Distinguishing Two Different Mental States of Human Thought Using Soft Computing Approaches

Akshansh Gupta, Dhirendra Kumar, Anirban Chakraborti and Vinod Kumar Singh

Abstract Electroencephalograph (EEG) is useful modality nowadays which is utilized to capture cognitive activities in the form of a signal representing the potential for a given period. Brain–Computer Interface (BCI) systems are one of the practical application of EEG signal. Response to mental task is a well-known type of BCI systems which augments the life of disabled persons to communicate their core needs to machines that can able to distinguish among mental states corresponding to thought responses to the EEG. The success of classification of these mental tasks depends on the pertinent set formation of features (analysis, extraction, and selection) of the EEG signals for the classification process. In the recent past, a filter-based heuristic technique, Empirical Mode Decomposition (EMD), is employed to analyze EEG signal. EMD is a mathematical technique which is suitable to analyze a nonstationary and nonlinear signal such as EEG. In this work, three-stage feature set formation from EEG signal for building classification model is suggested to distinguish different mental states. In the first stage, the signal is broken into a number of oscillatory functions through EMD algorithm. The second stage involves compact representation in terms of eight different statistics (features) obtained from each oscillatory function. It has also observed that not all features are relevant, therefore, there is need to select most relevant features from the pool of the formed features which is carried out in the third stage. Four well-known univariate feature selection algorithms are investigated in combination with EMD algorithm for forming the feature vectors for further classification. Classification is carried out with help of learning

A. Gupta (✉) · A. Chakraborti · V. Kumar Singh
School of Computational and Integrative Sciences, Jawaharlal Nehru University,
New Delhi, India
e-mail: akshanshgupta@jnu.ac.in

A. Chakraborti
e-mail: anirban@jnu.ac.in

V. Kumar Singh
e-mail: vinod.acear@gmail.com

D. Kumar
AIM & ACT Banasthali Vidyapith, Niwai, Rajasthan, India
e-mail: dhirendrakumar@banasthali.in

M. Tanveer and R. B. Pachori (eds.), *Machine Intelligence and Signal Analysis*,
Advances in Intelligent Systems and Computing 748,
https://doi.org/10.1007/978-981-13-0923-6_28

the support vector machine (SVM) classification model. Experimental result on a publicly available dataset shows the superior performance of the proposed approach.

1 Introduction

The Brain–Computer Interface (BCI) is one of the regions which has sponsored up in developing techniques for assisting neurotechnologies for ailment prediction and manage motion [1, 2, 15]. BCIs are rudimentary geared toward availing, augmenting, or rehabilitating human cognitive or motor-sensory characteristic [14, 19]. To capture brain activities, EEG is one of the prevalent technology as it provides a signal with high temporal resolution in a noninvasive way [14, 15]. Mental task classification (MTC)-based BCI is one of the famed categories of BCI technology which does no longer involve any muscular activities [3], i.e., EEG responses to mental tasks.

In the literature, the EEG signals have been analyzed especially in three domains specifically temporal, spectral, and hybrid domain. In a hybrid domain, both the frequency and temporal information is utilized for analysis of the EEG signals simultaneously. Empirical mode decomposition (EMD) is this sort of heuristic hybrid approach that can examine the signal in both domains by decomposing the signal in distinctive frequency components termed as Intrinsic Mode Function (IMF) [13]. In literature, EMD has been incorporated for data analysis followed by using these decomposed signals for parametric feature vector formation for building classification model [6, 10].

In this work, final set of feature vectors for the classification process is obtained in three stages. In the first stage, the raw EEG signal is analyzed using EMD algorithms which results into a number of IMFs. A compact representation of these IMFs with the parametric feature coding has been introduced with the help of eight well-known parameters namely root mean square, variance, skewness, kurtosis, Lampleziv Complexity, central and maximum frequency, and Shannon entropy. Further to select only relevant features, four univariate feature selection methods are investigated which is the third stage of the proposed method for obtaining the final feature vectors for classification.

