Channel Selection Methods Evaluation for EEG Motor Imagery Classification with Neural Network Architectures for Evidence of a Global Channel Set

Carlos Simon Amador[1], Dusthon Llorente-Vidrio[1], and Rita Q. Fuentes-Aguilar[2][0000−0003−2559−539X]

[1] School of Engineering and Sciences, Tecnologico de Monterrey
[2] Institute of Advanced Materials for Sustainable Manufacturing, Tecnologico de Monterrey
rita.fuentes@tec.mx

Abstract. Motion imagery is a concept that has been widely researched for its application in EEG-controlled Brain-Computer Interfaces (BCI). These interfaces offer several applications in rehabilitation. From diseases like amyotrophic lateral sclerosis (ALS) to brain trauma. Along different researchers, there are several techniques for selecting the electrode channels for the motion imagery BCI. However, little effort has been put into figuring out what is the better approach for motion imagery channel selection. This paper aims to evaluate the accuracy of principal component analysis (PCA) and sequential selection algorithms for two different deep learning architectures, multiple-layer perceptron neural network (MLPNN) and convolutional neural network (CNN). The evaluation is unique in its aim to evaluate also the channels most commonly opted for, and consider the possibility of a global channel set for different architectures.

Keywords: Channel Selection, EEG, Motor Imagery, Deep Learning, Brain-Computer Interface.

1 Introduction

Motion imagery (MI) is a mental process that consists of a subject imagining a movement without executing said movement [1, 2]. This mental process activates the primary motor cortex and the additional motor areas in the same manner as real movement, and it can be analyzed by electroencephalography (EEG) recordings [3, 4]. This type of EEG signal has been widely researched for its application in brain-computer interfaces (BCI). BCIs are systems that allow direct communication between a subject and their environment through the use of brain signals, which includes MI-based EEG signals [5]. BCIs typically consist of four phases which are signal acquisition, feature extraction, feature classification, and device control interface [6]. The recent development of deep learning (DL) has allowed an increase in performance in MI classification, as they hold the capacity of adapting non-linear and non-stationary signals and extracting feature information automatically [6, 7]. MLP are the most common feed-forward neural networks. This is because they are fast, easy to implement and do not require training sets that are too big [8]. The structure of this architecture consists of three types of layers: An input layer, an output layer, and multiple hidden layers [9]. The hidden layers receive data from the previous layer and conduct a nonlinear or linear activation function on the weighted sum of all inputs from the previous layer. CNN architectures have been widely used for MI classification purposes for their ability to learn features from a local receptive field, which allows them to classify complex EEG classification tasks [10]. One important aspect of task classification of MI is channel selection, that is, what number of electrode channels and which channels may yield better results from the classification. Regarding the number of channels, an increase in the number might result in higher accuracy, but the probability of electrode-related error increases [11]. This is why it is important to consider an approach to select optimal channels and channel sets based on the requirements of the BCI. Another aspect to take into consideration is electrode placement, as the different electrode areas are related to different functions [12, 13]. EEG experiments on MI classification are usually conducted using the 10-20 system [14]. There are different approaches when it comes to channel selection methods. These can be separated into filtering techniques, wrapper techniques, hybrid techniques, and embedded techniques [15]. Filtering techniques include correlation criteria [16], mutual information [17], chi-squared [18]. Wrapper techniques include sequential selection, examples being sequential feature selection (SFS) [19] and Plus-L-Minus-r search method [20], as well as heuristic search algorithms. This paper focuses on the evaluation of a sequential selection algorithm in comparison with a PCA approach. It aims to find commonalities in the optimal channel sets, for both channel selection methods and examine whether or not this can be significant enough to consider a global channel set for all subjects in order to avoid the use of channel selection methods and reduce the number of electrode channels used during experimentation. In addition, it aims to conduct comparisons in accuracy for sequential selection algorithm, and a PCA approach in a raw 64-channel set, for MI classification.

2 Methods

To identify the algorithm with the higher accuracy for MI classification, a bi-classification paradigm was chosen, and a public database was used to get training and testing sets. The algorithm was evaluated with two separate classifiers.

