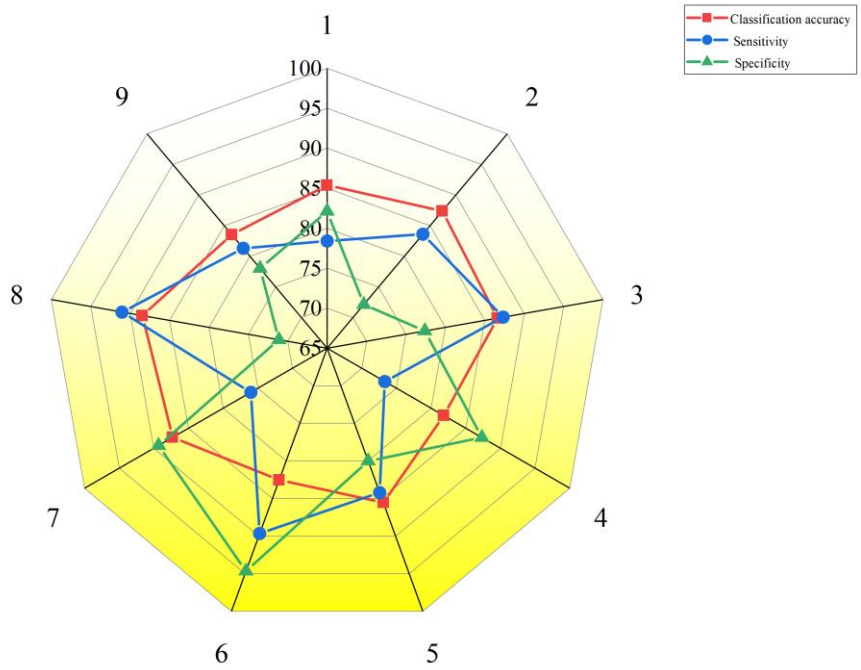

Figure 4.6. Classification accuracy, sensitivity, and specificity of all the channels for the BCI dataset 1

Support Vector Machine Recursive Feature Elimination (SVM-RFE) is a popular feature selection method that selects the feature with the highest-ranking criterion based on respective weights. In the current work, SVM-RFE was employed as a channel selection method, and its results are compared with the proposed method. For each subject, the top r ranked channels are selected and used for feature extraction, and finally, the classification accuracy is computed. SFFS selects a feature from the original set that results in the highest objective function value and dynamically eliminates the most useless feature from the previously selected features. Improved Sequential Floating Forward Selection (Improved -SFFS) is a relatively new channel selection algorithm proposed by Qiu et al. (2016) [101]**.** They improved the performance of the conventional SFFS algorithm by introducing spatial distribution as an important criterion for selecting new channels.

In our work, a 5-fold cross-validation scheme [64] is applied to both datasets to 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, another 20% of feature vectors are used for testing, and the rest of the 80% are employed for the training set. This process is repeated until all the feature vectors are used for testing the proposed algorithm. The best classification accuracies from all the participants of datasets 2 and 3 are shown in Tables 4.5 and Table 4.6. For dataset 2, the proposed DCR algorithm achieved an average classification accuracy of 85.59%, significantly higher than the rest of the four methods.

Compared to the SVM-RFE, our algorithm yielded an 18% improvement in average classification accuracy but used two more channels. In addition, five out of seven participants obtained the best classification accuracy among all the investigated algorithms. In contrast, the SVM-RFE has shown the best channel reduction rate (CRR) by eliminating more redundant and irrelevant channels. For dataset 3, the performance of our method is similar to dataset 2. This dataset represents a dense distribution of electrodes. The avg. classification accuracy of the DCR method was 87.57% and yielded an average improvement of 5.12%, 4.00%, and 11.41% compared to the Improved SFFS, SFFS, and SVM-RFE methods, respectively, using only 25 channels. The interesting point is that, similar to dataset 2; the SVM-RFE method achieved the best CRR among all the investigated channel selection algorithms.

