Optimization of Irrelevant Features for Brain-Computer Interface (BCI) System

Ong Zhi Ying\textsuperscript{1*}, Saidatul Ardeenawatie binti Awang\textsuperscript{1} and Vikneswaran A/L Vijean\textsuperscript{1}

\textsuperscript{1} Biosignal Processing Research Group (BioSIM), School of Mechatronic Engineering, Universiti Malaysia Perlis, Arau, Perlis, Malaysia

* jean-624@hotmail.com / ongzhiying624@gmail.com

Abstract. The brain is the most important body part for human. The brain controls all activities of the body such as movement, imagine, and response. Therefore, it is believed that the signals which collected from human scalp contain a lot of useful information. This useful information known as features can be extracted by applying advanced signal processing. Then, the features used for a brain-computer interface (BCI) system. However, the most suitable and relevant features for the BCI system still not investigate. In this paper, ten healthy subjects were involved in data collection. Threshold method, notch filter and wavelet decomposition were applied during pre-processing. Then, the signals were normalised. Hilbert-Huang Transform (HHT) and Power Spectral Density (PSD) were implemented. The features such as statistical-based features, approximate entropy (ApEn), sample entropy (SampEn), fuzzy entropy (FuzEn), permutation entropy (PermEn), distribution entropy (DistEn), Hjorth parameter, and Hurst exponent (HE) were extracted from PSD and HHT separately. Genetic algorithm (GA) and reliefF were carried out to select the most suitable and relevant features for the BCI system. The prediction rate before and after feature selection were compared. The performance after feature selection is improved in term of prediction rate and training time. The best classifier, in this case, is the bagged tree which can achieve 99.30%.

1. Introduction

The technology that allows people to control the external device without using any muscles and nerves called a brain-computer interface (BCI). It provides a novel communication pathway between the brain and the computer. BCI helps disabled people to restore lost motor functions [1]. It is a non-invasive system that has no risk of surgery [2].

The signals collected from human scalps are known as an electroencephalogram (EEG). The EEG signals were used for BCI system. Features that extracted from EEG signals are the main key for a BCI system. It influences the performance of a BCI system. There are many methods used to extract features in the frequency domain [3], [4] or time-frequency domain [4], [5], [6]. However, not all features are useful for BCI system. Different features are required by different cases of neurology. Irrelevant features will make the system confuse and the performance of the system is being influenced [7].
In this paper, the classification results for features before and after applying feature selection were compared. This is to find out which features are the most suitable and relevant for a BCI system. The features were extracted from power spectral density (PSD) and Hilbert-Huang Transform (HHT) which represented frequency domain and time-frequency domain respectively.

2. Data Acquisition
Ten Universiti Malaysia Perlis (UniMAP) students were involved in this study. They are healthy and right-handed. Their ages are in the range of 20 to 30 years old. They are in good mental condition with no neurological disorder.

The EEG signals were collected from human scalps by using EEGO™ sports device (ANT Neuro, Enschede, Netherlands) with 32 channels. The sampling frequency was set to 512 Hz. Nyquist theorem stated that the sampling rate should be at least 2 times greater than the original rate [8]. The low cut-off frequency is 1.0 Hz and the high cut-off frequency is 70.0 Hz. The impedance kept below 5 kΩ by inserting the gel into all channels.

During the experiment, the subjects were required to perform four different tasks. They had to close eyes, relax and avoid body movement throughout the experiment. For relaxation task, they were asked to sit without doing anything for 1 minute. Then they were asked to imagine opening and clenching their fist for 3 minutes. The first minute is left hand, the second minute is right hand and last minute is both hands. There are many other imaginary tasks such as mental math tasks and visual evoked potential tasks. However, the motor imagery task was used for BCI system due to the prosthetics arm is used in daily life for a disabled individual. It will be accidentally activated the prosthetics arm when the users calculating some math questions or seeing the visual stimuli in their daily lives.

3. EEG Signal Processing
In this paper, pre-processing, feature extraction, feature selection and classification were carried out to analyse the EEG signals.

3.1. Pre-processing
Firstly, the hard threshold method was used to remove eye blinks artefact [9]. The EEGO™ sports device is wireless, but the notch filter still applied to minimize the possibility of power line interference which is at 50 Hz [10], [11], [12], [13]. Then, Daubechies 8 wavelet family with 5th level decomposition was used to de-noise the baseline wander noise [14], [15]. After removing unwanted noise, the filtered signals were normalized and segmented into 50 segments in windowing size of 1 second each segment.

