

# EEG Signals of Motor Imagery Classification Using Adaptive Neuro-Fuzzy Inference System

Shereen A. El-aal, Rabie A. Ramadan and Neveen I. Ghali

**Abstract** Brain Computer Interface (BCI) techniques are used to help disabled people to translate brain signals to control commands imitating specific human thinking based on Electroencephalography (EEG) signal processing. This paper tries to accurately classify motor imagery imagination tasks: e.g. left and right hand movement using three different methods which are: (1) Adaptive Neuro Fuzzy Inference System (ANFIS), (2) Linear Discriminant Analysis (LDA) and (3) k-nearest neighbor (KNN) classifiers. With ANFIS, different clustering methods are utilized which are Subtractive, Fuzzy C-Mean (FCM) and K-means. These clustering methods are examined and compared in terms of their accuracy. Three features are studied in this paper including AR coefficients, Band Power Frequency, and Common Spatial pattern (CSP). The classification accuracies with two optimal channels C3 and C4 are investigated.

**Keywords** Brain Computer Interface · Classification · Adaptive neuro fuzzy inference system

## 1 Introduction

Brain Computer Interface (BCI) provides new communication and control technology to people who suffer from disabilities and neuromuscular disorders. It enables disabled persons to operate, for example, word processing programs and control devices in some special circumstances. Signal processing algorithms are used to translate brain signals to control commands without using muscles. Non-invasive EEG reads electrical signals using wired or wireless BCI [1] from cortical surface

---

S.A. El-aal (✉) · N.I. Ghali  
Faculty of Science, Alazhar University, Cairo, Egypt  
e-mail: shereen\_mhmd86@yahoo.com

R.A. Ramadan  
Computer Engineering Department, Cairo University, Cairo, Egypt  
Hail University, Hail, Kingdom of Saudi Arabia

through the scalp. Then, it describes band powers in terms of rhythmic brain waves: 8–13 Hz for Mu rhythm, 14–30 Hz for Beta rhythm.

Different classifiers are utilized for EEG classification including LDA [2], Support Vector Machine (SVM) [3], Fuzzy SVM (FSVM) [4], KNN [5] and ANFIS [6].

This paper is organized as follows: Sect. 2 introduces BCI categories, some of researches done in BCI signal classification and feature extraction, EEG features, different classifiers and different clustering methods used. Section 3 defines the problem. Section 4 surveys BCI processing phases and the methods used in every phase. Section 5 assess and compare classification accuracy using different classifiers and different features and finally Sect. 6 gives the conclusion.

## 2 Related Work

BCI is a communication system that allows disabled users to control computer application without a need to peripheral muscle and nerves, just by thoughts.

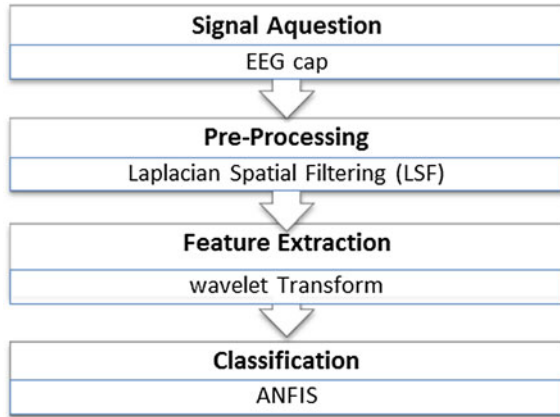
In all cases, brain waves are categorized to five categories: Delta, Theta, Alpha, Beta and Gamma [7]. They are described by their frequency where Delta waves falls in the range 0.5–4 Hz, Theta is in the range 4–7 Hz, Alpha is within the range 8–13 Hz, Beta is in the range 14–30 Hz and Gamma is supposed to be in the above 30 Hz range. Alpha and Beta are the two most prominent frequency bands used in motor Imagery task. Alpha peaks around 10 Hz and it appears in mediation and relaxation state. Beta rhythm is associated with conscious and focuses states as well as it occurs in motor task movements.

Due to the importance of the extracted data accuracy, different steps have introduced for this reason. In the signal preprocessing step filter techniques are used to convert recorded signals to a special form in order to be easier in categorization. Some of the used techniques are Laplacian Spatial Filtering (LSF) [8] and Kalman filter [9].