We compared the *CIF* scores of all the channel selection methods to investigate the effect of selected channels on respective averages. classification accuracies. In Figure 4.8, the subject-wise CIF scores are compared on datasets 2 and 3. Since the "all channels" method did not use any optimization method, it is not included in the comparison. It is clear that the SVM-RFE outperformed other algorithms in the selection of the most informative channels. However, the selected channels were not sufficient to achieve the best classification accuracy. Although, our proposed DCR method selected more channels compared to the SVM-RFE method but achieved the best classification accuracy among all the methods. Therefore, it can be concluded that the DCR method effectively selected informative channels without compromising classification accuracy on both datasets with a moderate and large number of electrodes. Considering an optimal balance between the selected channels and the classification accuracy, our method can be the best alternative to developing new BCI systems.

The execution complexity of the proposed approach depends on four steps: (1) Signal preprocessing, (2) Channel selection, (3) Feature engineering, and (4) Classification. In the preprocessing, the SG filter and sliding window-segmentation techniques were applied to improve the quality of raw signals and decompose them into small segments. It has been already discussed in Chapter 3 that the complexity of the SG filter can be asymptotically defined $O(n^2)$ where n is the size of the filtering window. Similarly, the asymptotic complexity of sliding window segmentation techniques depends on the segmentation window which is linear $\{O(n)\}$in nature. In the channel selection process, two measures: (1) the GINI index, and (2) the MIC score. Calculating the GINI Index for a dataset typically requires iterating over all possible splits, which results in a time complexity of O(n^2), where n is the number of data points. The Maximal Information Coefficient (MIC) measures the strength of the relationship between two variables and requires n^2 operations. Therefore the complexity of this step will be O(n^2). The feature engineering step consists of feature extraction (MEMD) having the complexity of $O(n^{11})$ and feature selection (PCA) with the complexity of $O(n^3)$ because of eigenvalue computations and multiple matrix multiplication operations. Hence, the overall complexity of this step will be $O(n^{11})$. Finally, the classification step with the SVM technique will execute in $O(n^2)$ (LibSVM).

Table 4.5. Performance comparison of different channel selection algorithms applied on dataset 2 with 59 channels

Subject	All channels Accuracy (%)	Proposed method		Improved SFFS		SFFS		SVM-RFE	
		Accuracy (%)	channels	Accuracy (%)	channels	Accuracy (%)	channels	Accuracy (%)	channels
A	43	75.02	11	69	6	60	9	57	4
B	42	82.80	8	63	15	66	19	54	2
C	62	90.93	21	87	26	91	21	84	32
D	81	97.06	23	94	29	94	34	88	26
E	91	82.8	20	96	19	96	21	95	10
F	49	78.75	7	65	8	58	19	58	4
G	66	91.77	19	72	22	83	21	71	18
Mean	62	**85.59**	15.28	78.0	17.9	78.3	20.6	72.5	13.7
STd	18.9			14.0	8.7	16.5	7.3	16.6	11.8
p-value				0.0025	--	0.002	--	0.0045	--

Table 4.6. Performance comparison of different channel selection algorithms applied on dataset 3 with 118 channels

Subject	All channels Accuracy (%)	Proposed method		Improved SFFS		SFFS		SVM-RFE	
		Accuracy (%)	channels	Accuracy (%)	channels	Accuracy (%)	channels	Accuracy (%)	channels
aa	75.7	93.6	24	76.4	27	**78.3**	26	68.6	**7**
al	85.7	79.2	33	**94.3**	47	93.6	43	92.9	**8**
av	62.1	94.6	11	65.0	18	68.6	18	62.9	14
aw	87.1	85.54	26	89.5	27	87.9	**20**	80.0	35
ay	87.1	84.94	31	91.4	35	**92.7**	35	88.6	**18**
Mean	79.5	**87.57**	25	83.3	30.8	84.2	28.4	78.6	16.4
STd	10.9			12.3	10.9	10.6	10.5	12.8	11.3
p-value				0.047		**0.023**		0.75	

