Relevant Feature Selection from EEG Signal for Mental Task Classification

Akshansh Gupta and R.K. Agrawal

School of Computer and Systems Sciences
Jawaharlal Nehru University, New Delhi India 1110067
{akshansgupta83,rkajnu}@gmail.com

Abstract. In last few years, the research community has shown interest in the development of Brain Computer Interface which may assists physically challenged people to communicate with the help of brain signal. The two important components of such BCI system are to determine appropriate features and classification method to achieve better performance. In literature, Empirical Mode Decomposition is suggested for feature extraction from EEG which is suitable for the analysis of non-linear and non-stationary time series. However, the features obtained from EEG may contain irrelevant and redundant features which make them inefficient for machine learning. Relevant features not only decrease the processing time to train a classifier but also provide better generalization. Hence, relevant features which provide maximum classification accuracy are selected using ratio of scatter matrices, Chernoff distance measure and linear regression. The performance of different mental task using different measures used for feature selection is compared and evaluated in terms of classification accuracy. Experimental results show that there is significant improvement in classification accuracy with features selected using all feature selection methods and in particular with ratio of scatter matrices.

Keywords: Empirical Mode Decomposition, Brain Computer Interface, Feature Selection, Chernoff distance measure, Scatter Matrices, Linear regression.

1 Introduction

Last few years have witnessed the advancement of technologies which has made possible the use of brain signals for communication between human and computer. This growth in technologies allows research community to develop a system called Brain Computer Interface (BCI) which can control a device such as computer or wheel chair by human intentions rather than mechanical power of human. It may be very useful to physically challenged persons who are suffering from locomotor syndrome, Amyotrophic Lateral Sclerosis, Head trauma, severe cerebral palsy or multiple disorders affect in body, which restricts such persons to operate any electronics device smoothly and freely. With the development of BCI, these people can operate any electronics device with the help of just

P.-N. Tan et al. (Eds.): PAKDD 2012, Part II, LNAI 7302, pp. 431–442, 2012.

brain signals and does not depend on the brain's normal output pathways of peripheral nerves and muscles. BCIs are often aimed for assisting, augmenting or repairing human cognitive or motor sensory function. Various techniques such as Electroencephalogram (EEG), Electrocardiogram, functional magnetic resonance imaging, Magneto encephalographic (MEG) and Positron emission tomography (PET) are used for monitoring brain signals activities.

EEG is commonly used for BCI implementation due to its low cost, ability to record brain signals and non-invasive nature. There are many components of a BCI system. However, the success of BCI system mainly depends on two components: feature extraction and classification method. The feature extracted/selected from EEG should have high discriminative power to distinguish the different tasks and the classification methods used to distinguish the different tasks should be efficient in real time. There are many classification methods available in the field of data mining and machine learning[16,20,31]. The research work[22] discusses pros and cons of linear and classification methods for BCI research.

In literature, autoregressive (AR) models or adaptive AR models (AAR)[1,3,7,9,20,26] and power spectral density (PSD)[2,27] are commonly used for feature extraction from EEG for BCI system. However, these methods assume linearity, Gaussianality and minimum-phase within EEG signals, i.e., the amplitudes of EEG signals are normally distributed, their statistical properties do not vary over time, and their frequency components are uncorrelated. Under these assumptions, the EEG signal is considered as a linear superposition of statistically independent sinusoidal or other wave components, and only frequency and power estimates are considered while phase information is lost. Recently, Empirical Mode Decomposition (EMD) is suggested for feature extraction from EEG signal which is suitable for the analysis of non-linear and non-stationary time series. A disadvantage arising at this point is that the feature vector so obtained with EMD would be too large and the number of training samples available are in general relatively small number. Consequently, it is essential to do a feature selection in order to solve the problem of curse-of-dimensionality which arises due to small sample and large number of features[17]. Also, the resultant features may contain noisy, irrelevant or redundant features which make them inefficient for machine learning. In fact, the presence of irrelevant and redundant features may deteriorate the performance of the classifier and requires high computation time and other resources for training and testing the data. Hence, in order to enhance the performance of BCI system in terms of accuracy and time required to detect, there is need to identify a set of relevant features.