In feature extraction step, suitable features with dimension reducing are used in order to gets good classification accuracy. The most prominent feature for imagery task is band power frequency. One of the issues of EEG signals is its nature is non-stationary signals; that mean its frequency changes over a time. Classical Fourier transform is not appropriate for such EEG signals because it produces only frequency component variable. In addition, Discrete Wavelet Transform (DWT) [4] was able to decompose the EEG signal into different level of resolution (frequency bands) using wavelet functions.

In classification step, several methods are utilized for BCI classification in which they are employing to predict class label for EEG signals patterns. In addition, supervised linear classifiers are commonly used especially, LDA [2] and Support Vector Machine (SVM) [3]. It depends on linear functions to discriminate classes. SVM is good when using features with high dimension, however it is sensitive to noise or outliers. So Fuzzy SVM (FSVM) [4] is used which promotes the effect of noise. In addition, it employs fuzzy membership for each training sample which is used for

**Fig. 1** Brain signal processing



demonstrating decision. ANFIS is a hybridization method in which the system is trained using back propagation learning method.

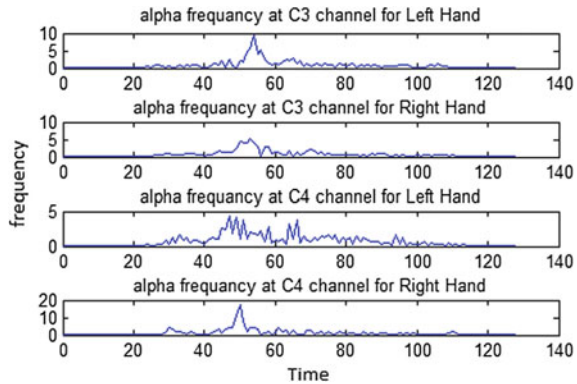
### 3 Problem Definition

BCI becomes a new challenge in many of the recent applications. One of these applications is helping disabled users to move just by thinking of movement. The process requires brain signals related to motor imagery movement to be extracted and classified especially, hand movement. The purpose is to sense if disabled person able to move his/her hands through imagination. Consequently, the subject can use the same system for playing video games such as games depends on joystick movements. Certainly, this will help people who suffer from Amyotrophic lateral sclerosis (ALS) or spinal injuries. The purpose of this paper is to introduce a complete framework for the brain signal processing including different modules for accurate classification to EEG signals. The framework utilizes the Laplacian Spatial Filtering (LSF), wavelet transform, and the ANFIS for accurate classification. In addition, we tend to report the best EEG signal feature for imagery tasks based on our experiments.

### 4 EEG Proposed Classification and Clustering Algorithms

Before going deeply into the utilized techniques and algorithms, Fig. 1 elaborates on our approach in extracting accurate EEG signals. As can be seen in the figure, our proposed approach consists of four major phases which are signal acquisition, signal preprocessing, feature extraction, and classification.

**Fig. 2** Alpha in C3 and C4 channels for *left* and *right* hand



In the signal acquisition phase, the EEG signals are collected using very sensitive sensors on the scalp. The second phase is the preprocessing signals in which the data noise is filtered using specific filters such as: Kalman [9] and LSF [8] using high pass filter. In addition, channel selection is used in preprocessing.

In feature extraction phase, suitable features with dimension reducing are extracted to show differences between various classes. Band power feature, especially Alpha and Beta, is one of the most important features for motor imagery classification [8]. Alpha also known as mu rhythm and it is increased in cortex comparable to the arm used for imagination. In addition, Beta power linked to motor movement and appears in anxious thinking and active concentration. Wavelet Transform (WT) [4] is the most prominent method for extracting band power features because it considers the frequency of the signal in different band power. Four WT coefficients can be extracted which are D1:D4 and final approximation A4. WT coefficients are extracted using 'db8' wavelet function.

After analysis, as we will see later, we discovered that, if the user imagines moving left hand, the signal frequency in channel C3 becomes quite stronger than the signal frequency in channel C4 as shown in Fig. 2. In addition, if the user imagine moves his/her right hand, the signal frequency in channel C3 becomes quite weaker than signal frequency in channel C4 while channel C3 located in left side and C4 in right side.

The last phase is the classification in which the classifier is used to specify the brain signals for each class. In this research, ANFIS is the essential classifier used for data categorization. Moreover, LDA and KNN are used for comparing results. Performance of classifiers is measured by classification accuracy in which it is presented by the percentage of the number of correctly classified test patterns.

