

Review

Not peer-reviewed version

A Mathematical Review on EEG Channel Selection Techniques for Motor Imagery Classification

[Seungjun Lee](#) *

Posted Date: 16 January 2025

doi: 10.20944/preprints202501.1148.v1

Keywords: Electroencephalography; EEG; Motor Imagery; Channel Selection



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Review

A Mathematical Review on EEG Channel Selection Techniques for Motor Imagery Classification

Seungjun Lee 

Korea Science Academy of KAIST, 105-47 Baegyongwanmun-ro, Busanjin-gu, 47162, Busan, Republic of Korea;
22-086@ksa.hs.kr

Abstract: Brain Computer Interface (BCI) technology is recently being spotlighted as the performance of Artificial Intelligence (AI) increased drastically. Although there are several instruments to measure brain's activity, Electroencephalography(EEG) signal is in limelight as it has high temporal resolution and is non-invasive, while being extremely portable. However, the raw EEG signals can not be directly used in BCI systems as it includes a lot of artifacts, and has numerous channels. Therefore, selecting only the needed channels and rejecting ones that include big noises are crucial to increase the performance of classification. Motor Imagery EEG (MI-EEG) is a type of EEG where subjects imagine that they are moving their body. MI-EEG channel selection techniques can be divided into two big groups: Common Spatial Pattern (CSP) based models and non-CSP based models. Therefore in this paper, we introduce the models classified in the criteria above, using detailed and strict mathematical terms, and compare each approach's accuracies.

Keywords: electroencephalography; EEG; motor imagery; channel selection

1. Introduction

Brain Computer Interface(BCI) systems measure the activities of the brain in order to perform specific tasks such as interpreting one's will to speak or move. While there are numerous methods to measure the brain's activity, such as EEG (electroencephalography), fMRI (functional magnetic resonance imaging), and MEG (Magnetoencephalography), BCI systems utilizing scalp EEG(henceforth called EEG) is trending as it has high temporal resolution, is non-invasive, and extremely portable.

Classification of EEG signals follows the procedure shown in Figure 1. [1,2]

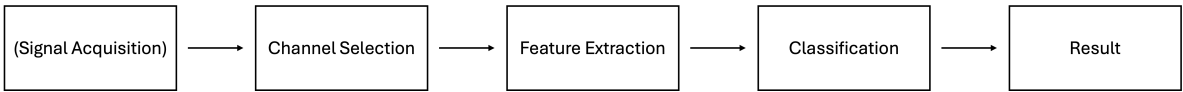


Figure 1. Procedure of EEG Classification

EEG signals are measured by attaching electrodes on the scalp according to the montage, which is a standardized arrangement of electrodes and selection of channel pairs. Montages are divided into two groups: referential and bipolar, by how channels are formed. In referential montage, the voltage difference between each electrode and one reference electrode is recorded, and forms a channel. In Bipolar montage, the voltage difference among consecutive two electrodes are recorded, and forms a channel. The recorded signals then go through artifact and noise removal process to remove undesired signals.[3]

Then, channel selection reduces the number of channels used in feature extraction and classification and ensures that only meaningful channels are included in the process afterwards. After the channel selection, only the necessary channels' data are needed to be acquired from subjects, so it can help reduce the time consuming setup process and inconvenience for subjects. Also, reducing the number of channels helps reduce computational complexity and increases classification accuracy. [2] Channel selection usually follows the following Figure 2:

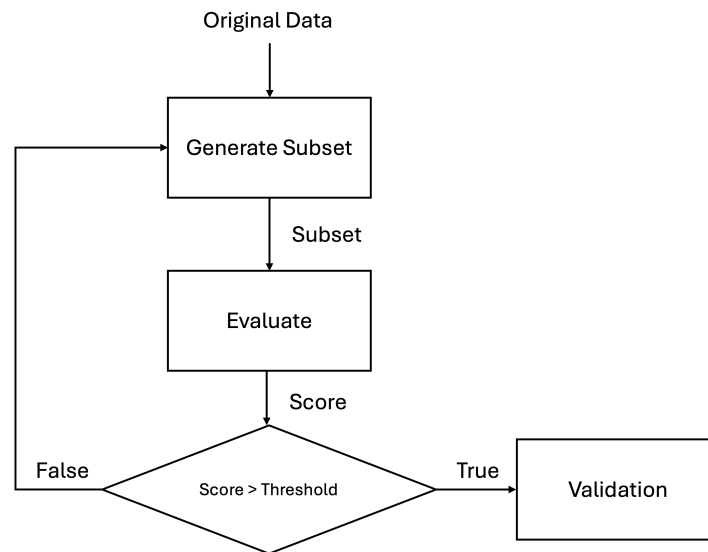


Figure 2. Procedure of Channel Selection

Motor Imagery EEG (MI-EEG) is a measurement of EEG signals while subject simulates its movement without actually moving. Therefore MI-EEG based BCIs have been suggested as a substitute method for control, especially for people who are disabled or suffering from locked-in syndrome.[4] It is widely accepted that imagination of a movement stimulates same neural pathway as real movement[5], and can even lead to muscle growth[6]. Traditional Statistical filters used for MI-EEG analysis can be divided into CSP (Common Spatial pattern)-based and non CSP-based filters.[2] Therefore, this paper will discuss about CSP based, filters and non-CSP based statistical filters used in MI-EEG classification in the following chapters.

2. Common Spatial Pattern(CSP) Based Filters

2.1. Original CSP Filter

The Common Spatial Pattern algorithm is known as a spatial filter effective for EEG filtering by maximizing one class's variance while minimizing other class's variance in n orthogonal bases as Figure 3, and can be mathematically induced as follows **Derivation**.



Figure 3. Example of CSP filtering ($n = 2$)

Derivation^[7-9]

Let $\mathbf{X} \in \mathbb{R}^{C \times N \times T}$ and $\mathbf{X}(t) \in \mathbb{R}^{C \times N}$ where C is the number of channels, N is the number of samples per epoch, and T is the number of epochs. $\mathbf{X}(t)$ represents the t th epoch's signal. In this paper, $\mathbf{X}$ only consists of two classes: cl_1 (class 1) and cl_2 (class 2) for simplicity. Then normalized covariance matrix of each classes: $\mathbf{C}_1$ and $\mathbf{C}_2$ is defined as

$$\mathbf{C}_i = \sum_{j \in cl_i} \frac{\mathbf{X}(j)\mathbf{X}(j)^T}{\text{tr}(\mathbf{X}(j)\mathbf{X}(j)^T)} \quad (1)$$

The covariance matrix is normalized with the sum of the diagonal elements of $\mathbf{X}(j)\mathbf{X}(j)^T$ to remove variances among trials.