Feature selection is used to remove such noisy, irrelevant, and redundant features.There are two major approaches to feature selection: filter and wrapper approach[10,14,22]. Most filter methods employ statistical characteristics of data for feature selection which requires less computation. It independently measures the importance of features without involving any classifier. Since, the filter approach does not take into account the learning bias introduced by the final learning algorithm, it may not be able to select the most relevant set of features for the learning algorithm. On the other hand, wrapper methods tend to find features better suited to the predetermined learning algorithm resulting in better performance. But, it tends to be computationally more expensive since the classifier must be trained for each candidate subset.

Feature ranking approaches have been widely investigated for feature selection[10,21,23] in literature. Since in most of feature ranking approaches, features are evaluated using statistical characteristics of the data, different feature ranking methods measure different characteristics of data. Therefore, the informative features selected by different ranking methods may be different. In literature to remove redundancy a forward/backward feature selection method or its combinations are used with a measure that selects relevant and non redundant features. Among the most widely used filter methods for feature selection, there are techniques based on statistical separability measures which allow one to select a suitable subset of features by assigning the degree of interclass separability associated with each subset. In particular, ratio of scatter matrices, Chernoff distance measures[19] and linear regression[21] are commonly employed by research community in various area of data mining and pattern classification field but yet to be explored in feature selection of EEG data for mental task classification. In this paper, we compare and evaluate these measures to determine relevant features for BCI system.

Our work is organized as follows: Feature extraction using empirical mode decomposition is included in Sect. 2. A brief introduction of separability measures employed for features selection are discussed in Sect. 3. Experimental data and results are discussed in Sect. 4 and Sect. 5 contains conclusions.

2 Feature Extraction from EEG

The feature extraction is carried out in two phases [12]: in the first phase, the empirical mode decomposition is used, and the second phase estimates different time and frequency parameters.

2.1 Empirical Mode Decomposition (EMD)

Under the assumption that any signal is composed of a series of different intrinsic oscillation modes, the EMD can be used to decompose the incoming signal into its different Intrinsic Mode Function (IMF). An IMF is a function that satisfies two conditions[12]:

1. In the entire signal, the number of extremes and the zero-crossings must be equal or differ at most by one.
2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima must be zero.

Given the incoming signal x(t), the algorithm of EMD is based on a sifting process that can be summarized as[12,14]:

1. Interpolate all the local maxima and minima in the signal with a cubic spline line, to produce the upper and lower envelope.
2. Repeat for the local minima to produce the lower envelope.
3. Compute the mean of both envelopes m_1.
4. Extract the detail $h_1 = x(t) - m_1$
5. Repeat the steps 1 to 4, and consider the detail h_i as the data, until detail h_1 can be considered an IMF.
6. After k iterations, the detail h_k is an IMF and is designated as: $IMF_1 = h_k$
7. Iterate steps 1 to 6 on the residual r_j in order to obtain all the IMFs of the signal:

$$r_j = x\,(t) - IMF_1 - IMF_2 ... IMF_m \tag{1}$$

The procedure terminates when the residual r_j is either a constant, a monotonic slope, or a function with only one extreme. The result of the EMD process produces n IMFs and a residue signal r_n. The original signal x(t) can be reconstructed summing up the n extracted IMF and the residue:

$$x(t) = \sum_{j=1}^{n} IMF_j + r_j \tag{2}$$

2.2 Estimation of Various Parameters

In order to obtain the IMFs of the signal, publicly available EMD toolbox for Matlab was utilized. The lower-order IMFs capture the faster oscillation modes of the signal, whereas the higher-order IMFs capture the slower oscillation modes. The EMD algorithm can be applied to each EEG 1 s segments. Afterward, the EMD is able to extract no more than five IMFs and the residue for each 1 s EEG segment. For each one of these five IMFs, different parameters can be computed. The following parameters can be used to represent each EMD[5]:

1. Root Mean Square (RMS),
2. Variance,
3. Shannon entropy[23]
4. Lempel-Ziv Complexity Measure[13],
5. Central Frequency (50 % of spectrum energy)
6. Maximum Frequency (95 % of spectrum energy)

Some parameters were chosen since they are commonly used in BCI (RMS, variance), LZ quantifies the complexity of a signal analysing its spatial-temporal patterns and was used to analyse EEG signals in other areas[10]. The central and maximum frequencies were used as descriptors of the bandwidth of each IMF. Entropy was used to measure the average amount of information in a signal.