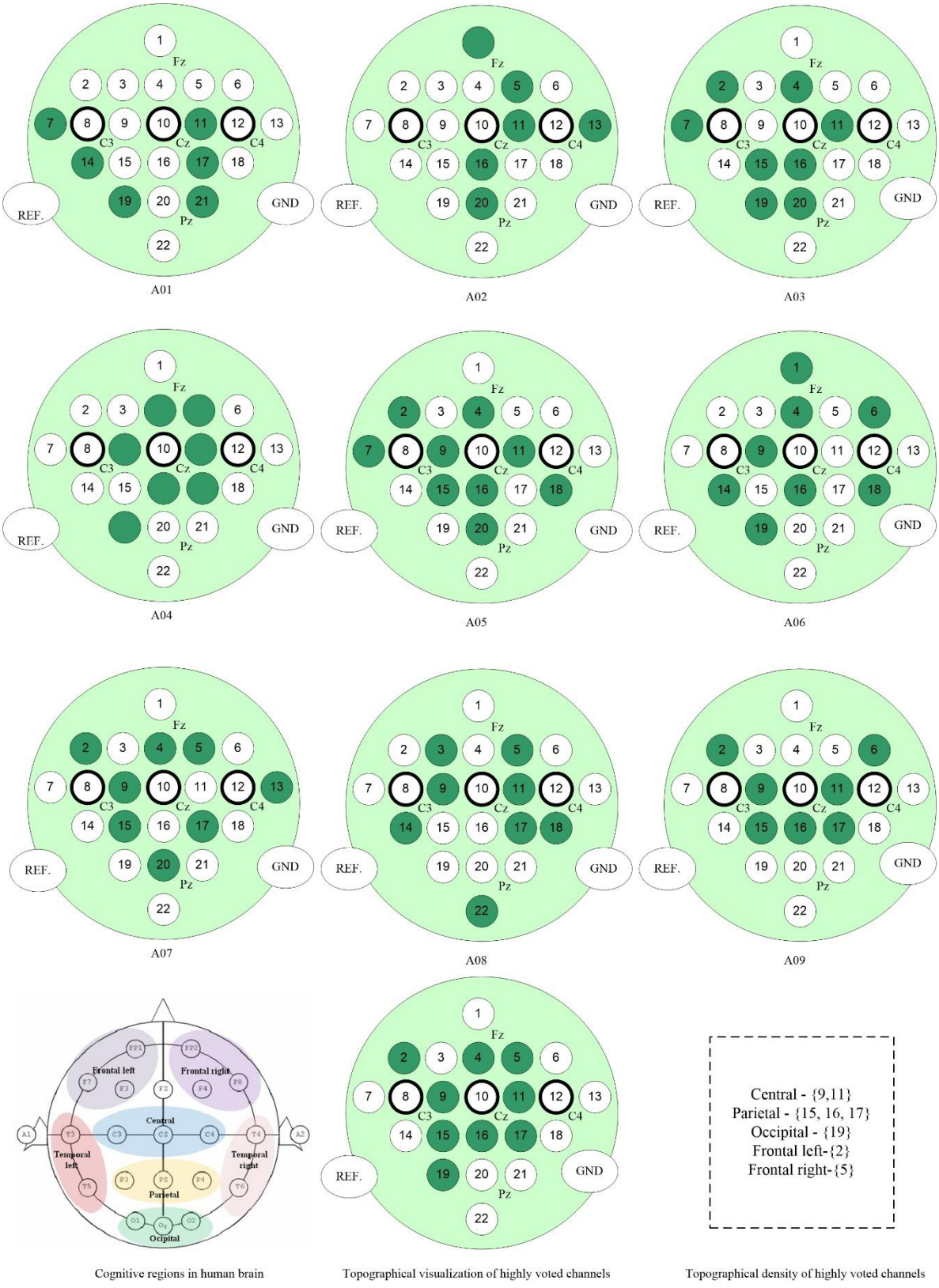

Figure 4.7. Subject-wise topographical visualization of all the selected channels and regional density of highly voted channels in the human brain for dataset 1