2.1 Referred dataset

The algorithms were evaluated on the public EEG dataset 'EEG Motor Movement/Imagery Dataset' from Schalk et al [21, 22]. It consists of a set of 64-channel

EEGs from 109 subjects recorded over 1500 recordings using the BCI2000 system, being presented in EDF+ format. The montage for the 64 electrodes followed the extended international 10-20 system (excluding Nz, F9, F10, Ft10, A1, A2, Tp9, Tp10, P9, P10).

The data was sampled at 160 samples per second and included an extra annotation channel for segmentation. The subjects were recorded performing 3 repetitions of 4 tasks of motion and MI in 14 experimental runs (including 2 baseline runs, no task). The tasks were indicated by the appearance of a target in a certain position on-screen:

1. Target appears on the left/right side of the screen, opening, and closing of the respective fist.
2. Target appears on the left/right side of the screen, with imagery of the opening and closing of the respective fist.
3. Target appears on the top/bottom of the screen, opening and closing of either both fists (top) or both feet (bottom).
4. Target appears on the top/bottom of the screen, imagery of opening and closing of either both fists (top) or both feet (bottom).

From the over 1500 recordings contained in the dataset, only 327 were implemented to reduce the problem to a bi-class evaluation. Task 2, which is the imaginary opening and closing of the left/right fist, was chosen for training and evaluating the classification models and channel selection algorithms.

2.2 Pre-processing

An analysis was conducted on the referred EEG database prior to the implementation of channel selection and classification. A frecuential and ERD/ERS analisys was made to verify that the MI de-sincronization was present in the database. The first part of Figure 1 and Figure 2 shows the pre-processing stage. First, a fourth-order bandpass butterworth filter was applied to the data. The data was filtered between 8 and 33 Hz, aiming to keep only β and μ bands, as they are most significant for MI classification [23].

Once the data is filtered, segmentation occurs extracting the data segments using information contained in the annotations channel. MI events are encoded in the annotation in the three following labels [21, 22]:

- T0, corresponds to rest
- T1, correspond to onset of motion in left fist
- T2, corresponds to onset of motion in right fist

Normalization of the events is performed by a min-max normalization, which is implemented to transform the data to fit the range $(0, 1)$ [24].

2.3 Sequential selection algorithm

The sequential channel selection method is based on the algorithm published by Mawata-Velu et al [13]. This algorithm operates on discriminating the active from inactive electrode channels for classification based on accuracy rating, as it can be seen in Figure 1 and Figure 2, for MLP and CNN architectures respectively. It first processes every single channel independently from one another through the classifier, generating a pool with the highest accuracy channels in the data-set. Next, it proceeds to conduct combinations with the obtained pool to generate a set of n electrodes, which runs through the classifier to obtain the highest accuracy set. Each channel set is used as a reference for the next one, which means combinations for a channel set of n channels will be made using the results from the previous iteration of $n-1$ channels. For the processing of each channel or channel sets, the splitting of data following an 80-20 rule for training-testing is done, and the data is sent to the classifier. The final shape of the input matrix given to the classifier for training/testing is of $(x, n, 113)$, where x is the number of events in the training/testing set, n is the number of channels, and 113 is the number of samples in the event. The first classifier

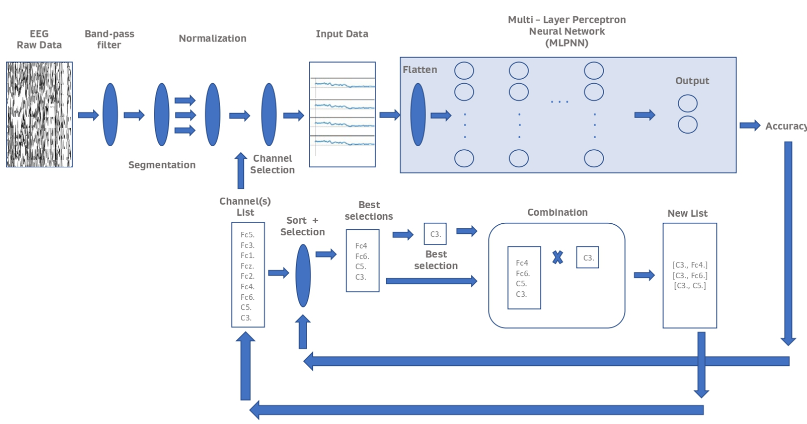


Fig. 1: Overview of the sequential selection algorithm on a MLP classifier.

architecture for the sequential selection algorithm is an MLP. The structure for this classifier can be observed in Table 1.