3.2. Feature Extraction
The filtered signals were undergone feature extraction techniques which are power spectral density (PSD) and Hilbert-Huang Transform (HHT). Then, 13 features were extracted from PSD and HHT separately. 13 features are minimum (Min), maximum (Max), mean, variance (Var), standard deviation (Std), approximate entropy (ApEn), sample entropy (SampEn), fuzzy entropy (FuzEn), permutation entropy (PermEn), distribution entropy (DistEn), Hjorth parameter, and Hurst exponent (HE). In the end, there are a total of 91 features being extracted.

3.2.1. Power Spectral Density (PSD). Power spectral density (PSD) describes the distribution of power into frequency components composing the signals [16], [17]. In this paper, the Welch method was implemented. In the Welch method, in order to determine the estimate of the PSD, the periodograms averaged over time. The periodograms are formed from the successive blocks. A data
window is applied to each segment of the time sequence. It divided the time sequence into successive blocks [18].

3.2.2. Hilbert-Huang Transform (HHT). Hilbert-Huang transforms (HHT) also known as a Hilbert spectrum (HS). It is a combination of Empirical Mode Decomposition (EMD) and Hilbert transform. EMD decomposed the signals into Intrinsic Mode Functions (IMF). Then, IMF components such as instantaneous frequency and amplitude were calculated by using Hilbert transform [18], [19], [20], [21]. In this study, the first three IMF components were used to extract the features.

3.2.3. Statistical-based Features. Minimum (Min) is the features with the smallest values. Maximum (Max) is the features with the largest values. Mean is the features with the average values. Variance (Var) is a quantity equal to the square of the standard deviation. Standard deviation (Std) is a quantity expressing by how much the variables of a group differ from the mean value for the group.

3.2.4. Entropy-based Features. Approximate entropy (ApEn) is a complexity or irregularity data. It is a measure of regularity of data. A higher non-negative value for ApEn obtained from an irregular time series while a lower ApEn value obtained from a regular and predictable time-series signal [18]. Equation (1) shows the calculation of approximate entropy.

\[
ApEn(m, r, N) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \log \frac{C_i^m(r)}{C_i^{m+1}(r)} - \frac{1}{N - m} \sum_{i=1}^{N-m} \log \frac{C_i^{m+1}(r)}{C_i^{m+1}(r)}
\]

(1)

\[
C_i^m(r) = \frac{2}{N_m(N_m-1)} \sum_{j=1, j \neq i}^{N_m} \Theta(r - ||x_i - x_j||)
\]

(2)

where \( m \) is embedding dimension, \( r \) is tolerance window and \( N \) is length.

Sample entropy (SampEn) is a measure of self-similarity. It also measures the complexity and regularity of the time-series data. Higher values registered for more irregular data while lower value implies high self-similarity [18]. The formula of sample entropy showed in equation (3) [22].

\[
SampEn(m, r, N) = \begin{cases} \rightarrow \infty, & \text{when } A = 0 \\ \ln \frac{B}{A}, & \text{otherwise.} \end{cases}
\]

(3)

where \( m \) is the length, \( r \) is distance, \( A \) and \( B \) are the measures of similarity between templates of length \( m \) and \( m + 1 \) respectively.

Fuzzy entropy (FuzEn) is used to express the mathematical values of the fuzziness of fuzzy sets [23]. The significance of the different characteristics of signals was evaluated by fuzzy entropy [24]. The formula of fuzzy entropy is given by equation (4).

\[
H(f, c_i) = - P(f, c_i) \log P(f, c_i)
\]

(4)

where \( f \) is fuzzy set, \( c_i \) is class \( i \) and \( P(f, c_i) \) can be interpreted as the degree to which the sample is pre-defined to belong to class \( i \) really contributes to that specific class.

Permutation entropy (PermEn) provides a quantification measure of the complexity of how the series behave according to the sequence of ordinal patterns [25]. It captures the order relations between values of a time series and extracting a probability distribution of the ordinal patterns. The formula of permutation entropy is given by equation (5).
\[ H(C^X(d)_k) = -\sum_{(\pi_l)_k \in \Omega_d} P(\pi_l)_k \ln P(\pi_l)_k \]  

(5)

\[ P(\pi_l)_k = P(\{\omega \in \Omega | X(T^{\omega l+d-1}(\omega)), ..., X(T^{\omega l}(\omega)), X(T^{\omega l-1}(\omega)) \} \text{ has ordinal pattern } \pi_l \text{ for } l = 1, 2, ..., k) \]  

(6)

where \( T \) is a piecewise monotone interval map, \( \Omega \) is an interval and \( X \) is identity on \( \Omega \).