3 Feature Selection

A disadvantage arising at this point is feature vector contains 180 parameters (5 IMFs x 6 parameters 6 channels). Consequently, it is essential to do a feature selection in order to solve the curse-of-dimensionality inconvenience[24]. Feature ranking is commonly used to determine a subset of relevant features. However, the disadvantage of feature ranking method is that they ignore the correlations between features. Hence the features selected may contain redundant information and influences the classification capabilities of the feature subset that is selected. Some of the methods suggested in literature for removing redundancy are Chernoff distance measure[22], ratio of inter-class and with-in class scatter, and linear regression[19]. In order to obtain a quantitative measure of how separable are two classes, a distance measure can be easily extracted from some parameters of the data. A very important aspect of probabilistic distance measures is that a number of these criteria can be analytically simplified in the case when the class conditional p.d.f.s follows multivariate normal distribution. The class conditional probability densities functions $p(X_k|C_i)$ of k-dimensional samples $\mathbf{X_k} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k]$ for a given class $C_i, i = 1, 2, 3, ..., k$ is given by

$$p(\mathbf{X}_k|C_i) = \frac{1}{(2\pi)^{\frac{d}{2}} \left|\sum_{\mathbf{k}}^{\mathbf{i}}\right|} \exp\left[-\frac{1}{2}(\mathbf{X}_k - \mu_k^i)^T (\sum_k^i)^{-1} (\mathbf{X}_k - \mu_{\mathbf{k}})\right] \tag{3}$$

where $\mu_{\mathbf{k}}^{\mathbf{i}}$ is a mean vector and $\sum_k^i$ is a covariance matrix for class C_i . In literature, for multivariate normal distribution for two classes, CD measure is given as follows[5]:

$$J_k^c = \frac{1}{2}\boldsymbol{\beta}(1-\boldsymbol{\beta})(\mu_k^2 - \mu_k^1)^T[(1-\boldsymbol{\beta})\sum_k^1 + \boldsymbol{\beta}\sum_k^2]^{-1}(\mu_k^2 - \mu_k^1) + \frac{1}{2}log\frac{|(1-\boldsymbol{\beta})\sum_k^1 + \boldsymbol{\beta}\sum_k^2|}{\left|\sum_k^1\right|^{1-\boldsymbol{\beta}}\left|\sum_k^2\right|^{\boldsymbol{\beta}}} \tag{4}$$

A major disadvantage of the class separability measure CD is that it is not easily computed, unless the Gaussian assumption is employed. In literature, a simpler criteria based on the scatters of feature vector samples is employed. To this end, the scatter matrices: within-class scatter and between-class scatter are respectively defined as:

$$S_w = \sum_k^1 + \sum_k^2 \tag{5}$$

$$S_b = (\mu_k^2 - \mu_k^1)(\mu_k^2 - \mu_k^1) \tag{6}$$

From these definition of scatter matrices, it is straightforward to observe that the criterion

$$J = \frac{|S_b|}{|S_w|} \tag{7}$$

takes large values when samples of the selected features space are well clustered around their mean within each class, and the clusters of the different classes are well separated. Also, the criteria J have the advantage of being invariant under linear transformation.

The regression analysis considers the relations between the selected features which minimizes redundancy. While using regression analysis for data, a multiple regression model is considered because there can be many features which could affect the presence or absence of samples from a particular class. A multiple regression model with a target variable y and multiple variables X is given by [15]:

$$y = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_{i1} + \boldsymbol{\beta}_2 x_{i2} + ... + \boldsymbol{\beta}_m x_m + \varepsilon_i, i = 1, 2, ...m \quad (8)$$

Where $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, ...\boldsymbol{\beta}_m$ are constants estimated by observed values of X and class label y and is estimated by normal distribution having mean zero and a variance σ^2. The sum of square errors (SSE) is given by

$$SSE = \sum_{i=0}^{n} (y_i - y_i^p) \quad (9)$$

Where y and $y^p$are observed and predicted values respectively. A large value of SSE means that the regression is predicted poorly. The total sum of squares is given by

$$SSTO = \sum_{i=0}^{n} (y_i - \bar{y}) \quad (10)$$

Where $\bar{y}$ is the average of y_i. In a regression model the choice of features which best explains the class label depends on the value of R^2 which is given by

$$R^2 = 1 - \frac{SSE}{SSTO} \quad (11)$$

4 Experimental Set-Up and Results

The EEG data used in our experiment was acquired by Keirn and Aunon[29] using the following procedure. The subjects were seated in an Industrial Acoustics Company sound controlled booth with dim lighting and noiseless fans for ventilation. An Electro-Cap elastic electrode cap was used to record from positions C3, C4, P3, P4, O1, and O2, defined by the 10-20 system of electrode placement. The electrodes were connected through a bank of Grass 7P511 amplifiers and bandpass filtered from 0.1100Hz. Data was recorded at a sampling rate of 250 Hz with a Lab Master 12 bit A/D converter mounted in an IBM-AT computer. Eye blinks were detected by means of a separate channel of data recorded from two electrodes placed above and below the subjects left eye.