A CNN architecture is also considered for evaluating the algorithms. The structure used for the implementation of the CNN classifier can be observed in Table 2.

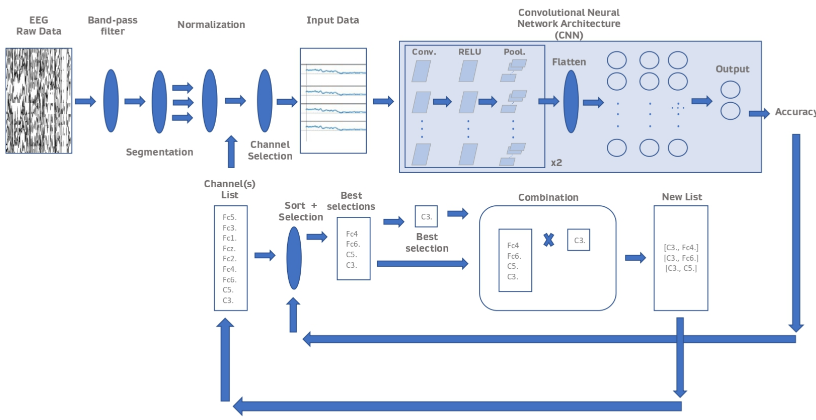


Fig. 2: Overview of the sequential selection algorithm on a CNN classifier.

Table 1: Summary and number of the parameters implemented in the MLP classification model.

Layer (type)	**Output Shape**	**Parameters**
Input Layer	(None, k, 113, 3)	0
Flatten	(None, 113)	0
Dense	(None, 25)	2850
Dense	(None, 25)	650
Dense	(None, 25)	650
Dense	(None, 2)	52
Softmax	(None, 2)	0

Table 2: Summary and number of the parameters implemented in the CNN classification model.

Layer (Type)	Output Shape	Parameters
Input Layer	(None, k, 113, 3)	0
Conv2D	(None, k, 113, 3)	6
Conv2D	(None, k, 113, 3)	12
Flatten	(None, 339)	0
Dense	(None, 25)	8500
Dense	(None, 25)	650
Dense	(None, 25)	650
Dense	(None, 2)	52
Softmax	(None, 2)	0

Both classifiers were trained using an Adam optimizer, using accuracy as the only metric, for 10 epochs for each channel or channel set.

2.4 PCA approach

PCA is a well-known technique for multivariate analysis used extensively in EEG and BCI applications. In neural computing applications, it aims to either reduce the data size in the form of feature selection, extracting the most significant components for classification, or using the significance acquired from said components to select the most significant channels. To accomplish this, the data for each subject X is arranged into a two-dimensional matrix $n \times p$ of n channels and p data samples, which for this case is 113. Once the data is arranged, the mean u_j of each row n is subtracted B as is shown in equations (1).

$$\begin{aligned} u_j = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} X_{ij}, \\ B = X - u^T. \end{aligned} \tag{1}$$

The next step to finding the PC of our data is to obtain the covariance of the matrix B, shown in equation (2).

$$C = \frac{1}{n-1} BB^T. \tag{2}$$

Where C is the covariance matrix, then the equation in (3) is used to get the PCA elements.

$$\begin{aligned} D = V^{-1}CV, \\ D_{kk} = d_k, \quad k = 1 \ldots n. \end{aligned} \tag{3}$$

Where V is the matrix of eigenvectors of C arranged on n columns, V^T is the matrix of principal components (PC) of the original matrix X, D is a diagonal matrix with the values d_k of the variance of X on the PC [25].