### 4.1 Autoregressive (AR) Model

AR model method [10] is used to extract linear feature for EEG signals which it describes the signal characteristics. AR model of order  $p$  is given by:

$$x[n] = - \sum_{k=1}^p (a_k \cdot x[n-k] + e[n]) \quad (1)$$

where  $p$  is the model order,  $e[n]$  represents white noise error and  $x[n]$  is signal data at point  $n$ . If EEG data is recorded using multiple channels, AR coefficients are produced for every channel and then combined to be a feature vector. Consequently, if the brain signals are recorded using C3 and C4 channels, AR coefficients for these signals are defined by  $a_{c3}[n]$ ,  $a_{c4}[n]$  respectively.

### 4.2 Common Spatial Pattern(CSP)

CSP [11] is a spatial filter method for preprocessing EEG signals. The main idea is to decompose brain signals into spatial patterns. Signal decomposition aims to minimize similarity between both classes and maximize differences by assigning the signal into patterns. This spatial pattern can be determined using covariance matrices conjoint diagonal of EEG signals for both classes. If the brain signals have length  $N$  and recorded through  $n$  electrodes, the dimension for covariance matrix will be  $n \times N$ .

CSP only suits two classes of discrimination because it depends on Fisher Discriminant Analysis (FDA) criteria. So CSP is modified to Bisearch CSP (BCSP) to discriminate four classes [12]. Details of CSP Equations are illustrated in [11].

### 4.3 Adaptive Neuro-Fuzzy Inference System (ANFIS) Classifier

Fuzzy inference system is a model that maps input characteristic to output decision. The shape of input/output membership function changes according to the input/ output parameters (fuzzy sets and rules); so it is easy to model fuzzy inference system. However, occasionally it is hard to extract parameters to tailor input/output membership functions when looking at data. Therefore, it will be appropriate if learning algorithm such as NN is used along with the fuzzy logic.

ANFIS [13] is a Sugeno model implementation for fuzzy inference in which it has the same view of NN with multiple hidden layers in between. ANFIS has five layers [14] as shown in Fig. 3. Details for each layer are explained in [13].

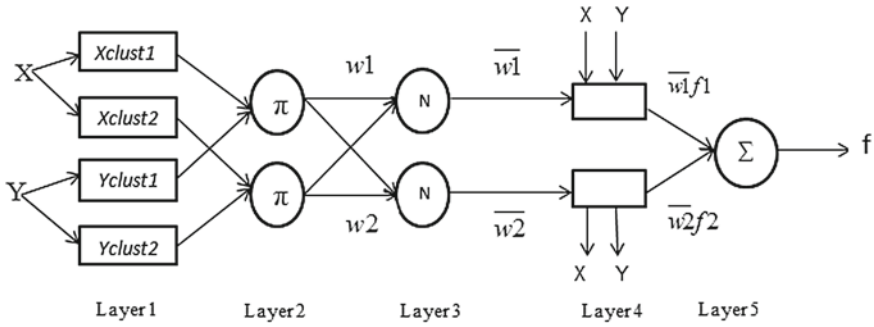


Fig. 3 ANFIS architecture for Sugeno method

Clustering process partitions the data into number of groups to facilitate data analysis where data in the same group have similar properties.

In the following subsection, different clustering techniques are explored including Subtractive, FCM and K-means.

### 4.3.1 Subtractive Clustering

Subtractive clustering [6, 15] fits under ANFIS technique. It defines cluster center based on the density around the data points. Candidates' data points are considered as cluster center then density of data point  $x_i$ , for instance, is computed by:

$$D_i = \sum_{j=1}^n \exp\left(-\frac{\|x_j - x_i\|^2}{(r_a/2)^2}\right) \tag{2}$$

where  $r_a$  is a positive constant specifying neighborhood for every cluster. After measuring all data points density, the largest density is selected as first cluster center. Assume the density for first data points  $x_{c1}$  is  $D_{c1}$ , density measure for data point  $x_i$  is revised by:

$$D_i = D_i - D_{c1} \exp\left(-\frac{\|x_i - x_{c1}\|^2}{(r_b/2)^2}\right) \tag{3}$$

where  $r_b$  is a positive constant determine neighborhood that has measurable reductions in density measure, generally  $r_b = 1.5r_a$ .