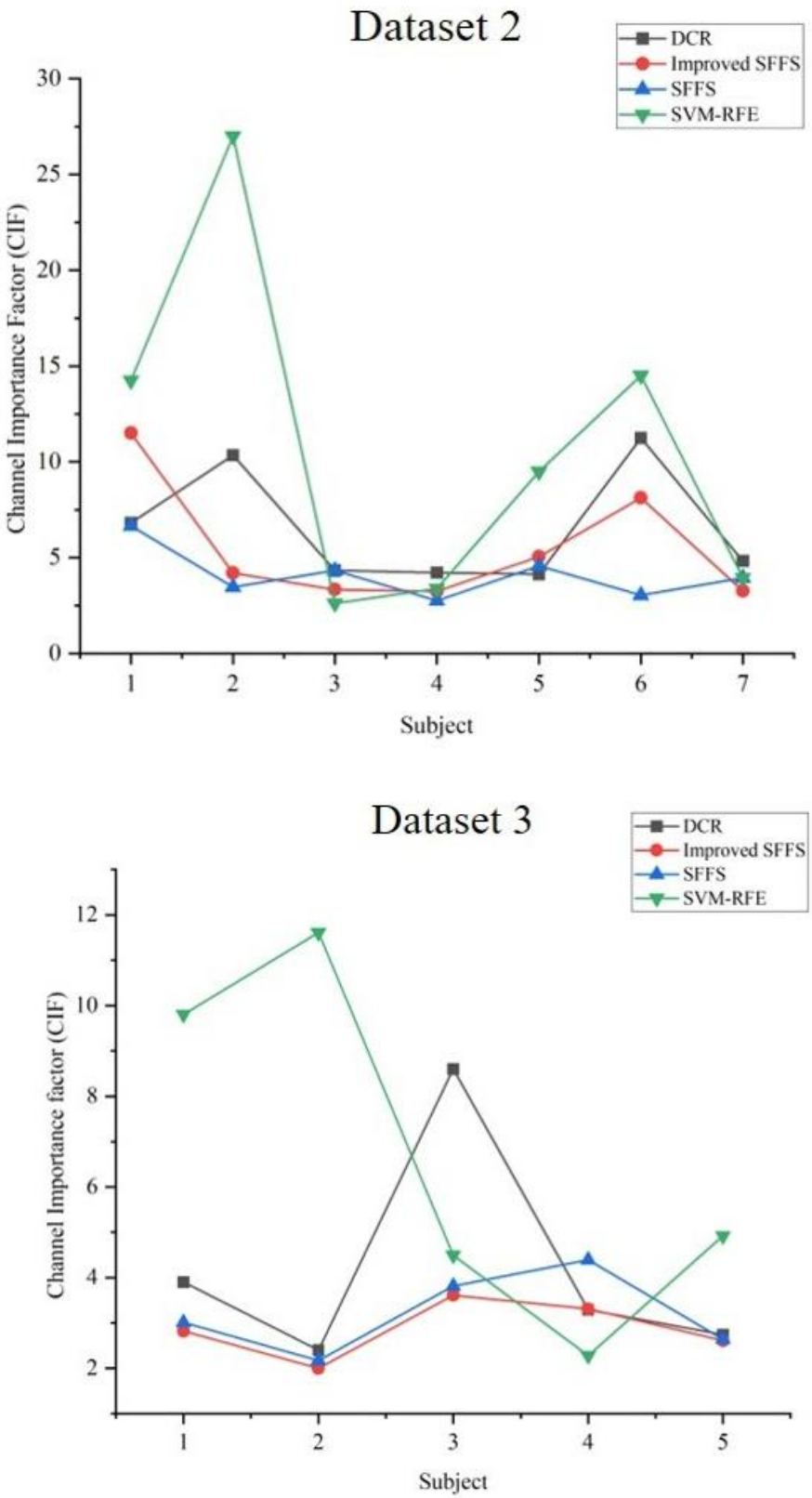

Figure 4.8. Subject-wise comparative analysis of Channel Importance Factor (CIF) scores on datasets 2 and 3.

4.8.3. Discussion

4.8.3.1. Distribution of the Selected Channels

The distribution of the channels selected by our method in dataset 2 is shown in Figure 4.9. Here, the selected channels are marked with different colors according to their location on the human skull. For three out of seven participants (A, D, and F), the selected channels (shown in the red colors) are equally distributed in the frontal and parietal lobes of the cerebrum while the rest of the participants' (B, C, E, G) selected channels (Yellow color marked) were mainly distributed in the frontal lobe of the human brain. Therefore, it can be concluded that the frontal and parietal cortex significantly contribute to MI activity performance along with the motor cortex (central cognitive region).