As $\mathbf{X}(j)\mathbf{X}(j)^T$ is symmetric, $\mathbf{C}_i$ is symmetric and therefore if we define $\mathbf{C} \triangleq (\mathbf{C}_1 + \mathbf{C}_2)$, it can be orthogonally diagonalized as

$$\mathbf{C} = (\mathbf{C}_1 + \mathbf{C}_2) = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \quad (2)$$

where $\mathbf{\Lambda}$ is a diagonal matrix consisting eigenvalues and $\mathbf{U}$ is an orthonormal matrix whose columns are eigenvectors of $\mathbf{C}$ and $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ as $\mathbf{U}^T = \mathbf{U}^{-1}$.

Therefore,

$$\mathbf{U}^T(\mathbf{C}_1 + \mathbf{C}_2)\mathbf{U} = \mathbf{U}^T\mathbf{C}\mathbf{U} = \mathbf{\Lambda} \quad (3)$$

By applying whitening transform and simultaneously diagonalizing using method proposed in [9], we get

$$\bar{\mathbf{U}}^T(\mathbf{C}_1 + \mathbf{C}_2)\bar{\mathbf{U}} = \mathbf{I} \quad (4)$$

where $\bar{\mathbf{U}} = \mathbf{U}\mathbf{\Lambda}^{-\frac{1}{2}}$, and letting $\mathbf{S}_1 \triangleq \bar{\mathbf{U}}^T\mathbf{C}_1\bar{\mathbf{U}}$ and $\mathbf{S}_2 \triangleq \bar{\mathbf{U}}^T\mathbf{C}_2\bar{\mathbf{U}}$, $\mathbf{S}_1$ and $\mathbf{S}_2$ can be simultaneously diagonalized as

$$\mathbf{S}_1 = \mathbf{V}\mathbf{\Lambda}_1\mathbf{V}^T, \mathbf{S}_2 = \mathbf{V}\mathbf{\Lambda}_2\mathbf{V}^T \quad (5)$$

sharing same eigenvectors satisfying

$$\Lambda_1 + \Lambda_2 = \mathbf{I} \quad (6)$$

by properties of simultaneous diagonalization.

Equation (5), (6) are important properties as it shows that when spanned by $\mathbf{V}$, the bigger the variance of one class is in one axis, the smaller the variance of other class in corresponding axis.

By transforming Λ_1 and Λ_2 back to equation about $\mathbf{C}_1$ and $\mathbf{C}_2$, we get

$$\mathbf{Q}^T \mathbf{C}_1 \mathbf{Q} + \mathbf{Q}^T \mathbf{C}_2 \mathbf{Q} = \mathbf{I} \quad (7)$$

where $\mathbf{Q} = \mathbf{U} \Lambda^{-\frac{1}{2}} \mathbf{V}$. For an epoch $\mathbf{X} \in \mathbb{R}^{C \times N}$, we can calculate its CSP transformation as

$$\mathbf{X}_{\text{CSP}} = \frac{\text{diag}(\overline{\mathbf{Q}}^T \mathbf{X} \mathbf{X}^T \overline{\mathbf{Q}})}{\text{sum}(\text{diag}(\overline{\mathbf{Q}}^T \mathbf{X} \mathbf{X}^T \overline{\mathbf{Q}}))} \quad (8)$$

where $\overline{\mathbf{Q}}$ is a $2k \times C$ sized matrix which is composed of k best columns of $\mathbf{Q}$ that makes class 1's variance large, and k best columns that makes class 2's variance large. $\square$

2.2. Filter Bank CSP^[10]

Typical frequency band analyzed in MI-EEG is 8-30Hz[11,12]. However, as EEG signals have variance among trials and subjects, algorithms' performance can be enhanced by selecting specific frequency bands per subjects[13]. Therefore, the authors suggested to divide the frequency bands in small bands(4Hz in this paper), and apply CSP filters to each bands. This technique allowed automatic selection of meaningful frequency bands per subjects.

Although mutual information is not used very common in MI-EEG classification, MIBIF(Mutual Information-based Best Individual Feature) and MIRSR(Mutual Information-based Rough Set Reduction) was used to select features to be trained in this paper, which calculates mutual information among features and selects k features that best represent the data.

2.3. Sparse Common Spatial Pattern (SCSP)^[14]

CSP coefficients calculated above, $\mathbf{Q}$, is dense, meaning that most of the entries in the matrix can be considered meaningful, and only some have negligible effect. However, the matrix $\overline{\mathbf{Q}}$ that includes the selected channels' coefficients select only some of the columns of $\mathbf{Q}$; therefore the original signal can not be transformed into a shape that can be discriminated clearest. Therefore, the authors have sparsified the CSP spatial filters, which can increase performance as the removed features include relatively unnecessary information compared to original CSP filters.

As CSP algorithm's projection matrix $\mathbf{W}$ can be written as

$$\begin{aligned} & \min_{\mathbf{W}_i} \sum_{i=1}^{i=m} \mathbf{W}_i \mathbf{C}_1 \mathbf{W}_i^T + \sum_{i=m+1}^{i=2m} \mathbf{W}_i \mathbf{C}_2 \mathbf{W}_i^T \\ & \text{while } \mathbf{W}_i (\mathbf{C}_1 + \mathbf{C}_2) \mathbf{W}_i^T = 1, i = \{1, 2, \dots, 2m\} \\ & \text{and } \mathbf{W}_i (\mathbf{C}_1 + \mathbf{C}_2) \mathbf{W}_j^T = 0, i, j = \{1, 2, \dots, 2m\} i \neq j \end{aligned} \quad (9)$$

Sparsity of $\mathbf{W}$ can be increased with several methods, including L_0 norm, L_p norm, and L_1/L_2 norm, where L_p norm of $\mathbf{X}$ is defined as

$$\|\mathbf{X}\|_p = \left(\sum_{i=1}^n |\mathbf{X}|^p \right)^{1/p} \quad (10)$$

where $|\cdot|$ represents magnitude of vector $(\cdot)$, making L_0 norm the number of non-zero entries of $\mathbf{X}$ and L_1 norm the sum of magnitudes of vectors, also known as Manhattan distance.

According to [15], the L_1/L_2 norm is more suit as a measure of non-sparsity than L_1 and L_p norm as it is bounded within $(0, \sqrt{n})$, while L_1 and L_p norm does not have an upper bound. Also, L_1/L_2 norm does not depend on vectors' size, while L_1 and L_p depends on it.