The calculation of Distribution entropy (DistEn) based on the empirical probability distribution function of distances among vectors formed from a given time series. It is an irregularity measure [26]. Equation (7) denotes distribution entropy.

\[ DistEn(m, M) = \frac{-1}{\log_2(M)} \sum_{t=1}^{M} p_t \log_2(p_t) \]  

(7)

where \( m \) is embedding dimension, \( M \) is the number of bins used and \( p_t \) is the probability of each bin.

3.2.5. Hjorth Parameter. Hjorth parameter has three kinds of parameters such as activity, mobility, and complexity. In this paper, only mobility and complexity was used because the activity parameter is the variance of the time function. Mobility parameter is defined as the square root of the ratio of the variance of the first derivative of the signal and the variance of the signal. Complexity parameter indicates how the shape of the signal is similar to a pure sine wave. When the shape of the signal gets more similar to a pure sine wave, the complexity value converges to 1 [27]. Equation (8) and equation (9) show the calculation of mobility and complexity respectively.

\[ Mobility = \sqrt{\frac{\text{var}(\frac{dy(t)}{dt})}{\text{var}(y(t))}} \]  

(8)

\[ Complexity = \frac{Mobility(\frac{dy(t)}{dt})}{Mobility(y(t))} \]  

(9)

where \( \frac{dy(t)}{dt} \) is the first derivative of the signals and \( y(t) \) is the signals.

3.2.6. Hurst Exponent (HE). Hurst exponent (HE) is a measure of smoothness of a fractal time-series based on asymptotic behaviour. It is also a measure of self-similarity, predictability and long-range correlation of a time-series [18]. Equation (10) denotes the Hurst exponent, \( H \).

\[ H = \frac{\log(R)}{\log(T)} \]  

(10)

where \( T \) is the sample duration, \( \frac{R}{S} \) is the corresponding value of the rescaled range, \( R \) is the difference in maximum and minimum of the deviation from mean and \( S \) is the standard deviation.

3.3. Feature Selection
Genetic algorithm (GA) and reliefF were used for feature selection. Only the features that had been selected by both GA and reliefF were used for the following studies. This shows better validation for the selected relevant features.

3.3.1. Genetic Algorithm (GA). Genetic algorithm (GA) is a wrapper-based feature selection [28]. It finds the minimum of the function. It is a heuristic searching algorithm. It also used to find the optimum combination of feature extraction methods and the classifiers [29]. In this study, k-nearest neighbours (kNN) used for the fitness function.

3.3.2. ReliefF. ReliefF is a filter-based feature selection method. It ranked the features by calculating the feature weights [7]. ReliefF is a simple and effective approach to feature weight estimation. The feature weight is defined in terms of feature relevance [7]. In this study, all positive weights were selected.

Equation (9) is the formula of weights if \(x_r\) and \(x_q\) are in the same class and equation (10) is the formula of weights if \(x_r\) and \(x_q\) are in a different class.

\[
W^i_j = W^{i-1}_j - \frac{\Delta_j(x_r, x_q)}{m}.d_{rq} \quad (9)
\]

\[
W^i_j = W^{i-1}_j + \frac{(P_{y_q}/1-P_{y_r}).\Delta_j(x_r, x_q)}{m}.d_{rq} \quad (10)
\]

where \(x_r\) is a random observation that selected by the algorithm iteratively, finds the k-nearest observations to \(x_r\) for each class, and updates, for each nearest neighbour \(x_q\), \(m\) is the number of iterations specified by ‘updates’, \(\Delta_j(x_r,x_q)\) is the difference in the value of the predictor \(F_j\) between observations \(x_r\) and \(x_q\). \(W^i_j\) is the weight of the predictor \(F_j\) at the \(i\)th iteration step, \(P_{y_r}\) is the prior probability of the class to which \(x_r\) belongs, and \(P_{y_q}\) is the prior probability of the class to which \(x_q\) belongs.