The subtractive clustering selects the highest density as the next cluster center after revising. Then, it continues until reaching sufficient number of clusters.

### 4.3.2 K-Means Clustering

One of data clustering algorithm is the K-means. K-means clustering [16] partitions data collection of  $n$  vectors  $x_j, j = 1, 2, \dots, n$  to a certain number of groups  $G_i, i = 1, 2, \dots, c$  and it defines  $k$  centroids one for each cluster.  $k$  centroids change their location step by step until they reach suitable locations. It is based on minimizing the cost (objection) function of dissimilarity using Euclidean distance to measure distance between data point  $x_k$  and cluster center  $c_i$ . Cost function is given by:

$$J = \sum_{i=1}^c J_i = \sum_{i=1}^c \left( \sum_{k, x_k \in G_i} \|x_k - c_i\|^2 \right) \tag{4}$$

where  $J_i$  is the cost function for cluster  $i$ . Details of K-means clustering algorithm equations and steps are illustrated in [16].

### 4.3.3 Fuzzy C-Mean Clustering

(FCM) [16] is also classified under clustering algorithm. FCM is usually used in pattern recognition applications. In FCM, each data point belongs to more than one cluster with membership grades between 0 and 1. FCM objective function is represented by:

$$J_m = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m \leq \infty \tag{5}$$

where  $m$  is the weighting exponent and  $U_{ij}$  is the membership grade for data point  $x_i$  in the cluster  $j$  and  $c_j$  is the centroid for the cluster  $j$ . Details of FCM clustering algorithm equations and steps are illustrated in [16].

## 4.4 K-Nearest Neighbor (KNN) Classifier

KNN [3] is simple machine learning method based on the close neighbors with the same properties may belong to the same class. Each data point is assigned to a class according to the majority voting of its neighbors where KNN specifies  $k$  neighbors training points for every test data point using metrics as Euclidean, Manhattan, Mahalanobis distances. KNN calculates the distance for all training data vector to each test data vector then it selects the closest  $k$  neighbor. Therefore, the output value is the average of  $k$  nearest neighbors' values.

KNN is a simple classifier but it is known as lazy learner because it uses training data for classification. It takes the most common label in  $k$  training neighbors to be predicted label for new test data point. In addition, changes in the  $k$  closest nearest neighbor will change the output class.

## 4.5 Linear Discriminant Analysis (LDA) classifier

LDA [2] is a supervised linear classifier algorithm is used for dimensionality reduction. It is based on finding linear discrimination function to maximize multiple class separation by representing axes component. LDA Specifies data point to a class according to decision hyperplane function  $g(x)$  given by:

$$g(x) = w^T x + w_0 \quad (6)$$

where  $w$  is a weight vector for combining features,  $x$  is the feature vector and  $w_0$  called Bias that determines the position of decision surface. Consider we have two classes  $C1$  and  $C2$ , then  $x \in C1$  if  $g(x) \geq 0$  or  $x \in C2$  otherwise.

## 5 Experimental Results

In this section, data set used for testing our approach is described and achieved results are analyzed according to classifiers performance and features used.

### 5.1 Graz Dataset

The used EEG data set in this paper is provided by Medical Informatics department, University of Technology at Graz. Female subjects with 25 years old were recorded imagery left and right hand movement dataset during sessions. Each session have 7 runs with 40 trials of 9 seconds length i.e. (280 trials). The subjects sat in a relaxing chair putting their arms on armrests. First two seconds was quite; at  $t = 2$  an acoustic stimulus refers to the starting of the trial and cross '+' displayed on the screen for only one second. At  $t = 3$  an arrow (left or right) was displayed for 1s, then subject asked to imagine left or right hand. The recording feedback was based on C3, Cz and C4 channels.

### 5.2 Classification with Varying Signal Intervals

In this subsection, classification is performed for hand imagination movement using C3 and C4 channels. Subsequently LSF is used to filter the EEG signals. Thereafter, WT is used to produce band power feature. In the present research, data set is digitized at 128 Hz; so fourth level of decomposition is chosen which is based on D1:D4 and A4. Moreover, the results are obtained by applying ANFIS, KNN and LDA using wavelet based features, AR coefficients and CSP.