Therefore, L_1/L_2 norm is adopted to (9) with the scaling factor r and $(1 - r)$ multiplied to it, forming

$$\begin{aligned} \min_{\mathbf{W}_i} (1-r) \sum_{i=1}^{i=m} \mathbf{W}_i \mathbf{C}_1 \mathbf{W}_i^T + \sum_{i=m+1}^{i=2m} \mathbf{W}_i \mathbf{C}_2 \mathbf{W}_i^T + r \sum_{i=1}^{2m} \frac{\|\mathbf{w}_i\|_1}{\|\mathbf{w}_i\|_2} \\ \text{while } \mathbf{W}_i (\mathbf{C}_1 + \mathbf{C}_2) \mathbf{W}_i^T = 1, i = \{1, 2, \dots, 2m\} \\ \text{and } \mathbf{W}_i (\mathbf{C}_1 + \mathbf{C}_2) \mathbf{W}_j^T = 0, i, j = \{1, 2, \dots, 2m\} i \neq j \end{aligned} \quad (11)$$

The original CSP algorithm is a complementary case when $r = 0$.

2.4. Robust Sparsd Common Spatial Pattern(RSCSP)^[16]

The RSCSP algorithm is a derivative algorithm of the SCSP algorithm developed to reduce the effect of outliers in the EEG signal. The non-robust SCSP algorithm calculates covariance matrix of class i : $\mathbf{C}_i$ as (1). As EEG signals are regularized, $\mathbf{C}_i$ can be written in the form of

$$\mathbf{C}_w = \frac{1}{(t \times n_w) - 1} \mathbf{E}_w \mathbf{E}_w^T \quad (12)$$

, where $w = 1, 2$ and $\mathbf{E}_w \in \mathbb{R}^{C \times (t \cdot n_w)}$ denotes EEG signals of class w concatenated by channels. As outliers can distort covariance matrices, MCD(Minimum Covariance Determinant) is used. MCD can be described as a process to find a subset H of original matrix $X = [c_1, c_2, \dots, c_n]$, such that its determinant $\det(\Sigma_H)$ is reduced, and can be mathematically represented as an optimization problem of $\min_{H \subset X, |H|=h} \det(\Sigma_H)$ where

$$\mathbf{C}_w = \frac{1}{(\alpha \times t \times n_w) - 1} \mathbf{E}_w \mathbf{E}_w^T \quad (13)$$

when $1 - \alpha$ is the breakdown point. FASTMCD algorithm can also be adopted to decrease the complexity of the calculation. FASTMCD algorithm first selects random h points and updates the points by ranking the data by their Mahalanobis distances and select h points with the smallest distances to form a new subset. Then if the new subset has a smaller determinant of covariance matrix, it functions as a new initial set and continues this process until the covariance matrix converges.

3. CSP Based Channel Selection Techniques

3.1. Bhattacharyya Bound of CSP Features^[7]

In [7], the authors reduced the number of channels by selecting a set of channels using forward search. Bhattacharyya bound, which is the upper bound of bayes error probability was used as evaluation, as the calculation of Bayes error probability is challenging. The statement: *Bhattacharyya bound is the upper bound of Bayes error probability* can be proved by **Proof**.

To summarize, the channels are selected by **Algorithm 1**

Proof.

The Bayes error probability P_e can be written as

$$P_e \triangleq \int_X \min(P(C_1)p(x|C_1), P(C_2)p(x|C_2))dx \quad (14)$$

where $P(C_1)$ and $P(C_2)$ are the prior probabilities of class 1 and 2, respectively. $p(x|C_1)$ and $p(x|C_2)$ are are class conditional probability density functions, and x represents an observed data point.

Since $\min(a, b) \leq \sqrt{ab}$,

$$\min(P(C_1)p(x|C_1), P(C_2)p(x|C_2)) \leq \sqrt{P(C_1)P(C_2)p(x|C_1)p(x|C_2)} \quad (15)$$

and integrating both sides over $\mathbf{X}$ results

$$P_e \leq \int_{\mathbf{X}} \sqrt{P(C_1)P(C_2)p(x|C_1)p(x|C_2)} dx \quad (16)$$

$$= \sqrt{P(C_1)P(C_2)} \int_{\mathbf{X}} \sqrt{p(x|C_1)p(x|C_2)} dx \quad (17)$$

By [17], (17) is a Bhattacharyya bound, which is a complementary case of Chernoff bound where $\beta = 0.5$, yielding

$$P_e \leq \sqrt{P(C_1)P(C_2)} e^{-k(1/2)} \quad (18)$$

where

$$k(1/2) = \frac{1}{8}(\mu_2 - \mu_1)^T \left[\frac{\Sigma_1 + \Sigma_2}{2} \right]^{-1} (\mu_2 - \mu_1) + \frac{1}{2} \ln \frac{|\frac{\Sigma_1 + \Sigma_2}{2}|}{\sqrt{|\Sigma_1||\Sigma_2|}} \quad (19)$$

and μ_i and Σ_i represent mean vectors and covariance matrix of class i 's CSP feature vectors, respectively.

Algorithm 1 Bhattacharyya Bound Based Channel Selection

```
// Initialization
S ← set of channels
S0 ← top ni channels with highest Fisher ratio of CSP features
nw ← number of channel wanted
Bmin ← Bhattacharyya(S0)

// Forward Search
for j = 0 ; j < nw - ni ; j++
    St ← S \ S0
    Bt ← Bhattacharyya(St[0] ∪ S0)
    Ct ← St[0]
    foreach C in S \ S0
        Btmp ← Bhattacharyya(C ∪ S0)
        if Btmp < Bt
            Bt ← Btmp
            Ct ← C
    S0 ← S0 ∪ Ct
return S0
```

3.2. CSP-Rank^[18]

CSP-Rank is based on the idea that the matrix $\bar{Q}$ in (8) includes the importance of each channel in classification as it acts as a weight to the original signal. The authors of this paper therefore set $k = 1$ and selected two vectors: SF_1 and SF_2 , where SF_i maximizes class i 's variance. Therefore, the channels with the largest entries in each of $SF_1, SF_2, SF_3, \dots$ are consecutively selected until the model trained with the selected channels reaches the desired accuracy.

The results are statistically significant at $p < 0.05$ significance compared to random selection. Also, the number of channels needed to reach 90% accuracy was 8-38 channels, while SVM-RFE(Support-vector machine recursive feature elimination, explained in Section 4.1) required 12-28 channels, therefore being better than SVM-RFE.

3.3. L₁ Norm of CSP Filter^[19]

Since it is widely accepted that the effectiveness of the CSP algorithm decreases as the number of channels increases due to overfitting, the authors first applied the CSP filter to all channels and scored

each channel based on its L_1 norm. Then, the channels with high scores were selected and used to train the model. The following score is used for calculating channels' weight:

$$SC(i) = \frac{\|\bar{Q}_i\|_1}{\|\bar{Q}\|_1} \quad (20)$$

where $\bar{Q} \in \mathbb{R}^{2k \times C}$ is the CSP projection matrix from (8), and $\bar{Q}_i$ is the i th column of $\bar{Q}$, and can be ordered in decreasing order to select wanted number of channels.

3.4. Cohen's d Effect Size CSP(E-CSP)^[20]

The E-CSP algorithm aims to remove channels that contain redundant information. The algorithm is divided into two major parts: part 1, where the noisy channels are removed, and part 2, where the remaining channels are evaluated to determine if they are meaningful.