Outline of this article is as follows: Sect. 2 contains an overview of feature extraction and parametric feature formation. Feature selection approach is discussed in Sect. 3. In Sect. 4, a brief description of dataset and Experimental result are discussed. The conclusion is discussed in Sect. 5.

2 Feature Extraction

In this work, feature extraction from EEG signal has been carried out in two stages: First stage involves the decomposition of EEG signal from each channel into k number of intrinsic mode functions (IMFs) using Empirical Mode Decomposition (EMD) algorithm (discussed in Sect. 2.1). Later, in the second stage, these decomposed IMFs obtained from each channel were used to calculate eight parametric features. Hence, each signal can be transformed to more compact form. A brief description of EMD and parametric Feature vector construction are described in the following subsections.

2.1 Empirical Mode Decomposition (EMD)

EMD is a mathematical technique which is utilized to analyze a nonstationary and nonlinear signal. EMD assumes that a signal is composed of a series of different IMFs and decompose the signal into these continuous functions. Each IMFs have the following properties [13]:

1. Number of zero crossings and number of extrema are either equal or differ at most by one.
2. Local maxima and local minima produces the envelope whose mean value is equal to zero at a given point.

Figure 1 showed the plot of first four IMFs of an EEG segment using EMD algorithm. More details of this algorithm can be found in [13].

2.2 Parametric Feature Vector Construction

For constructing feature vector from the decomposed EEG signal, we have calculated eight parameters using moment values, complexity measure, and uncertainty values of the decomposed signal. The moments characterize the decomposed signal by certain statistical properties, a complexity measure shows repetitive nature in

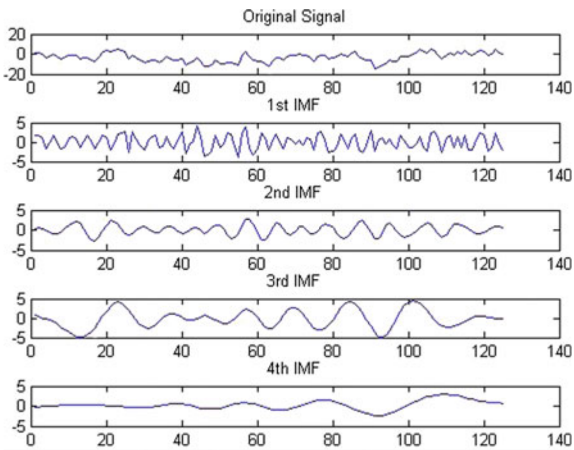

Fig. 1 IMF plot obtained for a given EEG signal

the time-series signal of decomposed signal and the uncertainty value denotes how much information contained by the signal. These parameters are root mean square, variance, skewness, kurtosis, Lampleziv Complexity, central & maximum frequency, and Shannon entropy of the signal.

3 Feature Selection

Feature selection [12, 16] is one of the approach to determine relevant features. Inspite of available rich research works on feature selection, not much work has been done in the area of mental task classification. The feature selection can be done using two methods. First method is classifier independent and relevance of the feature is measured by its inherent statistical properties, such as distance measure, correlation, etc.This approach is also known as filter method of feature selection. The second is wrapper method, where feature selection is classifier dependent and choose optimal subset of features to enhance accuracy of classifier. The wrapper based methods [16] find optimal or relevant subset of features from all possible combination of subsets of features and require classifier to evaluate the performance of the subset. Therefore, the computational cost of wrapper methods is much higher than filter methods.

Univariate (ranking) method, a filter feature selection method, is simple to compute. In these methods, a scoring function is used to measure the relevance, information content, discriminatory capability, or quality index of each feature separately without involving any classifier. Features are then ranked on the basis of their score in the order of its relevance. Many univariate feature selection methods are suggested in the literature. We have investigated four well known and commonly used univariate filter feature selection methods, namely, Fisher discriminant ratio, Pearson's correlation coefficient, Mutual information, and Wilcoxons rank sum test to determine a set of relevant features for mental task classification. A brief description of these methods is given below.