For our experiment, the data from six subjects except subject 5 performing five different mental tasks were analyzed. The five mental tasks are: the baseline(B) task, for which the subjects were asked to relax as much as possible; the letter(L) task, for which the subjects were instructed to mentally compose a letter to a friend or relative without vocalizing; the math(M) task, for which the subjects were given non-trivial multiplication problems, such as 49 times 78, and were asked to solve them without vocalizing or making any other physical movements; the visual counting(C) task, for which the subjects were asked to imagine a blackboard and to visualize numbers being written on the board sequentially; and the geometric figure rotation(R), for which the subjects were asked to visualize a particular three-dimensional block figure being rotated about an axis.

Data was recorded for 10 seconds during each task and each task was repeated five times per session. Most subjects attended two such sessions recorded on separate weeks, resulting in a total of 10 trials for each task. With a 250 Hz sampling rate, each 10 second trial produces 2,500 samples per channel. These are divided into half-second segments that overlap by one quarter-second, producing at most 39 segments per trial segments containing eye blinks are discarded.

Features are extracted from each one of signal using EMD. Each signal is represented in terms of 180 statistics (5 IMFs x 6 parameters 6 channels). To remove redundancy from the selected pool of features, three feature selection measures are investigated: Chernoff distance measure, ratio of with-in class scatter and between class scatter and linear regression. For Chernoff distance measure, features are selected using 3 different values ranging from 0.1 to 0.9 with an increment of 0.4. We have used linear discriminate classifier (LDC), Quadratic discriminate classifier (QDC), k-nearest neighbor (KNNC) and Support vector machine (SVC) to evaluate the performance of the feature selection methods. The average classification accuracy is computed using ten cross-validations. All the simulations are done using matlab. Tables 1 and Table 2 show the minimum classification accuracy achieved with different classifiers and the number of features for different measures respectively. For Chernoff distance measure, the maximum classification accuracy achieved over different values of β is shown in Table 1. The best results in each category are indicated in bold. Figures 1-2 and Tables 1-2 show the variation of classification accuracies and minimum number of features for the different mental tasks respectively. Figures 1-2 shows at end of this manuscript. We observe the following from Tables 1-2:

1. The classification accuracy of all mental tasks classification improved significantly with the use of feature selection.
2. The maximum average classification accuracy of mental tasks is achieved with feature selection method using ratio of scatter matrices for all classifiers except KNN.
3. The average classification accuracy of mental tasks with SVC and LDC are similar and better in comparison to QDC and KNN using all feature selection methods.
4. The performance of ratio of scatter matrices in combination of both LDC and SVC is better in terms of classification accuracy in comparison to other combination of a classifier and feature selection method.
5. The number of features required to obtain maximum classification accuracy is significantly smaller using feature selection methods in comparison to baseline using all classifiers. In particular, the number of features selected in combination of KNN is relatively smaller in comparison to other classifiers. However, the classification accuracy is significantly less in comparison to other classifiers.
6. As the number of features required to obtain maximum classification accuracy is significantly smaller using feature selection methods, the computation time by all the learning methods will be significantly reduced.