First, to remove trials with large artifacts, the z-score calculated by equation, z-score, calculated by (21) is used.

$$Z_{li}^j = \frac{|C_{li}^j - \bar{C}_{li}|}{\sigma} \quad (21)$$

where i, j, l represents channel number, trial, and class, respectively, and σ is the standard deviation of l th task's i th channel, is calculated. The z-score was adopted as it measures how far the signal deviates from the average. Therefore, the i th channel's j th trial can be considered noisy if $z_{li}^j > \alpha$, where α is the criterion of noisy channel. Therefore, let $s_{ij}^l = \begin{cases} 1, & (z_{li}^j > \alpha) \\ 0, & (else) \end{cases}$ and let the frequency of l th class' j th trial being noisy as

$$f_l^j = \frac{\sum_{i=1}^N s_{ij}^l}{N} \times 100 \quad (22)$$

If the trial's frequency exceeds the threshold $\beta\%$, it is considered as a noisy trial. Therefore, we can define the set of noisy trials as

$$K_l = \{j | f_l^j > \beta\}; \forall j \quad (23)$$

Therefore, after removing the noisy channels, the i th channel's mean for j th task can be calculated as

$$\bar{c}_{li} = \frac{1}{|J_s^l|} \sum_{j \in J_s^l} c_{li}^j \quad (24)$$

where J_s^l represents non-noisy channels.

Second, the effect size d_i of channel i is evaluated as a meaningful channel if

$$d_i = \frac{|\bar{c}_{1i} - \bar{c}_{2i}|}{\sigma} > \gamma \quad (25)$$

where

$$\sigma = \frac{\sigma_{1i}^j + \sigma_{2i}^j}{2}; \forall j \in J_s^l \quad (26)$$

and σ_{li}^j stands for the standard deviation of l th class' i th channel, and γ represents the criterion for a channel to be thought as meaningless, as the variance among two class is small.

Finally, the leftover channels L , are selected and CSP filter is applied.

4. Non-CSP Based Methods

4.1. Support-Vector Machine Recursive Feature Elimination (SVM-RFE)^[21,22]

SVM(Hard Margin Support Vector Machine) Algorithm

Support vector machine is an algorithm that aims to find a hyperplane that maximizes the margin among the given classes. Let there be N data points: $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ where d

is the dimension of the point, $x_i \in \mathbb{R}^d$ and $y_i \in \{1, 2\}$ be a label. The hyperplane can be written as $follows(x) = w^T x + b$ where w is the weight and b is the bias. The optimization problem of w and b to maximize the margin can be written as

$$J = \min_{w,b} \frac{1}{2} \|w\|^2 \text{ subject to } y_i(w^T x_i + b) \geq 1, \forall i \quad (27)$$

, and this problem can be solved using the Lagrange multiplier.

Let Lagrange function

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i (y_i (w^T x_i + b) - 1) \quad (28)$$

, where $\alpha \geq 0$ is the Lagrange multiplier. By partial differentiating with w , we get

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^N \alpha_i x_i y_i = 0 \implies w = \sum_{i=1}^N \alpha_i x_i y_i \quad (29)$$

By partial differentiating with b ,

$$\frac{\partial L}{\partial b} = - \sum_{i=1}^N \alpha_i y_i = 0 \implies \sum_{i=1}^N \alpha_i y_i = 0 \quad (30)$$

Therefore, we get a dual Lagrangian function

$$L_D(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i^T x_j, \quad (31)$$

making (27)

$$\max \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \text{ subject to } \sum_{i=1}^N \alpha_i y_i = 0 \quad (32)$$

Soft Margin SVM Algorithm

However, as real datasets are not discriminated using linearly, slack variable ξ_i is adopted, changing the objective (27) into

$$J = \min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \text{ subject to } y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \forall i \quad (33)$$

and can be derived into a same shape as (32) using the same method from above as $C \sum_{i=1}^N \xi_i$ is a constant term.

SVM-RFE Algorithm

SVM-RFE algorithm aims to find a subset of size r among n variables ($r < n$) which maximizes the performance of the classifier using backward sequential selection.

From previous research [22,23], it is well known that ΔJ can be used as a criterion for elimination. The equation (33) can be transformed into

$$\Delta J(i) = \frac{\partial J}{\partial w_i} \Delta w_i + \frac{\partial^2 J}{\partial w_i^2} (\Delta w_i)^2 \quad (34)$$

using Taylor series to the second order. As the objective function J reaches its optimal point, $\frac{\partial J}{\partial w_i} \Delta w_i = 0$. Therefore,

$$\Delta J(i) = \frac{\partial^2 J}{\partial w_i^2} (\Delta w_i)^2 \quad (35)$$

. When removing i th feature, the weight w_i must be 0, therefore $\Delta w_i = w_i$. Therefore, the criterion for removing the i th feature can be approximated as

$$C_i = \Delta J(i) = (w_i)^2 \quad (36)$$

4.2. γ^2 Value Based Channel Selection

Similar to how the Bhattacharyya bound-based CSP selection used the F-ratio for selecting the initial subset of channels, the authors of [19] have utilized the γ^2 value as a control group. γ^2 is defined as

$$\gamma^2(i) = \left(\frac{\sqrt{n_1 n_2} \overline{\|x_i^{(1)}\|_2} - \overline{\|x_i^{(2)}\|_2}}{n_1 + n_2 \text{stdev}(\|x_i^{(1,2)}\|_2)} \right)^2 \quad (37)$$

where $k \in \{1(\text{alias of } cl_1), 2(\text{alias of } cl_2)\}$, $x_i^{(k)} \in \mathbb{R}^T$ denotes i th channel of $X^{(k)}$, which means $X^{(k)} = [x_1^{(k)}, x_2^{(k)}, \dots, x_N^{(k)}]^T$, and n_k denotes number of epochs in class k .

4.3. Correlation Based Channel Selection^[24,25]

EEG has a limitation that even though one is repeating the same task, due to other activities and thoughts in the brain, different EEG is recorded, and makes it hard to analyze the task performed. The authors assumed that as channels related to MI-EEG while performing same task will contain similar information across trials, thus having high correlation among trials, while channels that are not related will not have alike information, therefore having low correlation.

Pearson's correlation is used in both papers, which measures linear dependency among random variables, and is calculated by

$$\rho(X, Y) = \frac{1}{N-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma_X} \right) \left(\frac{Y_i - \bar{Y}}{\sigma_Y} \right) \quad (38)$$

where X, Y are random variables, $\bar{X}, \bar{Y}$ are means of random variables, and σ_X, σ_Y represents standard deviation of X and Y respectively. Correlations are calculated among every channels, forming a correlation matrix of size $C \times C$. By averaging each row, we can find the channel that is most effective. By repeating this procedure among every epochs, we can select N_s number of channels by selecting the channels that are considered to be effective over certain criteria times.