In this experiment, the subject was idle in the first 3 seconds; so there is no enough information for classification. Therefore, intervals used for classification are from  $t = 4$  to  $t = 8$ . Furthermore, these intervals are used in order to reduce the feature



**Table 1** Classification accuracy using different signal intervals

ANFIS clustering methods					
Signal length (%)	Subtractive (%)	FCM (%)	K-means (%)	KNN (%)	LDA (%)
600	95	90	90	87.5	90
500	82.5	80	80	70	80
300	72.5	70	70	75	75
200	92.5	92.5	92.5	82.5	85
Average	85.6	83.1	83.1	78.8	82.5

space to utilize signal information for classification. Every second is represented by 128 samples (signal length); so intervals used in classification have 640 samples. In this experiment, band power feature is extracted using WT and the classification was done with varying feature space. With hypothesis, subjects may lose their focus for moments so that affects acquired signals accuracy. Therefore, different intervals are selected randomly to add more classifiable information; then, the average will be taken after all.

Table 1 demonstrates ANFIS, LDA and KNN classifiers performance using band power feature with different signal intervals. It shows that, LDA accuracy is within the range from 75 % to 90 % with average 82.5 %. Likewise, ANFIS accuracy along different clustering algorithms varies from 70 % to 95 % and average about 85.6 %. As shown in Table 1, ANFIS and LDA classifiers achieved satisfactory results where, ANFIS produced the best result for subtractive method.

### 5.3 Classification for Every Second Interval

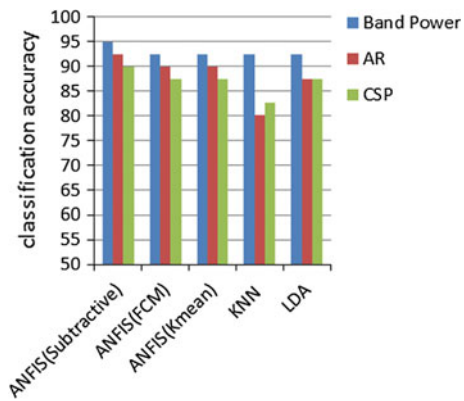
In this experiment, classification was done using band power feature while signal intervals are represented at every second from  $t = 4$  to  $t = 8$ . Furthermore, 4 seconds was used as one interval. Assuming that, signals may involve coincide with more thinking process so the next second interval may have beneficial information for classification. Therefore, the classification process is tested at every second interval; then the average is compared with interval  $t = 4$  to  $t = 8$ .

Table 2 demonstrates the accuracy for ANFIS, KNN and LDA classifiers. It shows that, the results at every second interval as well as the average. It seems that the average do not provide satisfactory result as shown in highlighted row. Consequently, all classifiers accuracy varies from 67 to 76 %. On the other hand, signal interval form  $t = 4$  to  $t = 8$  provides much better results for all classifiers. In particular, ANFIS using subtractive clustering algorithm is giving the best accuracy in which it achieves accuracy of 95 % compared to other researchers using the same data set [5].

**Table 2** Classification accuracy for every second

ANFIS clustering methods					
Time sample (%)	Subtractive (%)	FCM (%)	K-means (%)	KNN (%)	LDA (%)
T = 4	75	75	75	60	62.5
T = 5	80	82.5	82.5	77.5	77.5
T = 6	85	85	85	87.5	82.5
T = 7	72.5	75	75	57.5	67.5
T = 8	67.5	62.5	62.5	55	72.5
Average	76	76	76	67.5	72.5
T = 4:8	95	92.5	92.5	92.5	92.5

**Fig. 4** Classification accuracy using band power AR and CSP features



### 5.4 Classification Accuracy Using Different Features

In this set of experiments, classifiers performance is evaluated using band power, AR coefficients and CSP features.

**Band Power:** Alpha and Beta are the most bands frequency that modulate the EEG signals for imagination activity. Classification accuracy using band power feature is demonstrated in Table 2 and Fig. 4 using feature space from  $t = 4$  to  $t = 8$  as one interval. As can be noticed, LDA, KNN and ANFIS classifiers provide the same results in which the produced accuracy is equal to 92.5 % while the subtractive method achieves the peak accuracy equal to 95 %.

**AR Coefficients:** As can be seen in the Fig. 4, KNN classifier does not perform well using AR in which it achieves only 80 %. LDA, FCM and K-means produce good results where LDA has 87.5 % accuracy while FCM and K-means gives accuracy of 90 %. However, the best accuracy is provided by ANFIS using subtractive clustering in which its average accuracy is 92.5 %.