Table 1. Variation in Classification Accuracy Different Mental Task

Task	BC	BL	BM	BR	CL	CM	CR	LM	LR	MR	Avg
WFS+ LDC	54.9	53.2	62.2	59.7	57	61.3	57.8	61.1	57	60.6	58.5
Scatt +LDC	92	93.3	97.4	94.6	92	96.3	93	96.1	93.5	94.4	94.3
JC+LDC	90.9	87.5	94.6	91.6	90.9	93.7	88.4	94.6	90.5	89.8	91.3
Reg+LDC	92.8	90.6	96.4	93.2	92.8	96.4	91.3	96	92.8	94.7	93.7
WFS+QDC	49.5	49.9	51.3	49	47.7	48.5	48.5	49.6	49.8	48.3	49.2
Scatt+QDC	91.3	88.3	95	92.9	91.3	94.3	91.5	95.5	91.6	92.5	92.4
JC+QDC	89.9	84.9	91.9	91.1	89.8	93	90.3	93.3	92.6	93.3	91
Reg_+QDC	90.4	85.7	95.4	92.1	90.4	94.1	90.1	95.5	91.7	92.4	91.8
WFS+KNNC	47.7	47	54.3	49.8	54	56	50	54.7	49.4	55.2	51.8
Scatt+KNNC	87	82	88.1	91.1	87	91	86.6	90.6	88	86.6	87.8
JC+KNNC	89.3	82	87.7	92.1	89.3	90.8	86.7	90.6	93.1	89.4	89.1
Reg+KNNC	84.1	79.4	86.4	88.9	84.1	87.8	85	90.5	88.5	86	86.1
WFS+SVC	58.4	59.8	65.1	62	59.7	62.4	63.2	65.8	63.6	66.6	62.7
Scatt+SVC	92.4	93.6	97.3	92.8	92.4	96.3	93.3	96.5	93.3	94.6	94.3
JC+SVC	91.9	88.5	95	93.2	91.9	94.6	91.7	96.5	94.2	94	93.2
Reg+SVC	92.8	90.6	96.6	91.1	92.8	96.4	92.3	96.3	92.8	94.7	93.6

Table 2. Variation of Number of Features required for Different Mental Task

Tasks	BC	B L	B M	B R	C L	C M	C R	L M	L R	M R	Avg
Scatter+LDC	15.3	23.7	21.3	13.8	15.3	19.3	17.7	15.5	14.8	19.3	17.6
JC+LDC	21.3	21.2	20.2	16.3	21.3	17.8	20	12.7	11.7	20	18.3
Reg+LDC	14.2	16.5	14.3	9.7	14.2	16.3	15.3	13.8	12.2	18.3	14.5
Scatter+QDC	12.7	19	17.5	12.2	12.7	19.2	15	14.8	13.3	14.8	15.1
JC+QDC	15.2	14.8	15.8	12.7	15.2	17.8	17	12	9.7	15.8	14.6
Reg+QDC	14.7	15.2	18.2	12.3	14.7	15.5	14.2	11.3	11.3	12.7	14
Scatter+KNNC	3.8	7.7	3.8	6.7	3.8	4.3	2.8	3.3	3.7	5	4.5
JC+KNNC	2.7	3.7	2.2	5.2	2.7	2.7	2.7	3.2	3	4.3	3.2
Reg+KNNC	6.8	7.3	5.5	8	6.8	5.7	5	5	8.2	8	6.6
Scatter+SVC	15.3	24	21.7	11.5	15.3	19.3	14.8	12.3	13.2	19.2	16.7
JC+SVC	21.3	21.2	18.7	12.8	21.3	17.8	22	15.8	14.5	20	18.6
Reg+SVC	13	14.3	13.7	7	13	11.7	13.8	13.7	10.8	13.3	12.4