4.4. Genetic Algorithm (GA) Based Selection

Genetic Algorithm(henceforth GA) is a heuristic methodology for finding an optimal solution by mimicking Charles Darwin's natural evolution theory. Genetic algorithm mimics how genetic information are passed over through generations. Like DNA is formed with genes, a gene is an element that forms a chromosome. A chromosome is a set of genes that forms a potential solution to the problem. The fitness function evaluates how well each chromosome solves the problem. The scores calculated are used for selection, which is a process of selecting chromosomes that will create an offspring. The higher the score is, the higher chance it has to be selected. Like our DNA, an offspring is created by combining the genetic information of two chromosomes. The point where that crossover will be made is randomly selected. For example, aaaa and bbbb can become aaaa, abbb, aabb, abbb, and bbbb. Mutation happens randomly at each position, changing the value from 0 to 1 or 1 to 0.

The genetic algorithm follows the flow chart Figure 4:

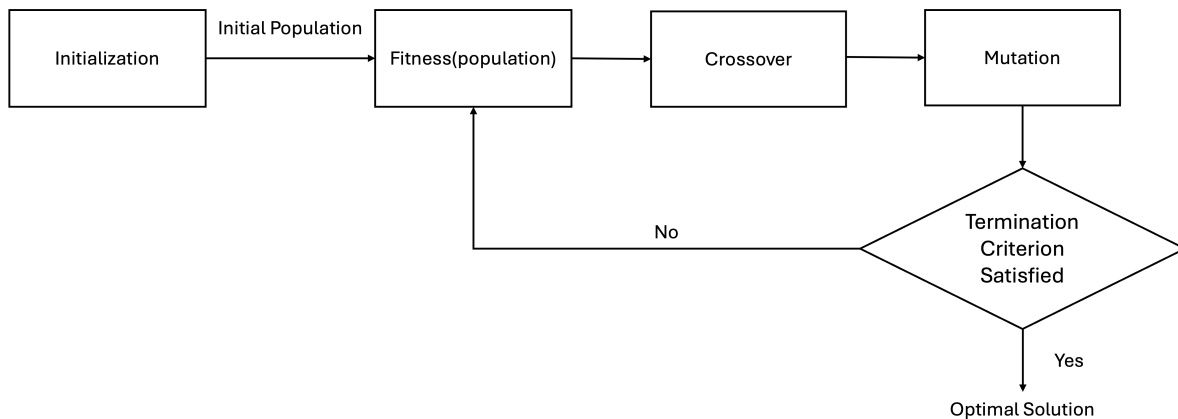


Figure 4. Flow Chart of Genetic Algorithm

Genetic Algorithm For Selecting Optimal Channels for P300 Based Classification [26,27]

EEG signals can be roughly divided into four categories: Visual Evoked Potential (VEP), Sensorimotor Rythm (SRM), Slow Cortical Potentials (SCP), and Event-Related Potentials (ERP). VEP is a signal acquired from three occipital electrodes, while forming two channels by setting the middle electrode as a reference, and represents the pathway of visual activity, from optic nerve to primary optic cortex. SRM is a signal recorded from sensorimotor cortex at mu(7–13 Hz), beta(13–30 Hz), and gamma(30–200 Hz) frequency bands, and is strongly modulated with simple movements and certain cognitive tasks. SCP is a measurement of a slowly changing electrical signal with frequency lower than 1Hz. ERP is not a term referencing an EEG measured in specific location, but is a fluctuation generated by neurons due to some event, such as motor or speech imagery, at a specific time. [26,28–34].

P300 is an unique ERP, which appears as a positive peak near 300ms after stimulation happens. A 64 channel EEG was recorded and four channels were selected in [26], by defining the fitness function $f(X) = \sum_{k=1}^n A_K$ where X is a chromosome consisting four genes, which are channels in this case. A_k is defined as the maximum r^2 among four channels, where r^2 is a value between 0 and 1, and higher the r^2 is, the higher the variance among classes are.

5. Result

In this study, we utilized the EEG dataset from BCI Competition IV Dataset 2a, which is specifically designed for benchmarking motor imagery classification algorithms. The dataset comprises EEG recordings from nine healthy subjects, each performing four different motor imagery tasks: imagining movements of the left hand, right hand, both feet, and tongue. EEG signals were recorded using 22 Ag/AgCl electrodes placed according to the international 10-20 system at a sampling rate of 250Hz [35].

For preprocessing, we applied a band-pass filter between 8Hz and 30Hz, a frequency band commonly used in previous research[11,12]. Baseline correction was performed by subtracting the mean of a 2 second pre-cue interval, mentioned as "Fixation cross" (from -2s to 0s relative to the cue onset), from the subsequent motor imagery data. The entire 4seconds of motor imagery data (from 0s to 4s after the cue onset) were used for classification.

Data were split into training and evaluation datasets using 5-fold cross-validation, ensuring that the classes were balanced in terms of the number of trials in each fold. Cross-validation was performed individually for each subject to account for inter-subject variability of EEG signals.

The accuracies of each algorithm were measured by training a support vector machine (SVM) classifier with a linear kernel using the data from the selected channels.

5.1. CSP and SCSP Algorithm

CSP algorithm can be thought of as a special case of the SCSP algorithm where $r = 0$. The accuracy of SCSP algorithm depends on its r value, so accuracy of SCSP algorithm where $r = 0, 0.1, 0.2, 0.3, 0.4, 0.5$ is compared as Figure 5.