4.8.3.2. Time Complexity Analysis

The total computation time of the proposed study mainly depends on three steps: (1) channel selection, (2) feature engineering, and (3) classification. Here, we showed the execution speed of all the algorithms in Figure 4.10 on the aforementioned datasets. In Figure 4.10.A, the operating speed of the proposed DCR method, SCSP1, SCSP 2, and IBGSA on dataset 1 is compared in terms of total execution time in seconds. It can be observed that the computation time of our method is significantly lesser than the rest of the channel reduction methods except for subjects 4 and 6. Using the DCR method, the avg. computation time was reduced by approximately 72% (30 seconds) compared to SCSP2. For the rest of the two algorithms, the DCR approach is 57% and 38% faster than SCSP1 and IBGSA methods, respectively. The time complexity of the four-channel selection methods (All channels, SVM-RFE, SFFS, Improved SFFS) on datasets 2 and 3 is compared with the proposed approach in Figure 4.10.B and Figure 4.10.C. It can be observed that the “all channels” method is the slowest among the remaining channel selection approaches for four participants because redundant and irrelevant data processing is also involved in the MI-based BCI classification. Here, the execution time of all five methods is relatively high compared to Dataset 1 because of more number of channels. Using the DCR method, the avg. computation time was reduced by approximately 89% (6890 seconds) compared to the “all channels” method. The execution speed of our method is exceptionally fast compared to the remaining three algorithms for four subjects out of five participants. Compared to the next algorithm, the

SVM-RFE approach, our method reduced the computation time by 59.63% (1131 seconds). Similarly, for the rest of the two algorithms, the SFFS and Improved SFFS schemes, our method significantly minimized the average. computation time by 43.08% (830 seconds) and 27.18% % (429 seconds), respectively.