Once the PC matrix is obtained from the data, one can use this to obtain the most significant channels from the signal. Method B4 [26, 27] was opted for our PCA channel selection. It consists of selecting a number p which will be equal to the number of components accounting for a certain proportion, λ_0, of the total variance. One variable is related to each component, the variable which has the largest coefficient in the component. Then all variables p are selected, and variables $K - p$ are rejected. A slight modification to the method is done, where instead the number of channels for a variance threshold, the number of maximum channels to be selected will be set to $p = 6$, since it is the maximum channel set implemented to the sequential channel algorithm approach.[28] This method was programmed in python using the PCA function from the Scikit-learn library. The function uses the Probabilistic PCA approach.

3 Results

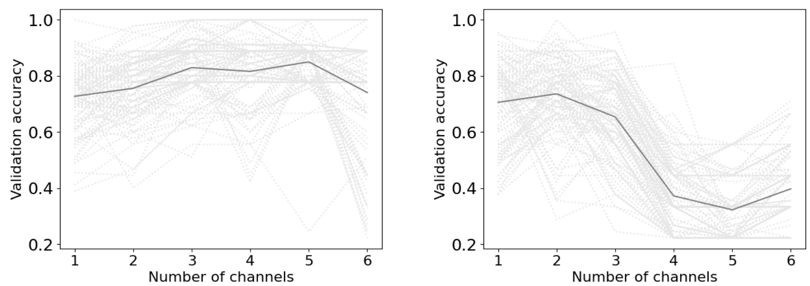
For the evaluation of channel methods, only accuracy was used as a evaluation metric, as it tends to be the most widely used metric in relation to MI classification. Moreover, as the intent of this evaluation is to find out which channels are preferred by the algorithms, channels are sorted by the instances in which they were selected, and the accuracy for each instance in which the channel appears is averaged.

3.1 Sequential selection evaluation

For the sequential selection algorithm, sorted lists of the top channels for each subject were obtained. Afterwards, the data was sorted first based on the instances in which the channel sets repeated, and second on the accuracy of the channel selection. The preferred channels by the algorithm for 1 channel can be seen in Table 3. In addition to this, the learning curves of MLP and CNN as a function of the number of channels used for the classifier can be seen in Figure 3. The preference of the channel selector can be seen spatially in Figure 4, where the color is dependant on the instances that each channel is selected.

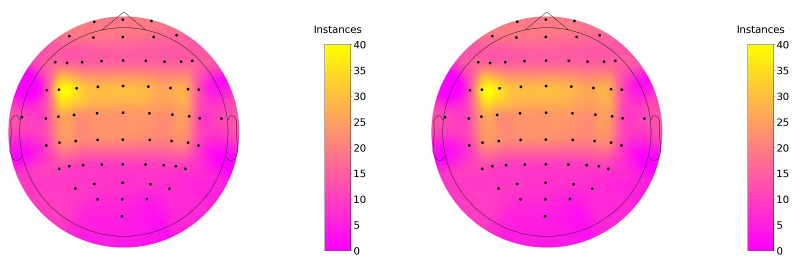
Table 3: Channels most commonly selected by the sequential selection algorithm for 1 channel.

MLP				CNN			
Channel	Instances	Percentage	Mean Accuracy	Channel	Instances	Percentage	Mean Accuracy
FC6	33	3.11%	70.37%	FC5	37	3.49%	77.78%
FC4	30	2.83%	79.63%	FCz	34	3.21%	79.37%
FCz	30	2.83%	76.39%	FC1	29	2.74%	83.33%
FC3	28	2.64%	75.00%	FC6	29	2.74%	66.67%
FC5	27	2.55%	73.33%	FC4	28	2.64%	68.25%
FT8	27	2.55%	68.52%	C3	26	2.45%	74.07%
FC1	26	2.45%	75.56%	C5	26	2.45%	71.43%
CP6	26	2.45%	73.61%	FC3	25	2.36%	74.07%
FC2	26	2.45%	69.84%	FC2	24	2.26%	73.02%
C5	25	2.36%	68.52%	C6	24	2.26%	71.11%



(a) MLP Classifier (b) CNN Classifier

Fig. 3: Learning curves of the classifiers through number of channels per subject. Line in bold reflects the mean of each number of channels. Individual results are the ones in grey.