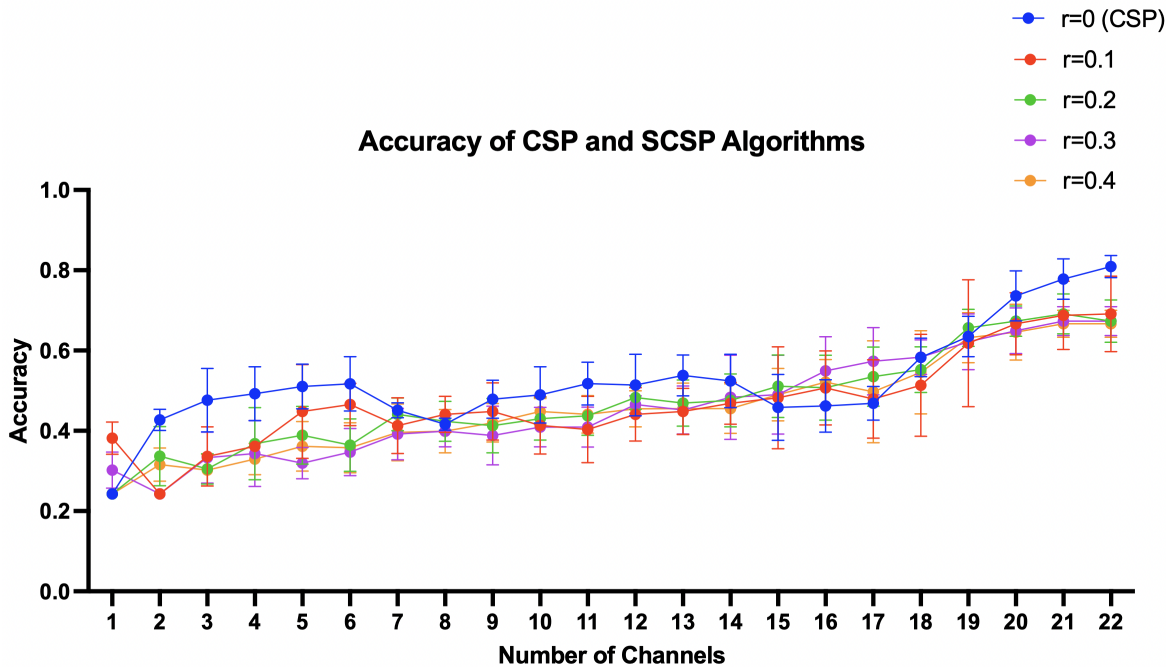


Figure 5. Performance Comparison of SCSP Algorithms With Different r Value Across Different Channels

5.2. Accuracy of CSP Based Channel Selection Methods

The accuracy of CSP based channel selection methods: CSPRank(Section 3.2), L1 Norm applied CSPRank(Section 3.3), SCSPRank(SCSP used instead of CSP in CSPRank), E-CSP(Section 3.4), and Bhattacharyya bounded CSP(Section 3.1) is as following Table 1 and Figure 6.

Table 1. Comparison of Classification Performance for Various CSP Based EEG Channel Selection Methods Across Different Numbers of Channels.

Number of Channels	1	2	3	4	5	6	
CSPRank	0.243±0.002	0.427±0.026	0.476±0.079	0.493±0.068	0.51±0.056	0.517±0.068	
L1 Norm applied CSPRank	0.243±0.002	0.344±0.055	0.433±0.038	0.458±0.044	0.451±0.032	0.453±0.014	
SCSPRank	0.243±0.002	0.303±0.029	0.361±0.071	0.384±0.065	0.365±0.055	0.422±0.019	
E-CSP	0.385±0.054	0.406±0.036	0.451±0.044	0.458±0.043	0.49±0.059	0.601±0.053	
Bhattacharyya Bounded CSP						0.712±0.009	
7	8	9	10	11	12	13	14
0.451±0.018	0.417±0.015	0.479±0.048	0.49±0.07	0.518±0.053	0.514±0.076	0.538±0.051	0.524±0.06
0.434±0.031	0.427±0.02	0.423±0.022	0.439±0.026	0.416±0.03	0.447±0.059	0.446±0.032	0.48±0.088
0.43±0.028	0.462±0.029	0.464±0.05	0.488±0.052	0.477±0.058	0.45±0.082	0.45±0.039	0.468±0.069
0.621±0.071	0.673±0.055	0.701±0.057	0.691±0.063	0.753±0.051	0.722±0.066	0.733±0.071	0.733±0.033
0.712±0.034	0.729±0.027	0.716±0.055	0.733±0.036	0.743±0.038	0.761±0.024	0.781±0.028	0.809±0.046
15	16	17	18	19	20	21	22
0.458±0.082	0.462±0.065	0.469±0.042	0.583±0.048	0.635±0.051	0.736±0.063	0.778±0.051	0.809±0.028
0.44±0.063	0.438±0.064	0.47±0.026	0.526±0.062	0.579±0.07	0.607±0.069	0.738±0.032	0.755±0.032
0.516±0.059	0.571±0.042	0.551±0.07	0.522±0.08	0.648±0.051	0.712±0.018	0.726±0.018	0.75±0.048
0.736±0.061	0.754±0.059	0.747±0.054	0.75±0.047	0.75±0.041	0.754±0.06	0.761±0.057	0.757±0.056
0.809±0.056	0.813±0.045	0.813±0.043	0.809±0.033	0.799±0.035	0.806±0.027	0.816±0.035	0.82±0.035

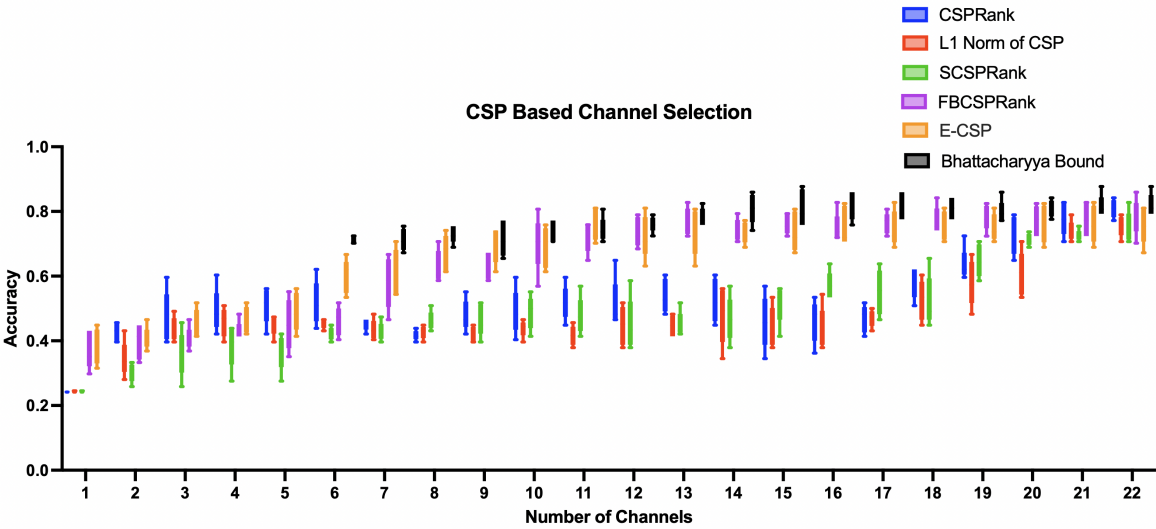


Figure 6. Performance of CSP Based Channel Selection Methods Across Different Channels

5.3. Accuracy of Non-CSP Based Statistical Channel Selection Methods

The accuracy of Non-CSP based channel selection methods: SVM-RFE(Section 4.1), γ^2 value based channel selection(Section 4.2), Correlation based channel selection(Section 4.3), and Genetic Algorithm based channel selection(Section 4.4) is as following Table 2 and Figure 7.

Table 2. Comparison of Classification Performance for Various Non-CSP Based Channel Selection Methods Across Different Numbers of Channels.