* WFS=Without feature selection

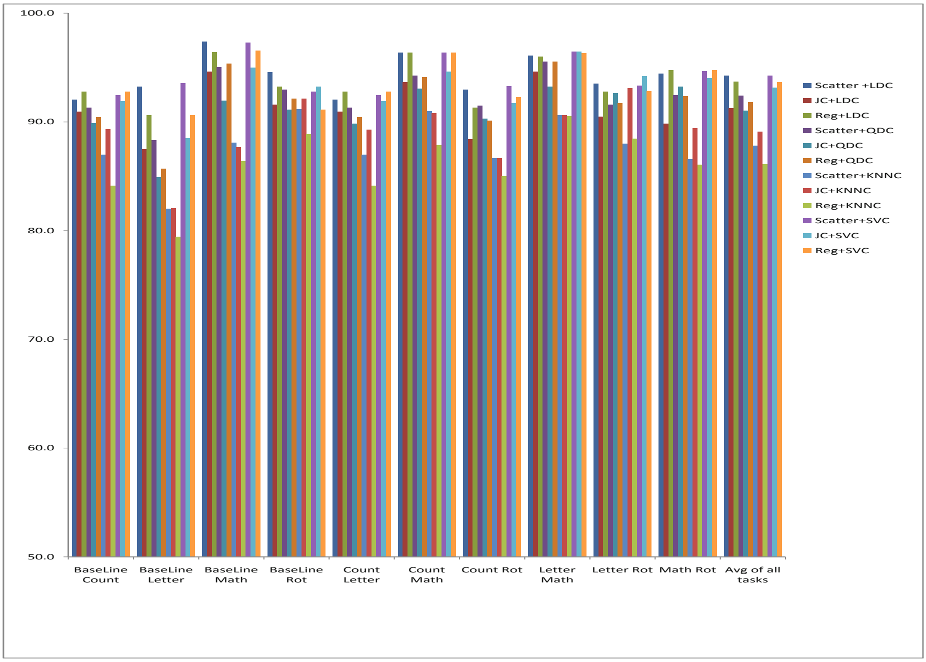

Fig. 1. Variation in Classification Accuracy for Different Mental Task

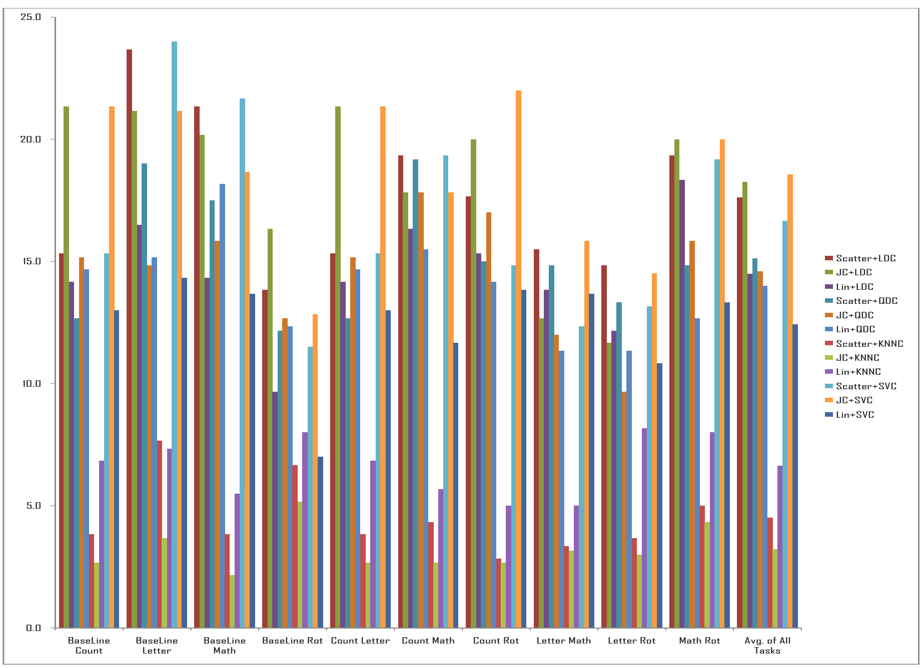

Fig. 2. Variation of Number of Features required for Different Mental Task

5 Conclusion

The performance of a classifier depends on the choice of features and classifier for any pattern recognition system. Features based on Empirical Mode Decomposition from EEG signal is extracted. These features may contain irrelevant and redundant features which makes them inefficient for machine learning. Hence, relevant features which provide maximum classification accuracy are selected using ratio of scatter matrices, Chernoff distance measure and linear regression. The performance of different mental task using different measures used for feature selection is compared and evaluated in terms of classification accuracy. Experimental results show that classification accuracy of all mental tasks classification improve significantly with the use of feature selection methods. In particular the performance of ratio of scatter matrix is better for all classifiers except KNN. The time required to learn the model will decrease significantly as the number of features reduces with the use of feature selections.In future, there is need to develop a feature selection method for mental task classification which gives better performance by all classifiers. It is also required to find out a method of feature extraction which extracts minimal and most relevant features from EEG signal for mental task classification and does not require any further feature selection.

References

1. Anderson, W.C., 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.G.: Linear classification of low-resolution EEG patterns produced by imagined hand movements. IEEE Trans. Rehabil. Eng. 8(2), 186–188 (2000)
3. Basseville, M., Benveniste, A.: Sequential segmentation of nonstationary digital signals using spectral analysis. Information Science 29(1), 57–73 (1983)
4. Richard, O., Peter, E., David, G.: Pattern Classification, 2nd edn. Wiley India (P) Ltd
5. Diez, P.F., Mut, V., Laciar, E.: A location of the empirical mode decomposition to the extraction of features from EEG signals for mental task classification. In: 31st Annual Int. Conference of the IEEE EMBS, Minnesota, pp. 2579–2582 (2009)
6. Fisher, A.R.: The use of multiple measurements in taxonomic problems. Ann. Eugen 7, 179–188 (1936)
7. Freeman, W.J.: Comparison of brain models for active vs. passive perception. Information Science 116, 97–107 (1999)
8. Garrett, D., Peterson, D.A., Anderson, C.W., Thaut, M.H.: Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Trans. Neural Syst. Rehabil. Eng. 11(2), 141–144 (2003)