The problem under consideration has n samples of EEG signal, d features, and m distinct classes for mental task problem. Lets assume that matrix $\mathbf{X}$ represents available EEG data of dimension $n \times d$, where n is total number of samples and d represents total number of features. Here, each row $\mathbf{x}_i$ in matrix represents sample from class label c_i where $i = 1, 2, \ldots, m$ and each column $\mathbf{f}_j$ in matrix represents feature vector. Thus, the matrix $\mathbf{X}$ is represented as.

3.1 Pearson's Correlation

Pearson's correlation coefficient (CORR) [7, 18] is employed to determine linear correlation between two variables or between a variable and class label. The Pearson's correlation coefficient (CORR) of kth feature vector ($\mathbf{f}_k$) with the class label vector ($\mathbf{c}$) is given by

$$CORR\,(\mathbf{f}_k, \mathbf{c}) = \frac{cov(\mathbf{f}_k, \mathbf{c})}{\sigma_{\mathbf{f}_k}\sigma_c} = \frac{E[(\mathbf{f}_k - \mu_k)(\mathbf{c} - \bar{c})]}{\sigma_{\mathbf{f}_k}\sigma_c}, \quad for\ k = 1,\ 2,\ \ldots,\ d \quad (1)$$

where $\sigma_{\mathbf{f}_k},\ \sigma_{\mathbf{c}}$ represent respectively the standard deviations of feature vector $\mathbf{f}_k$ and $\mathbf{c}$. $cov\,(\mathbf{f}_k,\ \mathbf{c})$ represents the covariance between $\mathbf{f}_k$ and $\mathbf{c}$, $\mu_k = \frac{1}{n}\sum_{i=1}^{n} X_{ik}$ and $\bar{c} = \frac{1}{n}\sum_{i=1}^{n} c_i$ are the mean of $\mathbf{f}_k$ and $\mathbf{c}$ respectively.

The value of $CORR\,(\mathbf{f}_k, \mathbf{c})$ lies between -1 and $+1$. The correlation value closer to -1 or 1, shows the stronger correlation among the prescribed variables while zero value implies no correlation between the two variables. It can measure both the degree as well as the trend of correlation. Also, it is invariant to linear transformations of underlying variables. However, the assumption of linear relationship between the variables is not always true. Also, sometimes the value of correlation coefficient may misinterpret the actual relation, as a high value does not always imply a close relationship between the two variables and it is sensitive to outliers too.

3.2 Mutual Information

Mutual information is an information theoretic-based ranking method which measures dependency between two variables. The mutual information of a feature vector $\mathbf{f}_k$ and the class vector $\mathbf{c}$ is given by [20]:

$$I\,(\mathbf{f}_k, \mathbf{c}) = \sum P(\mathbf{f}_k, \mathbf{c}) \log \frac{P(\mathbf{f}_k, \mathbf{c})}{P(\mathbf{f}_k)P(\mathbf{c})} \tag{2}$$

where $P(\mathbf{f}_k)$ and $P(\mathbf{c})$ are the marginal probability distribution functions for random variables $\mathbf{f}_k$ and $\mathbf{c}$ respectively and $P(\mathbf{f}_k, \mathbf{c})$ is joint probability distribution.

The maximum value of mutual information indicates the higher dependency of variable on the target class. The advantage of mutual information is that it can capture even the nonlinear relationship between the feature and the corresponding class label vector $\mathbf{c}$.