Number of Channels	1	2	3	4	5	6	
SVM-RFE	0.299±0.069	0.455±0.048	0.527±0.068	0.6±0.072	0.597±0.034	0.653±0.032	
γ^2 based selection						0.726±0.087	
Correlation based selection	0.292±0.035	0.444±0.038	0.562±0.111	0.521±0.086	0.615±0.067	0.68±0.047	
GA based selection	0.371±0.058	0.399±0.054	0.406±0.036	0.43±0.029	0.451±0.079	0.462±0.045	
7	8	9	10	11	12	13	14
0.667±0.03	0.67±0.033	0.684±0.023	0.677±0.006	0.684±0.021	0.712±0.047	0.76±0.059	0.757±0.059
0.698±0.042	0.736±0.047	0.74±0.075	0.753±0.077	0.729±0.046	0.746±0.053	0.719±0.077	0.736±0.058
0.712±0.03	0.705±0.043	0.74±0.058	0.729±0.083	0.722±0.066	0.74±0.067	0.747±0.058	0.74±0.053
0.58±0.08	0.628±0.05	0.639±0.044	0.702±0.085	0.722±0.046	0.736±0.045	0.764±0.043	0.74±0.035
15	16	17	18	19	20	21	22
0.75±0.06	0.743±0.072	0.753±0.082	0.771±0.085	0.771±0.081	0.747±0.049	0.746±0.043	0.757±0.056
0.694±0.054	0.757±0.047	0.764±0.042	0.757±0.047	0.743±0.046	0.76±0.046	0.757±0.058	0.757±0.047
0.732±0.053	0.75±0.06	0.75±0.062	0.754±0.059	0.754±0.059	0.747±0.067	0.764±0.05	0.764±0.053
0.76±0.03	0.75±0.044	0.764±0.032	0.771±0.042	0.781±0.039	0.771±0.046	0.774±0.051	0.781±0.056

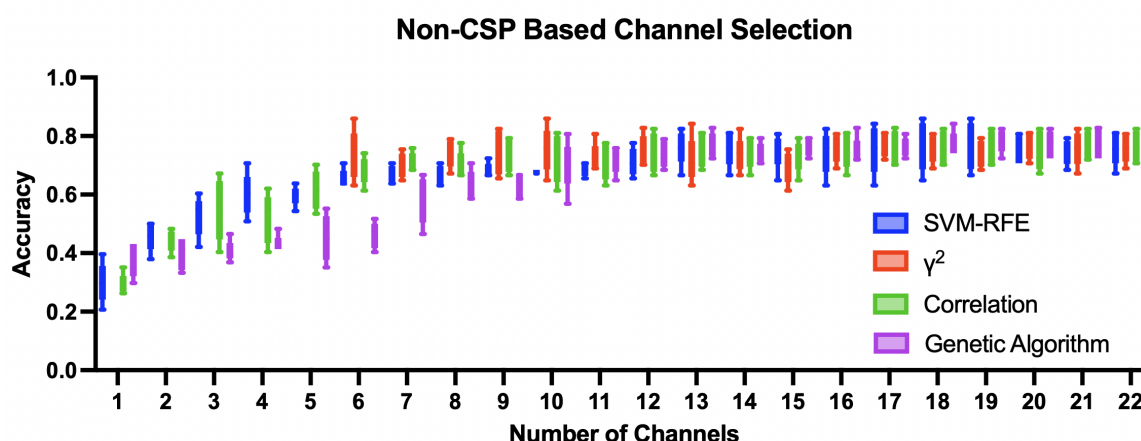


Figure 7. Performance of Non-CSP Based Channel Selection Methods Across Different Channels

References

- Alotaiby, T.; El-Samie, F.E.A.; Alshebeili, S.A.; Ahmad, I. A review of channel selection algorithms for EEG signal processing. *EURASIP Journal on Advances in Signal Processing* **2015**, *2015*, 1–21.
- Baig, M.Z.; Aslam, N.; Shum, H.P. Filtering techniques for channel selection in motor imagery EEG applications: a survey. *Artificial intelligence review* **2020**, *53*, 1207–1232.
- Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*.
- Choi, J.; Kaongoen, N.; Jo, S. Investigation on Effect of Speech Imagery EEG Data Augmentation with Actual Speech. In Proceedings of the 2022 10th International Winter Conference on Brain-Computer Interface (BCI), pp. 1–5. <https://doi.org/10.1109/BCI53720.2022.9735108>.
- Roth, M.; Decety, J.; Raybaudi, M.; Massarelli, R.; Delon-Martin, C.; Segebarth, C.; Morand, S.; Gemignani, A.; Décorps, M.; Jeannerod, M. Possible involvement of primary motor cortex in mentally simulated movement: a functional magnetic resonance imaging study. *NeuroReport* **1996**, *7*.
- Yue, G.; Cole, K.J. Strength increases from the motor program: Comparison of training with maximal voluntary and imagined muscle contractions. *Journal of Neurophysiology* **1992**, *67*, 1114–1123. <https://doi.org/10.1152/jn.1992.67.5.1114>.
- Lin, H.; Zhuliang, Y.; Zhenghui, G.; Yuanqing, L. Bhattacharyya bound based channel selection for classification of motor imageries in EEG signals. In Proceedings of the 2009 Chinese Control and Decision Conference, pp. 2353–2356. <https://doi.org/10.1109/CCDC.2009.5192711>.
- Ramoser, H.; Müller-Gerking, J.; Pfurtscheller, G. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans Rehabil Eng* **2000**, *8*, 441–6. <https://doi.org/10.1109/86.895946>.
- Fukunaga, K., Chapter 2 - RANDOM VECTORS AND THEIR PROPERTIES. In *Introduction to Statistical Pattern Recognition (Second Edition)*, Second ed.; Fukunaga, K., Ed.; Academic Press: Boston, 1990; pp. 11–50. <https://doi.org/https://doi.org/10.1016/B978-0-08-047865-4.50008-9>.
- Ang, K.K.; Chin, Z.Y.; Wang, C.C.; Guan, C.T.; Zhang, H.H. Filter bank common spatial pattern algorithm on BCI competition IV Datasets 2a and 2b. *Frontiers in Neuroscience* **2012**, *6*. <https://doi.org/10.3389/fnins.2012.00039>.
- Blankertz, B.; Tomioka, R.; Lemm, S.; Kawanabe, M.; Müller, K.R. Optimizing Spatial filters for Robust EEG Single-Trial Analysis. *IEEE Signal Processing Magazine* **2008**, *25*, 41–56. <https://doi.org/10.1109/MSP.2008.4408441>.
- Müller-Gerking, J.; Pfurtscheller, G.; Flyvbjerg, H. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical Neurophysiology* **1999**, *110*, 787–798. [https://doi.org/https://doi.org/10.1016/S1388-2457\(98\)00038-8](https://doi.org/https://doi.org/10.1016/S1388-2457(98)00038-8).
- Blankertz, B.; Dornhege, G.; Krauledat, M.; Müller, K.R.; Curio, G. The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage* **2007**, *37*, 539–550. <https://doi.org/https://doi.org/10.1016/j.neuroimage.2007.01.051>.