3.4. Classification

Several classifiers had been used for this study. There are k-nearest neighbours (kNN), decision tree and ensemble classifiers. Table 1 shows the parameters of classifiers.

| Model type                      | Fine KNN | Medium KNN | Weighted KNN | Fine Tree | Bagged Tree |
|---------------------------------|----------|------------|--------------|-----------|-------------|
| Number of neighbors            | 1        | 10         | 10           | -         | -           |
| Distance metric                | Euclidean| Euclidean  | Euclidean    | -         | -           |
| Distance weight                | Equal    | Equal      | Squared inverse| -        | -           |
| Standardize data               | True     | True       | True         | -         | -           |
| Maximum number of splits       | -        | -          | -            | 100       | -           |
| Split criterion                | -        | -          | -            | Gini’s diversity index | - |
| Surrogate decision splits      | -        | -          | -            | off       | -           |
| Ensemble method                | -        | -          | -            | - Bag     | -           |
| Learner type                   | -        | -          | -            | Decision tree | - |
| Number of learners             | -        | -          | -            | - 30      | -           |

4. Results and Discussions

Four different tasks represented four different classes. Before feature selection, there are 91 features with 32 channels that trained for classifiers. Table 2 shows the initial classification results before applying feature selection.
Table 2. Initial classification results before feature selection.

| Model type | Fine KNN | Medium KNN | Weighted KNN | Fine Tree | Bagged Tree |
|------------|----------|------------|--------------|-----------|-------------|
| Accuracy (%) | 71.60 | 74.90 | 76.30 | 90.00 | 99.00 |
| Training time (sec) | 1767.3 | 1687.7 | 1866.2 | 1297.5 | 1867.0 |

From the table, the accuracies for all classifiers are above 70%. The best classifier in term of accuracy is the bagged tree which can achieve 99%. However, it took a long time for training the data into all classifiers. The fastest classifier before feature selection is fine tree.

After feature selection, only 34 out of 91 features had been selected. Table 3 shows the features that had been selected by genetic algorithm and reliefF.

Table 3. Features that had been selected.

| No. | Features                  | No. | Features                  | No. | Features                  |
|-----|---------------------------|-----|---------------------------|-----|---------------------------|
| 1.  | PSD – Max                 | 13. | IMF 1 Amplitude – Mean    | 25. | IMF 2 Amplitude – PermEn  |
| 2.  | PSD – Var                 | 14. | IMF 1 Amplitude – Std     | 26. | IMF 3 Frequency – ApEn    |
| 3.  | PSD – Std                 | 15. | IMF 1 Amplitude – FuzEn   | 27. | IMF 3 Frequency – PermEn  |
| 4.  | PSD – FuzEn               | 16. | IMF 1 Amplitude – PermEn  | 28. | IMF 3 Amplitude – Min     |
| 5.  | PSD – PermEn              | 17. | IMF 1 Amplitude – Mobility| 29. | IMF 3 Amplitude – Max     |
| 6.  | PSD – DistEn              | 18. | IMF 2 Frequency – ApEn    | 30. | IMF 3 Amplitude – Mean    |
| 7.  | PSD – Mobility            | 19. | IMF 2 Frequency – PermEn  | 31. | IMF 3 Amplitude – Std     |
| 8.  | PSD – Complexity          | 20. | IMF 2 Amplitude – Min     | 32. | IMF 3 Amplitude – FuzEn   |
| 9.  | PSD – Hurst Exponent      | 21. | IMF 2 Amplitude – Max     | 33. | IMF 3 Amplitude – PermEn  |
|10.  | IMF 1 Frequency – ApEn    | 22. | IMF 2 Amplitude – Mean    | 34. | IMF 3 Amplitude – Mobility|
|11.  | IMF 1 Frequency – PermEn  | 23. | IMF 2 Amplitude – Std     |     |                           |
|12.  | IMF 1 Amplitude – Max     | 24. | IMF 2 Amplitude – FuzEn   |     |                           |

After feature selection, only 34 features with 32 channels trained for classifiers. From table 3, it can be concluded that most of the features extracted from PSD and instantaneous amplitude were selected. Sample Entropy was not a relevant feature in this case. Table 4 shows the classification results after applying feature selection.