3.3 Fisher Discriminant Ratio

Fisher Discriminant Ratio (FDR) is another univariate filter feature selection method which is based on the statistical properties of the features. FDR($\mathbf{f}_k$) for kth feature for two class i and j is defined as:

$$FDR(\mathbf{f}_k) = \frac{\left(\mu_{i(k)} - \mu_{j(k)}\right)^2}{\sigma_{i(k)}^2 + \sigma_{j(k)}^2} \tag{3}$$

where $\mu_{i(k)}$ and $\sigma^2_{i(k)}$ are the mean and dispersion of the data of class i, respectively, for kth feature. It takes maximum value when the square of the difference between mean of two classes for feature $\mathbf{f}_k$ is maximum and sum of variances of corresponding feature for both class i and j is minimum.

3.4 Wilcoxon's Ranksum Test

Wilcoxon Ranksum Test, proposed by [21], is a statistical test, which is carried out between data of two classes on the basis of median of the samples without assuming any probability distribution.

The statistical difference $t(\mathbf{f}_k)$ of feature $\mathbf{f}_k$ for given two classes i and j using Wilcoxon's statistics is defined as [17]:

$$t\left(\mathbf{f}_k\right) = \sum_{l=1}^{N_i}\sum_{m=1}^{N_j} DF((X_{lk} - X_{mk}) \leq 0) \tag{4}$$

where N_i and N_j are the number of the samples in ith and jth class respectively, DF is the logical distinguishing function between two classes which assigns a value of 1 or 0 corresponding to true or false and X_{lk}, is the expression values of kth feature for lth sample. The value of $t(\mathbf{f}_k)$ tends from zero to $(N_i \times N_j)$. The relevance of the feature is determined as:

$$R\left(t(\mathbf{f}_k)\right) = \max(t(\mathbf{f}_k), N_i \times N_j - t(\mathbf{f}_k)) \tag{5}$$

4 Experimental Setup and Result

4.1 Dataset and Constructing Feature Vector

In order to check the effectiveness of the proposed method, experiments have been performed on a publicly available dataset[1] [15] which consists of recordings of EEG signals using six electrode channels from seven subjects with the recording protocols. Each subject was asked to perform 5 different mental tasks as namely *Baseline task relax* (B), *Letter Composing task* (L), *Non trivial Mathematical task* (M), *Visualizing Counting* (C) of numbers written on a blackboard, and *Geometric Figure Rotation* (R) task. For conducting the experiment, data from all the subjects are utilized except Subject 4; as data recorded for Subject 4 is incomplete [8].

The EEG signal corresponding to each mental task of a particular subject is formed into half-second segments which yields into 20 segments (signal) per trial per channel.

[1] http://www.cs.colostate.edu/eeg

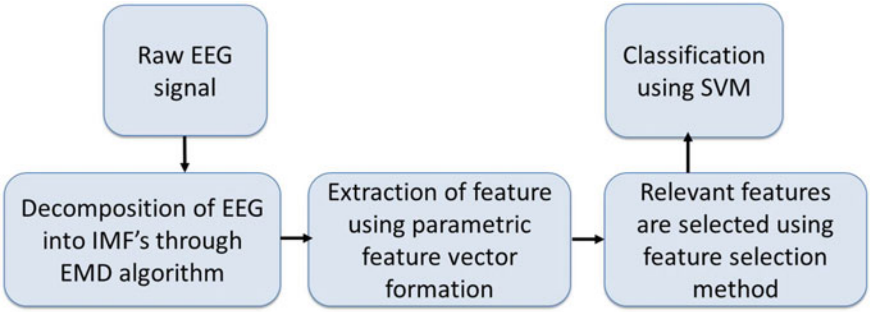

Fig. 2 Flow diagram of the proposed method

Thus, for every channel, each of 20 segments are decomposed using EMD algorithm into 4 IMFs. The eight parameters are extracted for each of these IMFs per segment per channel per trial for a given subject. A set of aforementioned eight statistical parameters is obtained for each of the six channels of the signal and these sets were concatenated to form a feature vector. Hence, the final feature vector is of 192 dimensions (four IMFs $\times$ 8 parameters $\times$ six channels) after applying the parametric feature vector formation step. As the dimension of feature vector are still high and not all features are relevant for classification so feature selection methods are utilized for selecting only relevant features for classification which results in lowering the time for building the classification model. Figure 2 shows complete pipeline for constructing the feature vector from each subject using all trial corresponding to each mental tasks labels (B, L, M, C, and R) for further classification using SVM classifier.