(a) MLP Classifier (b) CNN Classifier

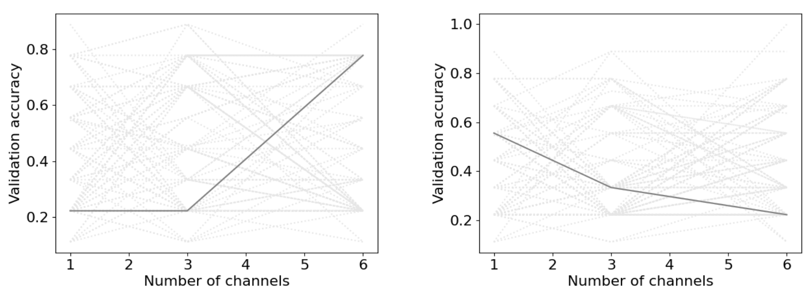
Fig. 4: Spatial maps of the number of instances in which channels appear as most significant in the dataset according to the sequential selection algorithm.

3.2 PCA evaluation

In the case of the PCA approach to channel selection, only one list of the 10 most significant channels was obtained for each subject. These channels were arranged into 1, 3, and 6 channel sets and the lists were sorted first based on the instances in which the channel sets repeated, and second on the accuracy of the channel selection. The preferred channels by the algorithm for 1 channel can be seen in Table 4. In addition to this, the learning curves of MLP and CNN as a function of the number of channels used for the classifier can be seen in Figure 5.

Table 4: Channels most commonly selected by the PCA algorithm for 1 channel.

MLP				CNN			
Channel	Instances	Percentage	Mean Accuracy	Channel	Instances	Percentage	Mean Accuracy
T9	28	2.64%	33.33%	T9	28	2.64%	22.22%
C6	11	1.04%	33.33%	C6	11	1.04%	77.78%
AF8	6	0.57%	55.56%	T8	6	0.57%	77.78%
T8	6	0.57%	55.56%	T10	6	0.57%	55.56%
T10	6	0.57%	22.22%	AF8	6	0.57%	22.22%
TP7	5	0.47%	66.67%	TP7	5	0.47%	44.44%
TP8	4	0.38%	44.44%	AF7	4	0.38%	66.67%
P7	4	0.38%	11.11%	F7	4	0.38%	44.44%
F7	4	0.38%	11.11%	TP8	4	0.38%	33.33%
AF7	4	0.38%	11.11%	P7	4	0.38%	22.22%



(a) MLP Classifier (b) CNN Classifier

Fig. 5: Learning curve of the classifiers through number of channels per subject. Line in bold reflects the mean of each number of channels.

4 Discussion and Conclusions

We can observe that at only one channel there seems to be a difference in accuracy between MLP and CNN classifiers for the sequential selection algorithm, where both accuracies are slightly above the 70% threshold (see Figure 3). For PCA, the difference is much more noticeable, with MLP having a close to 20% accuracy, while the one in the CNN classifier is close to 60%. As the channel selection algorithms begin to generate channel sets with a higher number of channels, there are notable differences between the classifiers. In PCA, MLP increases significantly, reaching almost 80% accuracy on average at 6 channels, as can be seen in Figure 5. However, in the sequential selection, although there is also an increase in accuracy, it seems to be of a lower slope, even decreasing after finding an above 80% peak in the 5-channel set. In the case of the CNN classifier, both for PCA and sequential selection algorithms, accuracy decreases as the number of channels increase. In the case of sequential selection, it seems to slightly increase from 1 to 2 channels, only to then decrease until the lowest point at 5 channel set, around 35%, only to slightly increase at the 6-channel set. Comparing PCA to the sequential selection algorithm, a clear advantage for the algorithm over sequential selection is that it requires only one iteration per subject to compute any number of channels, while sequential selection requires the same number of iterations as the number of channels in a channel set. PCA shows low accuracy for one channel set, though it seems to increase with a number of channels, as the two most preferred channels have accuracies of 33.33% (Table 4), and a mean of slightly more than 20% (Figure 5). On the other hand, sequential selection holds accuracies of 70.37% and 79.63% for the two most preferred channels (Table 3), and a mean of more than 70% (Figure 3). Overall, as the number of channels per set increases, sequential selection seems to have higher accuracy than PCA (Figure 5 and Figure 3). However, the sequential selection CNN model implented in our work did not reach the same accuracies as other works, which surpassed the 90% barrier [13].