Table 4. Classification results after feature selection.

| Model type | Fine KNN | Medium KNN | Weighted KNN | Fine Tree | Bagged Tree |
|------------|----------|------------|--------------|-----------|-------------|
| Accuracy (%) | 82.30 | 84.00 | 84.90 | 90.90 | 99.30 |
| Training time (sec) | 149.82 | 78.676 | 363.71 | 575.09 | 538.92 |

After feature selection, the performance of the classification improved in term of prediction rate and training time. For kNN, the prediction rate had been improved by around 11% to 15%. The prediction rates for fine tree and the bagged tree just improved by 1% and 0.3% respectively. The best classifier in term of prediction rate is bagged tree. The training time had been shortening around 81% to 95% for kNN classifiers. The training time for fine tree and the bagged tree had shortened 55.68% and 71.13% respectively. The fastest classifier after feature selection is medium kNN. The performances of kNN classifiers improved more than the fine tree and bagged tree. It may due to the fitness function of the genetic algorithm is k-nearest neighbours (kNN) and reliefF finds the k-nearest observations.
Feature selection helps classifiers to distinguish different classes. Other researchers had proven genetic algorithm and reliefF feature selections selected the relevant and suitable features that can improve the performances of classification in many different fields such as image processing [28], [7], [30], emotional stress test detection [31] and emotion recognition [32]. Similarly, this paper had proved the combination of feature selection methods performed well in the field of BCI system.

5. Conclusion
In conclusion, irrelevant features must be eliminated for the BCI system to enhance the performance. The prediction rate and training time improved in an effective way after applying feature selection. This shows that the irrelevant features confuse the classifier and make a false conclusion. For further work, feature level fusion technique and channel selection should be implemented to further reduce the data dimensionality. In addition, the number of used channels should be reduced for user-friendliness.

References
[1] J. Jiang, Z. Zhou, E. Yin, Y. Yu, Y. Liu, and D. Hu 2015 “A novel Morse code-inspired method for multiclass motor imagery brain – computer interface (BCI) design,” Comput. Biol. Med., 66, 11–19.
[2] H. A. Shedeed and M. F. Issa 2016 “Brain-EEG Signal Classification Based on Data Normalization for Controlling a Robotic Arm,” Int. J. Tomogr. Simul., 29(1), 72–85.
[3] B. Orset, K. Lee, R. Chavarriaga, and J. d. R. Millan 2019 “Reliable decoding of motor state transitions during imagined movement,” in 9th International IEEE EMBS Conference on Neural Engineering, 263–266.
[4] W. Yi et al. 2016 “EEG oscillatory patterns and classification of sequential compound limb motor imagery,” J. Neuroeng. Rehabil., 13(11), 1–12.
[5] J. Zhou, M. Meng, Y. Gao, Y. Ma, and Q. Zhang 2018 “Classification of motor imagery EEG using wavelet envelope analysis and LSTM networks,” in Proceedings of the 30th Chinese Control and Decision Conference, CCDC, 5600–5605.
[6] C. Tan, F. Sun, W. Zhang, S. Liu, and C. Liu 2017 “Spatial and spectral features fusion for EEG classification during motor imagery in BCI” in 2017 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2017, 309–312.
[7] R. P. L. DURGABAI and R. B. Y 2014 “Feature Selection using ReliefF Algorithm,” Ijarcc, 3(10), 8215–8218.
[8] C. S. Wang 2012 “Design of a 32-channel EEG system for brain control interface applications,” J. Biomed. Biotechnol., 2012, 1–10.
[9] P. P. Ngoc, V. D. Hai, N. C. Bach, and P. Van Binh 2015 “EEG Signal Analysis and Artifact Removal by Wavelet Transform,” in 5th International Conference on Biomedical Engineering in Vietnam, 2015, 46, 179–183.
[10] M. Teplan 2002 “Fundamentals of EEG measurement,” Meas. Sci. Rev., 2(2), 1–11.
[11] K. S. Bayram, M. K. Ayyüce, and B. Bolat 2013 “Classification of EEG Signals by using Support Vector Machines,” in 2013 IEEE Int. Symp. Innov. Intell. Syst. Appl. (INISTA), 1–3.
[12] S. Vaid, P. Singh, and C. Kaur 2015, “EEG SIGNAL ANALYSIS FOR BCI INTERFACE : A REVIEW,” in 2015 Fifth International Conference on Advanced Computing & Communication Technologies, 143–147.
[13] M. Ferdjallah and R. E. Barr 1994 “Adaptive Digital Notch Filter Design on the Unit Circle for the Removal of Powerline Noise from Biomedical Signals,” Eng. Educ., 41(6), 529–536.
[14] K. Asaduzzaman, M. B. I. Reaz, K. S. Sim, and M. S. Hussain 2010 “A Study on Discrete Wavelet-Based Noise Removal from EEG Signals,” in Advances in Computational Biology, 680(2010), 593–599.
[15] H. A. Akkar and F. Ali Jasim 2017 “Optimal Mother Wavelet Function for EEG Signal Analyze Based on Packet Wavelet Transform,” Int. J. Sci. Eng. Res., 8(2), 1222–1227.
[16] O. Dressler, G. Schneider, G. Stockmanns, and E. F. Kochs 2004 “Awareness and the EEG power spectrum: analysis of frequencies,” *Br. J. Anaesth.*, 93(6), 806–809, 2004.