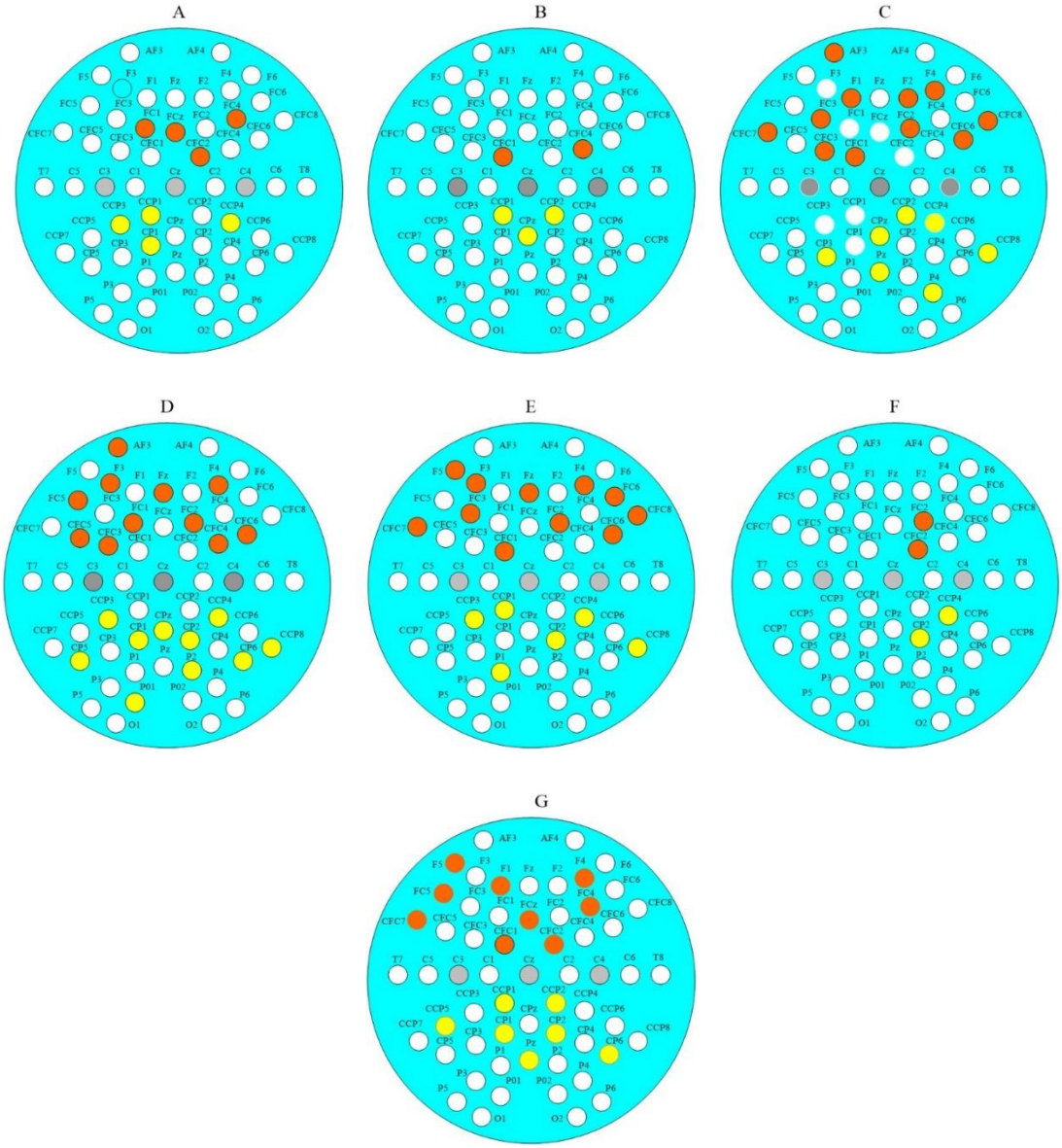

Figure 4.9. Distribution of the selected channels in dataset 2. Participants A, B, C, D, E, F, and G). The red circles represent selected channels from the frontal lobe while the yellow-colored circle shows the channel distribution in the parietal lobe.