4.2 Results and Discussion

As discussed in the previous subsection, a set of feature vectors have been obtained corresponding to every mental task labels (B, L, M, C, and R). Binary mental task classification problem has been formulated to distinguish the different mental state of different subjects. The optimal value of SVM regularization parameters, i.e., gamma and cost, were obtained with the help of grid search algorithm. The average classification accuracy of 10 runs of 10 cross-validations has been reported. Figure 3 shows average classification accuracy of different binary combination of mental tasks averaged over all subjects corresponding to different feature selection techniques. Number of relevant features selected corresponding to given feature selection method is summarized in Fig. 4. From these figures it can be noted that incorporating the feature selection techniques will lead to better accuracy in comparison to without feature selection.

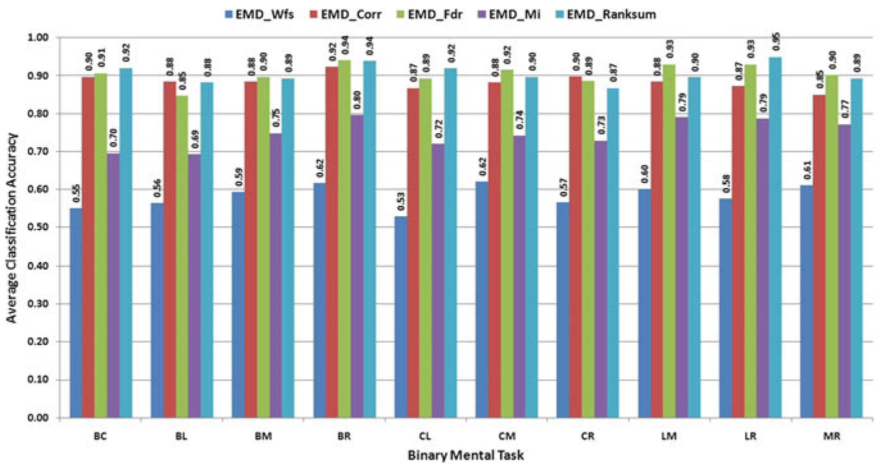

Fig. 3 Average classification accuracy for different binary mental task combinations over all subjects

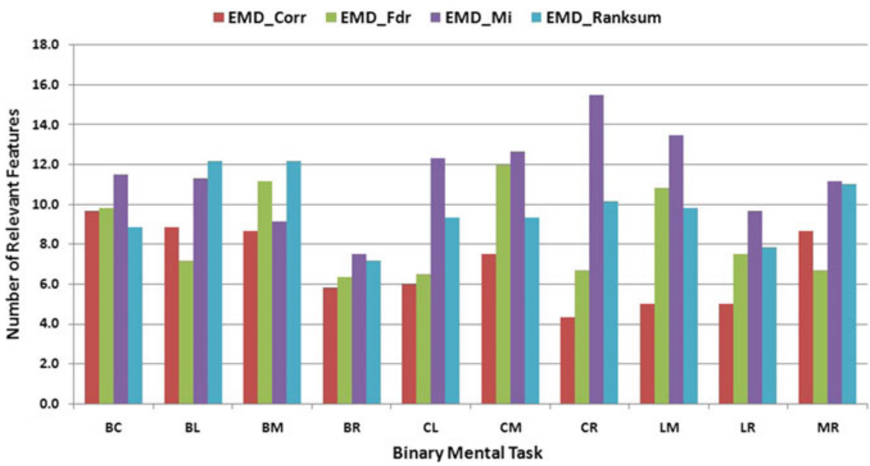

Fig. 4 Average number of relevant features corresponding to different binary mental task combinations over all subjects