14. Arvaneh, M.; Guan, C.; Ang, K.K.; Quek, C. Optimizing the Channel Selection and Classification Accuracy in EEG-Based BCI. *IEEE Transactions on Biomedical Engineering* **2011**, *58*, 1865–1873. <https://doi.org/10.1109/TBME.2011.2131142>.
15. Hurley, N.; Rickard, S. Comparing Measures of Sparsity. *IEEE Transactions on Information Theory* **2009**, *55*, 4723–4741. <https://doi.org/10.1109/TIT.2009.2027527>.
16. Arvaneh, M.; Guan, C.; Ang, K.K.; Quek, C. Robust EEG channel selection across sessions in brain-computer interface involving stroke patients. In Proceedings of the The 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1–6. <https://doi.org/10.1109/IJCNN.2012.6252687>.
17. Duda, R.O.; Hart, P.E.; Stork, D.G., Chapter2 - Bayesian decision theory. In *Pattern Classification, 2nd Edition*, 2 ed.; Duda, R.O.; Hart, P.E.; Stork, D.G., Eds.; Wiley, 2000; book section 2, pp. 1–32.
18. Tam, W.K.; Ke, Z.; Tong, K.Y. Performance of common spatial pattern under a smaller set of EEG electrodes in brain-computer interface on chronic stroke patients: A multi-session dataset study. In Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6344–6347. <https://doi.org/10.1109/IEMBS.2011.6091566>.
19. Meng, J.; Liu, G.; Huang, G.; Zhu, X. Automated selecting subset of channels based on CSP in motor imagery brain-computer interface system. In Proceedings of the 2009 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 2290–2294. <https://doi.org/10.1109/ROBIO.2009.5420462>.
20. Das, A.K.; Suresh, S. An Effect-Size Based Channel Selection Algorithm for Mental Task Classification in Brain Computer Interface. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, pp. 3140–3145. <https://doi.org/10.1109/SMC.2015.545>.
21. Qifeng, Z.; Wencai, H.; Guifang, S.; Weiyu, C. A new SVM-RFE approach towards ranking problem. In Proceedings of the 2009 IEEE International Conference on Intelligent Computing and Intelligent Systems, Vol. 4, pp. 270–273. <https://doi.org/10.1109/ICICISYS.2009.5357684>.
22. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* **2002**, *46*, 389–422. <https://doi.org/10.1023/A:1012487302797>.
23. LeCun, Y.; Denker, J.; Solla, S. Optimal brain damage. *Advances in neural information processing systems* **1989**, *2*.
24. Jin, J.; Miao, Y.Y.; Daly, I.; Zuo, C.L.; Hu, D.W.; Cichocki, A. Correlation-based channel selection and regularized feature optimization for MI-based BCI. *Neural Networks* **2019**, *118*, 262–270. <https://doi.org/10.1016/j.neunet.2019.07.008>.
25. Hsu, W.Y.; Cheng, Y.W. EEG-channel-temporal-spectral-attention correlation for motor imagery EEG classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **2023**, *31*, 1659–1669.
26. Hasan, I.H.; Ramli, A.R.; Ahmad, S.A. Utilization of Genetic Algorithm for Optimal EEG Channel Selection in Brain-Computer Interface Application. In Proceedings of the 2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology, pp. 97–102. <https://doi.org/10.1109/ICAJET.2014.25>.
27. Kee, C.Y.; Kuppan Chetty, R.M.; Khoo, B.H.; Ponnambalam, S.G. Genetic Algorithm and Bayesian Linear Discriminant Analysis Based Channel Selection Method for P300 BCI. In Proceedings of the Trends in Intelligent Robotics, Automation, and Manufacturing; Ponnambalam, S.G.; Parkkinen, J.; Ramanathan, K.C., Eds. Springer Berlin Heidelberg, pp. 226–235.
28. Brenner, R., CHAPTER 77 - INVESTIGATIONS IN MULTIPLE SCLEROSIS. In *Neurology and Clinical Neuroscience*; Schapira, A.H.V.; Byrne, E.; DiMauro, S.; Frackowiak, R.S.J.; Johnson, R.T.; Mizuno, Y.; Samuels, M.A.; Silberstein, S.D.; Wszolek, Z.K., Eds.; Mosby: Philadelphia, 2007; pp. 1031–1044. <https://doi.org/https://doi.org/10.1016/B978-0-323-03354-1.50081-X>.
29. Gibson, R.M.; Owen, A.M.; Cruse, D., Chapter 9 - Brain-computer interfaces for patients with disorders of consciousness. In *Progress in Brain Research*; Coyle, D., Ed.; Elsevier, 2016; Vol. 228, pp. 241–291. <https://doi.org/https://doi.org/10.1016/bs.pbr.2016.04.003>.
30. Samar, V.J., Evoked Potentials. In *Encyclopedia of Language & Linguistics (Second Edition)*; Brown, K., Ed.; Elsevier: Oxford, 2006; pp. 326–335. <https://doi.org/https://doi.org/10.1016/B0-08-044854-2/02397-X>.
31. Cinzia Baiano, M.Z. Visual Evoked Potential, 2023 May 11.
32. Schmidt, S.; Jo, H.G.; Wittmann, M.; Hinterberger, T. ‘Catching the waves’ – slow cortical potentials as moderator of voluntary action. *Neuroscience & Biobehavioral Reviews* **2016**, *68*, 639–650. <https://doi.org/https://doi.org/10.1016/j.neubiorev.2016.06.023>.
33. Cheyne, D.O. MEG studies of sensorimotor rhythms: A review. *Experimental Neurology* **2013**, *245*, 27–39. <https://doi.org/https://doi.org/10.1016/j.expneurol.2012.08.030>.

34. Mudgal, S.K.; Sharma, S.K.; Chaturvedi, J.; Sharma, A. Brain computer interface advancement in neurosciences: Applications and issues. *Interdisciplinary Neurosurgery* **2020**, *20*, 100694. <https://doi.org/https://doi.org/10.1016/j.inat.2020.100694>.
35. Tangermann, M.; Müller, K.R.; Aertsen, A.; Birbaumer, N.; Braun, C.; Brunner, C.; Leeb, R.; Mehring, C.; Miller, K.J.; Mueller-Putz, G.; et al. Review of the BCI Competition IV. *Frontiers in Neuroscience*, *6*. <https://doi.org/10.3389/fnins.2012.00055>.



Seungjun Lee is a senior student at Korea Science Academy of KAIST. His research interest is bio-signal processing, especially EEG data

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