9. Graimann, B., Huggins, J.E., Schlogl, A., Levine, S.P., Pfurtscheller, G.: Detection of movement-related desynchronization patterns in ongoing single-channel electrocardiogram. IEEE Trans. Neural Syst. Rehabil. Eng. 11(3), 276–281 (2003)
10. Guyon, I., Elisseeff, A.: An Introduction to Variable and feature Selection. Machine Learning Research (3), 1157–1182 (2003)
11. Guyon, I., Weston, J., Bernhill, S., Vapnik, V.: Gene Selection for cancer classification using support vector machine. Machine Learning (46), 389–422 (2002)
12. Huang, N.E., Shen, Z., Long, S.R., Wu, M.L., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H.: The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis. Proc. R. Soc. London A 454, 903–995 (1998)
13. 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. Rehalil. Eng. 14(2), 190–193 (2006)
14. Kohavi, R., John, G.: Wrapper for feature subset selection. Artificial Intelligence (1-2), 273–324 (1997)
15. Lempel, A., Ziv, J.: On the complexity of finite sequences. IEEE Trans. Inform. Theory IT-22, 75–81 (1976)
16. Lingras, P., Butz, C.: Rough set based 1-v-1 and 1-v-r a roaches to support vector machine multi-classification. Information Science 177, 3782–3798 (2007)
17. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B.: A review of classification algorithms for EEG-based brain computer interfaces. Neural Eng. 4, R1–R13 (2007)
18. Muller, K.-R., Anderson, C.W., Birch, G.E.: Linear and non-linear methods for brain computer interfaces. IEEE Trans. Neural Syst. Rehabil. Eng. 11(2), 165–169 (2003)
19. Park, H.-S., Yoo, S.-H.-Y., Cho, S.-B.: Forward selection Method with regression analysis for optimal gene selection in cancer classification. International Journal of Computer Mathematics 84(5), 653–668 (2007)
20. 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)
21. Pierre, A.D., Kittler, J.: Pattern Recognition: A Statistical Approach. PHI (1982)
22. Ruiz, R., Riquelme, J.C., Ruiz, S.A.: Incremental wrapper based gene selection from microarray data for cancer classification. Pattern Recognition 39(12), 2383–2392 (2006)
23. Shannon, C.E.: A Mathematical Theory of Communication. ATT Tech. J. 27, 379–423, 623–656 (1948)
24. Tibshiran, R., Hastie, T., Narasimha, B., Chu, G.: Diagnosis of multiple cancer types by shrunken centriods of gene expression. Proc. Natl. Acad. Sci., USA (99), 6567–6572 (2002)
25. Tsai, C.-Y.: On detecting nonlinear patterns in discriminate problems. Information Science 176, 772–798 (2006)
26. Tsoi, A.C., So, D.S.C., Sergejew, A.: Classification of electroencephalogram using artificial neural networks. In: Advances Neural Information Processing Systems, vol. 6, pp. 1151–1158. Morgan Kaufman, San Francisco (1994)
27. Yom-Tov, E., Inbar, G.F.: Feature selection for the classification of movements from single movement-related potentials. IEEE Trans. Neural Syst. Rehabil. Eng. 10(3), 170–177 (2002)

28. Zachary, A.K., Jorge, I.A.: A new mode of communication between man and his surroundings. IEEE Transactions on Biomedical Engineering 37(12), 1209–1214 (1990)
29. Zachary, A.K.: Alternative modes of communication between man and machine. Master's thesis, Purdue University (1988)
30. Zhang, X.S., Roy, R.J., Jensen, E.W.: EEG complexity as a measure of depth anesthesia for patients. IEEE Trans. Biomed. Eng. (48), 1424–1433 (2001)
31. Zhou, S.M., Gan, J.Q.: Constructing parsimonious fuzzy classifiers based on L2-SVM in high-dimensional space with automatic model selection and fuzzy rule ranking. IEEE Trans. Fuzzy Syst. 15(3), 398–409 (2007)