The research work of Gupta and Kirar [11] used 48 features using parametric feature extraction method for the classification. However, our current approach has generated 192 features from EMD signal, which may have some irrelevant features that may hinder classification performance in terms of accuracy and time. The feature selection techniques have huge advantage in terms of dimensionality reduction and removal of irrelevant features. Hence, reduced set of relevant features increase the performance of classifiers [4]. The study compared the results of some popular feature selection techniques (Corr, MI, FDR, and Ranksum) with complete set of features

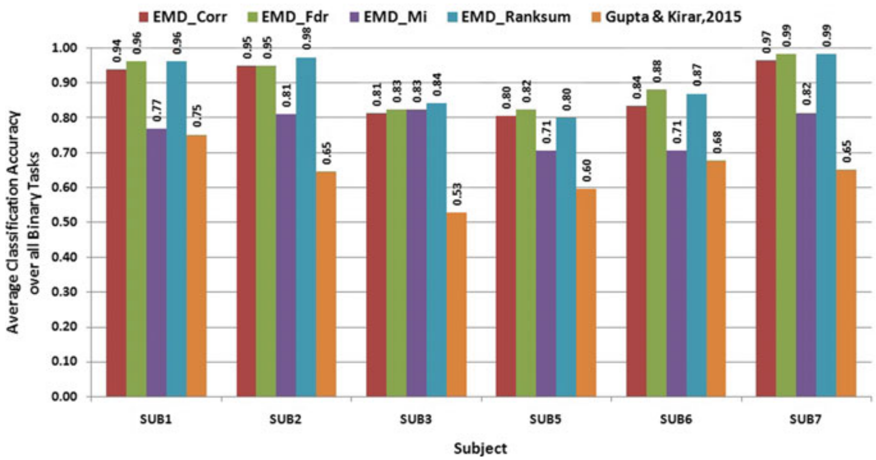

Fig. 5 Comparison of performance of various classification approaches on different subjects

(WFS) for mental task classification. The feature selection techniques had shown promising results in terms of accuracy over WFS (See Fig. 3) and the research work of Gupta & Kirar [11] (See Fig. 5). The training time of classifiers would also reduced significantly as the number of relevant features after applying the feature selection methods are much lesser than that of without feature selection (See Fig. 4). On the basis of two classification performance parameter, we observed that correlation-and FDR-based feature selection methods are well suited for mental task classification. Hence, our model can be beneficial for the differently abled persons to communicate with the machine more efficiently, i.e., quickly and accurately.

4.3 Friedman Statistical Test

For assessment of the significant difference among combinations of different feature selection methods with EMD and the research work [11], statistically, a two way [5] and nonparametric statistical test known as Friedman test [9] has been conducted. Where, H_0 was null hypothesis that assumes there is no significance difference in performance in among all approach. The H_0 was rejected at 95% of confidence level. Table 1 shows the ranking of different methods. Lowest rank for a given method shows its better performance compared to other methods. From Table 1, it can be noted that the combination of EMD feature extraction with FDR feature selection performs better in comparison to other methods.

Table 1 Friedman's ranking

Algorithms	Ranking
EMD_Corr	2.65
EMD_Fdr	1.65
EMD_Mi	4
EMD_Ranksum	1.7
Gupat & Kirar,2015	5

5 Conclusion

The EEG signals are used to capture the cognitive activities and each activity had embedded hidden patterns. Our study employed effective machine learning strategy to capture the hidden patterns from the EEG signal of different mental tasks and make prediction about the unknown mental task from the given signal. In this work, EMD algorithm is used to decomposed EEG signal into IMFs and parametric features are calculated for forming the feature vectors. Further, for selecting only relevant features, four well-known univariate feature selection techniques are investigated which reduces the dimension of feature vectors which results into reduction of time in building the classification model. The experiment has been performed on a publicly available EEG dataset which contains the responses to different mental thought regarding some task. The experimental results show the performance of the proposed approached for binary mental task classification problem is improved after incorporating the feature selection in conjunction with EMD algorithm.