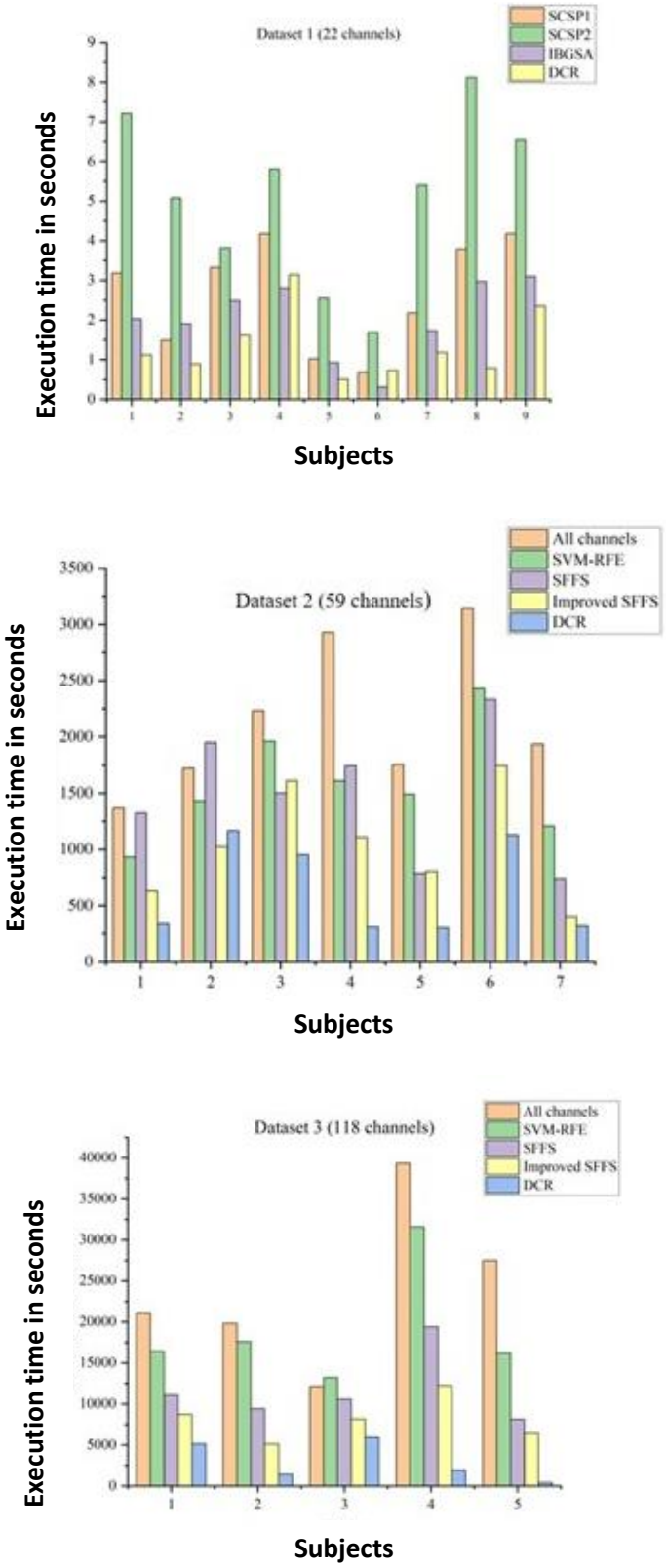

Figure 4.10. Computation time comparison between state-of-the-art channel selection methods on three BCI datasets with varying numbers of EEG channels

4.8.3.3. Results Implications

The proposed work aims to improve the classification accuracy of MI-based BCI systems using a minimum number of channels. This alleviates the human effort involved with the BCI preparation setup and reduces the computation burden due to a large number of channels. In the current research, our results suggest that the frontal and parietal cortex also play a significant role in the regulation of MI tasks. As we know BCI systems are used by individuals with motor disabilities and hence facilitate the best possible assistive support. Therefore, the results suggest that our approach may be useful to work on cognitive decline and motor control impairment simultaneously to boost the scope of conventional non-invasive brain-computer interface (BCI) systems. Similarly, our work can be extrapolated to study the Intracranial-EEG (icEEG) [102] behavior of frontal lobe epilepsy [103] where recurring seizures [104] arise in the frontal lobes of the brain. This process is time-consuming and might be erroneous because of manual methods. In such scenarios, our proposed classification model can effectively recognize the true cases with precise classification accuracy.

4.9. Limitations of the Study

The proposed DCR method yielded higher classification accuracy with a fewer number of channels. The performance of our method is validated on three standard BCI datasets with a different number of electrodes. Despite providing very robust results, our method has the following limitations:

I. The performance of the proposed method is primarily based on the effective link between neurophysiology and the candidate channels. In the case of a lack of information from the candidate channels, the performance of our method might degrade.

II. The proposed DCR method selects the relevant set of channels using the Mutual Information Coefficient (MIC). However, measuring and maximizing mutual information from finite data is a difficult training objective. Therefore, results may be less informative and error-prone.

III. Our classification model is implemented only on motor imagery datasets. Therefore, its scope is limited only to voluntary limb movements.

IV. Although we have validated our approach on three BCI datasets with different EEG channels, the number of participants and dataset size are 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.

4.10. Conclusion & Future Scope

Determining the optimal channel subset is crucial in developing a high-performance BCI system in terms of classification accuracy and computational complexity. In this work, we have developed a novel subject-specific channel selection scheme using Dynamic Channel Relevance (DCR) for the brain-computer interface. Dynamic channel relevance is a recently developed information-theoretic paradigm used for software feature selection. The advantage of the DCR implementation is its ability to investigate effective channels using the "maximum relevance with minimum redundancy" principle from a set of raw channels. The proposed method was applied to three different BCI competition datasets with a different number of electrodes to validate the performance. The results indicate the proposed method not only reduces computational complexity but also improves classification accuracy.

Finally, we have shown the relationship between selected channels and cognitive regions. Since the channel selection process is subject-specific, different sets of channels were found for each participant. Our experiment correlated frequently used channels with different cognitive regions and deduced that MI tasks are mostly distributed in the central, occipital, and parietal regions.

Some prominent applications of proposed methods can be used in different interdisciplinary research fields, such as sequence alignment, where positions of the new acid-base pairs are matched with true ones [105,106]. Here, the proposed candidate solution-based relevance search approach can be useful to find the importance of new acid-base pairs.