Analyzing the sequential selection algorithm preferred channels, the ones that have better results in accuracy correspond with the literature data [12], for the frontal-central (FC), and frontal-temporal (FT) channels, as well as central-parietal (CP) and central (C) channels, although they vary between MLP and CNN classifiers (see Figure 3). These are significant since they spatially cover the precentral gyrus of the brain, which is the area responsible for the generation of neural impulses for muscle control. The 10 most preferred channels hold together 26.22% of all channel selections for MLP, and 26.6% for CNN, across the 106 subjects. As seen in the spatial maps of the number of channel selections (Figure 4), the sequential selection algorithm tends to indicate that channels located in the most anterior portion of the precentral gyrus are preferred for classification in both MLP and CNN classifiers, with a tendency towards the left channels. However, the PCA algorithm yields a different set of channels as most preferred across the data set. Table 4, shows a preference of temporal (T9, T8, T10), temporal-parietal (TP7, TP8), and frontal (AF8, AF7) channels, although they also present frontal (F7), central (C6), and parietal (P7) channels. These related to auditory perception and high-level visual processing, not necessarily movement-related tasks. The discrepancy is related to the method of discrimination in the selection methods, as sequential selection chooses channels that yield the most accurate, while PCA finds the channels which explain the highest variance in the signal. The 10 most preferred channels hold 7.38% of all channel selections across the 106 subjects for both classifiers. Due to the discrepancy between both approaches, having a mostly frontal-central preference for sequential selection and temporal preference for PCA, it is not possible to state that a global, optimal channel set for MI classification exists. However, based on the information gathered, the frontal-central channels shown by the sequential selection method seem to be the most accurate ones, as well as the most selected group, holding a 26.22%

This research aimed to find a global optimal channel set to simplify the process of channel selection. It evaluated the performances of PCA and sequential selection methods by comparing their accuracies, as well as finding recurrent channel choices in both methods. These channel selection approaches were evaluated using a 106-subject database, with a left-right fist opening-closing paradigm. MLP and CNN architectures were implemented as the classifiers, and the resulting channel selections were sorted by the instances in which they were repeated across subjects and the accuracies they yielded. The results present different preferences in channels between channel selection methods, as PCA focused on channels in the temporal area while sequential selection opted for channels in the frontal-central area. Overall, sequential selection yielded a better mean accuracy in both classifiers and through different channel sets, with the exception of the 6-channel set for the MLP classifier. Based on the previously mentioned it cannot be concluded that there is a global channel set between channel selection methods, the best channels to be worked based on accuracy rely on the channel selection algorithm.

References

1. C. Zhang, Y.-K. Kim, and A. Eskandarian, "EEG-inception: An accurate and robust end-to-end neural network for EEG-based motor imagery classification" Journal of Neural Engineering, vol. 18, no. 4, p. 046014, 2021.
2. M. Jeannerod, "Mental imagery in the motor context" Neuropsychologia, vol. 33, no. 11, pp. 1419–1432, 1995.
3. C. Lindig-León, S. Rimbert, and L. Bougrain, "Multiclass classification based on combined motor imageries" Frontiers in Neuroscience, vol. 14, 2020.
4. S. Hétu, M. Grégoire, A. Saimpont, M.-P. Coll, F. Eugène, P.-E. Michon, and P. L. Jackson, "The neural network of motor imagery: An ale meta-analysis" Neuroscience & Biobehavioral Reviews, vol. 37, no. 5, pp. 930–949, 2013.
5. O.-Y. Kwon, M.-H. Lee, C. Guan, and S.-W. Lee, "Subject-independent brain–computer interfaces based on deep convolutional neural networks" IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 10, pp. 3839–3852, 2020.
6. A. M. Roy, "A CNN model with feature integration for MI EEG subject classification in BMI" 2022.
7. Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals" Journal of Neural Engineering, vol. 14, no. 1, p. 016003, 2016.
8. U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model" Expert Systems with Applications, vol. 38, no. 10, pp. 13475–13481, 2011.
9. Y. Lee and I. Sohn, "Reconstructing damaged complex networks based on Neural Networks" Symmetry, vol. 9, no. 12, p. 310, 2017
10. X. Lun, Z. Yu, T. Chen, F. Wang, and Y. Hou, "A simplified CNN classification method for Mi-EEG via the electrode pairs signals" Frontiers in Human Neuroscience, vol. 14, 2020.
11. R. Scherer, G. R. Müller-Putz, and G. Pfurtscheller, "Chapter 9 flexibility and practicality" International Review of Neurobiology, pp. 119–131, 2009.