Acknowledgements Authors express their gratitude to Cognitive Science Research Initiative (CSRI), DST & DBT, Govt. of India & CSIR, India for obtained research grant.

References

1. Anderson, C.W., Stolz, E.A., Shamsunder, S.: Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks. IEEE Trans. Biomed. Eng. **45**(3), 277–286 (1998)
2. Babiloni, F., Cincotti, F., Lazzarini, L., Millan, J., Mourino, J., Varsta, M., Heikkonen, J., Bianchi, L., Marciani, M.: Linear classification of low-resolution eeg patterns produced by imagined hand movements. IEEE Trans. Rehabil. Eng. **8**(2), 186–188 (2000)
3. Bashashati, A., Fatourechi, M., Ward, R.K., Birch, G.E.: A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. J. Neural Eng. **4**(2), R32 (2007)
4. Bennasar, M., Hicks, Y., Setchi, R.: Feature selection using joint mutual information maximisation. Expert Syst. Appl. **42**(22), 8520–8532 (2015)
5. Derrac, J., García, S., Molina, D., Herrera, F.: A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm Evol. Comput. **1**(1), 3–18 (2011)
6. Diez, P.F., Torres, A., Avila, E., Laciar, E., Mut, V.: Classification of Mental Tasks Using Different Spectral Estimation Methods. INTECH Open Access Publisher (2009)

7. Dowdy, S., Wearden, S., Chilko, D.: Statistics for Research, vol. 512. Wiley, New York (2011)
8. Faradji, F., Ward, R.K., Birch, G.E.: Plausibility assessment of a 2-state self-paced mental task-based bci using the no-control performance analysis. J. Neurosci. Methods **180**(2), 330–339 (2009)
9. Friedman, M.: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc. **32**(200), 675–701 (1937)
10. Gupta, A., Agrawal, R., Kaur, B.: Performance enhancement of mental task classification using eeg signal: a study of multivariate feature selection methods. Soft Comput. **19**(10), 2799–2812 (2015)
11. Gupta, A., Kirar, J.S.: A novel approach for extracting feature from eeg signal for mental task classification. 2015 International Conference on Computing and Network Communications (CoCoNet), pp. 829–832. IEEE (2015)
12. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)
13. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H.: The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. Lond. Ser. A: Math. Phys. Eng. Sci. (1971) **454**, 903–995 (1998)
14. Kauhanen, L., Nykopp, T., Lehtonen, J., Jylanki, P., Heikkonen, J., Rantanen, P., Alaranta, H., Sams, M.: Eeg and meg brain-computer interface for tetraplegic patients. IEEE Trans. Neural Syst. Rehabil. Eng. **14**(2), 190–193 (2006)
15. Keirn, Z.A., Aunon, J.I.: A new mode of communication between man and his surroundings. IEEE Trans. Biomed. Eng. **37**(12), 1209–1214 (1990)
16. Kohavi, R., John, G.H.: Wrappers for feature subset selection. Artif. Intell. **97**(1), 273–324 (1997)
17. Li, S., Wu, X., Tan, M.: Gene selection using hybrid particle swarm optimization and genetic algorithm. Soft Comput. **12**(11), 1039–1048 (2008)
18. Pearson, K.: Notes on the history of correlation. Biometrika, 25–45 (1920)
19. Pfurtscheller, G., Neuper, C., Schlogl, A., Lugger, K.: Separability of eeg signals recorded during right and left motor imagery using adaptive autoregressive parameters. IEEE Trans. Rehabil. Eng. **6**(3), 316–325 (1998)
20. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication (urbana, il) (1949)
21. Wilcoxon, F.: Individual comparisons by ranking methods. Biom. Bull. **1**(6), 80–83 (1945)