[17] B. Rajak, M. Gupta, D. Bhatia, and A. Mukherjee 2016 “Power Spectral Analysis of EEG as a Potential Marker in the Diagnosis of Spastic Cerebral Palsy Cases,” *Int. J. Biomed. Eng. Sci.*, 3(3), 23–29.

[18] U. R. Acharya, S. V. Sree, G. Swapna, R. Joy, and J. S. Suri 2013 “Automated EEG analysis of epilepsy: A review,” *Knowledge-Based Syst.*, 45, 147–165.

[19] N. E. Huang et al. 1998 “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proc. Roy. Soc. Lond. A*, 454, 903–995.

[20] M. K. I. Molla and K. Hirose 2010 “Hilbert Spectrum in Time-Frequency Representation of Audio Signals Considering Disjoint Orthogonality,” *Adv. Adapt. Data Anal.*, 2(3), 313–336.

[21] M. K. I. Molla, M. A. M. Shaikh, and K. Hirose 2008 “Time-frequency representation of audio signals using Hilbert spectrum with effective frequency scaling,” in *Proceedings of 11th International Conference on Computer and Information Technology, ICCIT 2008*, 2008, 335–340.

[22] G. Manis, M. Akhtaruzzaman, and R. Sassi 2018 “Low computational cost for sample entropy,” *Entropy*, 20(61), 1–15.

[23] Y. YENİYAYLA 2011 “Fuzzy entropy and its applications”.

[24] B. Medina Salgado and L. Duque Muñoz 2015 “Fuzzy entropy relevance analysis in DWT and EMD for BCI motor imagery applications,” *Ingeniería*, 20(1), 9–19.

[25] K. Keller, T. Mangold, I. Stolz, and J. Werner 2017 “Permutation Entropy : New Ideas and Challenges,” *Entropy*, 19(134), 1–16.

[26] C. Karmakar, R. K. Udhayakumar, and M. Palaniswami 2015 “Distribution Entropy ( DistEn ): A Complexity Measure to Detect Arrhythmia from Short Length RR Interval Time Series,” in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015, 61, 5207–5210.

[27] S. Oh, Y. Lee, and H. Kim 2014 “A Novel EEG Feature Extraction Method Using Hjorth Parameter,” *Int. J. Electron. Electr. Eng.*, 2(2), 106–110.

[28] L. Zhuo, J. Zheng, X. Li, F. Wang, B. Ai, and J. Qian 2008 “A genetic algorithm based wrapper feature selection method for classification of hyperspectral images using support vector machine,” *Geoinformatics 2008 Jt. Conf. GIS Built Environ. Classif. Remote Sens. Images*, 7147, 397–402.

[29] S. V. Eslahi, N. J. Dabanloo, and K. Maghooli 2019 “A GA-based feature selection of the EEG signals by classification evaluation: Application in BCI systems,” *Comput. Sci. - Neural Evol. Comput.*, 1–12.

[30] Z. Wang et al. 2016 “Application of ReliefF algorithm to selecting feature sets for classification of high resolution remote sensing image,” in *International Geoscience and Remote Sensing Symposium (IGARSS)*, 755–758.

[31] D. Shon, K. Im, J.-H. Park, D.-S. Lim, B. Jang, and J.-M. Kim 2018 “Emotional Stress State Detection Using Genetic Algorithm-Based Feature Selection on EEG Signals,” *Int. J. Environ. Res. Public Health*, 15(2461), 1–11.

[32] J. Zhang, M. Chen, S. Zhao, S. Hu, Z. Shi, and Y. Cao 2016 “ReliefF-based EEG sensor selection methods for emotion recognition,” *Sensors (Switzerland)*, 16(1558), 1–15.