12. S. Syakiylla, S. Syakiylla Sayed Daud, and R. Sudirman, "Decomposition level comparison of stationary wavelet transform filter for visual task electroencephalogram," Jurnal Teknologi, vol. 74, no. 6, 2015.
13. T. Mwata-Velu, J. G. Avina-Cervantes, J. Ruiz-Pinales, T. A. Garcia-Calva, E.-A. González-Barbosa, J. B. Hurtado-Ramos, and J.-J. González-Barbosa, "Improving motor imagery EEG classification based on channel selection using a deep learning architecture" Mathematics, vol. 10, no. 13, p. 2302, 2022.
14. R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of International 10–20 System Electrode Placement," Electroencephalography and Clinical Neurophysiology, vol. 66, no. 4, pp. 376–382, 1987.
15. M. Z. Baig, N. Aslam, and H. P. Shum, "Filtering techniques for channel selection in motor imagery EEG applications: A survey" Artificial Intelligence Review, vol. 53, no. 2, pp. 1207–1232, 2019.
16. I. Guyon and A. Elisseeff, "An introduction to feature extraction" Journal of machine learning research, pp. 1157–1182, Mar. 2003.
17. C. Lazar, J. Taminau, S. Meganck, D. Steenhoff, A. Coletta, C. Molter, V. de Schaetzen, R. Duque, H. Bersini, and A. Nowe, "A survey on filter techniques for feature selection in gene expression microarray analysis" IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 9, no. 4, pp. 1106–1119, 2012.
18. P. E. Greenwood and M. S. Nikulin, A guide to chi-squared testing, vol. 280. New York, New York: Wiley, 1996.
19. P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection" Pattern Recognition Letters, vol. 15, no. 11, pp. 1119–1125, 1994.
20. S. Nakariyakul and D. P. Casasent, "An improvement on floating search algorithms for feature subset selection" Pattern Recognition, vol. 42, no. 9, pp. 1932–1940, 2009.
21. G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system" IEEE Transactions on Biomedical Engineering, vol. 51, no. 6, pp. 1034–1043, 2004.
22. A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, PhysioToolkit, and PhysioNet" Circulation, vol. 101, no. 23, 2000.
23. S. R. Sreeja, J. Rabha, K. Y. Nagarjuna, D. Samanta, P. Mitra, and M. Sarma, "Motor imagery EEG signal processing and classification using Machine Learning Approach" 2017 International Conference on New Trends in Computing Sciences (ICTCS), 2017.
24. H. Shan, H. Xu, S. Zhu, and B. He, "A novel channel selection method for optimal classification in different motor imagery BCI paradigms" BioMedical Engineering OnLine, vol. 14, no. 1, 2015.
25. Kok-Meng Ong, Thung, K.-H., Chong-Yaw Wee, & Paramesran, R. (2005). Selection of a Subset of EEG Channels using PCA to classify Alcoholics and Non-alcoholics. 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. doi:10.1109/iembs.2005.1615389
26. Jolliffe, I. T. (1972). Discarding Variables in a Principal Component Analysis. I: Artificial Data. Applied Statistics, 21(2), 160. doi:10.2307/2346488
27. Jolliffe, I. T. (1973). Discarding Variables in a Principal Component Analysis. II: Real Data. Applied Statistics, 22(1), 21. doi:10.2307/2346300
28. M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers" Neural Computation, vol. 11, no. 2, pp. 443–482, 1999.