**CSP:** CSP is the most popular algorithm in BCI field for learning spatial filters for oscillatory processes using frequency band and time window. CSP classification results for different classifiers are shown in the Fig. 4. Here, almost all of the algorithms are giving the same results including ANFIS using FCM, K-means and LDA. Their results approach 87.5 % of classification accuracy. However, it seems that KNN did not reach the same accuracy; therefore, it is not recommended to be used with CSP.

## 6 Conclusion

This paper has presented various EEG features and BCI classifiers using motor imagery dataset. This dataset was proposed to discriminate left and right hand imagination movement. WT based features was used in which it decomposes the signals to different levels. The number of decomposition level is based on EEG signal frequency. In addition, AR coefficients and CSP features were used.

FCM and K-means clustering algorithms provide stable classification result when the number of clusters is only two. However they produce different results for more than two clusters. This is because k-means chooses centroid for each cluster randomly and FCM depends on membership grade initialization for every cluster. Subtractive clustering algorithm generates the numbers of clusters according to provided cluster radius. Subsequently, it calculates cluster centroid density based on surrounding data points. Subtractive clustering produces stable result whatever the numbers of clusters.

The paper investigated the differences of three classifiers (ANFIS, LDA, and KNN) and achieved classification accuracies were reported. ANFIS classifier implementation depends on using clustering algorithms (subtractive, FCM and K-means). Clustering algorithms are used to tailor input data to output rules. The maximum classification accuracy is obtained using ANFIS classifier. In particular, using subtractive clustering algorithm in which it is provided 95 % for band power and 92.5 % for AR coefficients features. Furthermore, FCM and K-means are almost provided the same results.

## References

1. Lin, C.T., Ko, L.W., et al.: Wearable and wireless brain-computer interface and its applications. *Found. Augment. Cognit.* **5638**, 741–748 (2009)
2. Lotte, F., Congedo, M., et al.: A review of classification algorithms for EEG-based brain-computer interfaces. *Neural Eng.* **4**, R1–R13 (2007)
3. Dokare, I., Kant, N.: Performance analysis of SVM, KNN and BPNN classifiers for motor imagery. *Eng. Trends Technol.* **10**, 19–23 (2014)
4. Xu, Q., Zhou, H., et al.: Fuzzy support vector machine for classification of EEG signals using wavelet-based features. *Med. Eng. Phys.* **31**, 858–865 (2009)

5. Yang, R., Gray, D. *et al*: Comparative analysis of signal processing in brain computer interface. *IEEE Ind. Electron. Appl.* 580–585 (2009)
6. Tawafan, A., Sulaiman, M., et al.: Adaptive neural subtractive clustering fuzzy inference system for the detection of high impedance fault on distribution power system. *AI* **1**, 63–72 (2012)
7. Larsen, E.A.: Classification of EEG Signals in a Brain Computer Interface System. University of Science and Technology, Norwegians (2011)
8. Qin, L., Ding, L., et al.: Motor imagery classification by means of source analysis for brain computer interface applications. *Neural Eng.* **1**, 144–153 (2004)
9. Aznan, NKN., Yang, YM.: Applying kalman filter in eeg-based brain computer interface for motor imagery classification. *ICT Converg.* 688–690 (2013)
10. Yang, R.: Signal Processing for a Brain Computer Interface. School of Electrical and Electronic Engineering, University of Adelaide, Adelaide (2009)
11. Gutierrez, D., Salazar, R.: EEG signal classification using time-varying autoregressive models and common spatial patterns. *IEEE EMBS*, 6585–6588 (2011)
12. Fang, Y., Chen, M., et al.: Extending CSP to detect motor imagery in a four-class BCI. *Inf. Comput. Sci.* **9**, 143–151 (2012)
13. Guler, I., Ubeyli, E.: Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. *Neurosci. Methods* **148**, 113–121 (2005)
14. bin Othman, MF., Shan Yau, TM.: Neuro fuzzy classification and detection technique for bioinformatics problems. *Model. Simul.* 375–380 (2007)
15. Priyono, A., Ridwan, M., et al.: Generation Of fuzzy rules with subtractive clustering. *Teknologi* **43**, 143–153 (2005)
16. Ghosh, S., Dubey, S.K.: Comparative analysis of K-means and fuzzy C-means algorithms. *J. Adv. Comput. Sci. Appl.* **4**, 34–39 (2013)