Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2013) 21: 2092 – 2109
© TÜBİTAK
doi:10.3906/elk-1203-9

Novel approaches for automated epileptic diagnosis using FCBF selection and classification algorithms

Baha ŞEN,[1,*] **Musa PEKER**[2]

[1]Department of computer Engineering, Faculty of Engineering, Yıldırım Beyazıt University, Ulus, Ankara, Turkey

[2]Department of Computer Engineering, Faculty of Engineering, Karabük University, Karabük, Turkey

Received: 09.03.2012 • **Accepted:** 05.06.2012 • **Published Online:** 24.10.2013 • **Printed:** 18.11.2013

Abstract: This paper presents a new application for automated epileptic detection using the fast correlation-based feature (FCBF) selection and classification algorithms. This study consists of 3 stages: feature extraction, feature selection from electroencephalography (EEG) signals, and the classification of these signals. In the feature extraction phase, 16 attribute algorithms are used in 5 categories, and 36 feature parameters are obtained from these algorithms. In the feature selection phase, the FCBF algorithm is chosen to select a set of attributes that best represent the EEG signals. The resulting attributes are used as input parameters for the classification algorithms. In the classification phase, the problem is classified with 6 different classification algorithms. The results obtained with the different classification algorithms are provided in order to compare the calculation times and the accuracy rates. The evolution of the proposed system is conducted using k-fold cross-validation, classification accuracy, sensitivity and specificity values, and a confusion matrix. The proposed approach enables 100% classification accuracy with the use of the multilayer perceptron neural network and naive Bayes algorithm. The stated results show that the proposed method is capable of designing a new intelligent assistance diagnostic system.

Key words: EEG signals, classification algorithms, FCBF selection algorithm, epilepsy disease

1. Introduction

Epilepsy can be seen in all ages, races, social classes, and countries. It is estimated that the ratio of the general population with active epilepsy (i.e. continuous seizures or the need for treatment) is 4 to 10 per 1000 people. However, in developing countries, this ratio can reach as high as 6 to 10 per 1000 people, as some studies suggest [1].

Electroencephalography (EEG) signals are generally used in the diagnosis of epilepsy. EEG signals record the electrical activity along the scalp with the help of electrodes and reflect the collective activity of brain cells. Generally, EEG comprises 4 main wave types. These are alpha waves (8–13 Hz), beta waves (13–30 Hz), delta waves (0–4 Hz), and theta waves (4–7 Hz) [2]. In EEG signals, these wave types show changes in pathological situations. It may be difficult to identify the components that show momentary changes in the EEG signal in pathological situations such as epilepsy seizures [2]. Researchers have stated that it is difficult to identify the epilepsy attacks of some patients, that the patient has to be monitored asleep and awake, and that their EEG

*Correspondence: bsen@ybu.edu.tr

records should be examined [3]. EEG data are analyzed visually by specialists. This is a stressful and exhausting process for the eye and for the mind. The EEG data recorded for hours, or even days, are analyzed from the computer screen in the form of the display of 10-s frames [2]. The EEG records interpreted by specialists with different types of training may create an inconsistent recording of the information obtained through the viewing [2]. Hence, it is imperative to analyze the EEG signals using a consistent and appropriate processing method in order to obtain correct diagnoses for the treatment of epilepsy. Recently, various models have been proposed for the identification of epileptic activity. Feature extraction and classification algorithms commonly used in the literature are examined. Algorithms such as standard deviation (SD) [4–6], median (MN) [5,6], arithmetic mean (AM) [5–7], zero-crossing (ZC) value [4,5], variance (V) value [5,8], values of maximum (MaxV) and minimum (MinV) [5], mean energy [5,9,10], mean Teager energy (MTE) [9], Petrosian fractal dimension (PFD) [10], Rényi entropy (RE) [10,11], spectral entropy (SpEn) [9–11], Wigner-Ville (WV) coefficients [8], wavelet transform [7,10,12], mean curve length (MCL) [9,10], and the Hjorth parameters [10,12] are commonly used for extracting features from EEG data. Various different approaches have been proposed in the literature as classification algorithms, such as the artificial neural network (ANN) [5–7,12–15], adaptive neuro fuzzy inference system (ANFIS) [11], decision tree [16], naive Bayes (NB) [17], logistic regression [18] and K-means clustering (KMC) algorithm [19].

The purpose of this study is to elicit classifiable information from human EEGs and to identify the algorithm that provides the highest classification accuracy. One of the important problems in biomedical signal processing is to identify the features that represent the data. There are various algorithms in the literature that are used for the identification of features that represent data; however, most of the time, the trial-and-error method is used to determine which algorithms identify the effective features for the specified problem. A system that can automatically undertake this process will facilitate the work of researchers to a great extent. This study proposes a method to best represent the data regarding the diagnosis of epilepsy. Previous studies are investigated with this aim and 36 features are extracted that have previously been used in the identification of epilepsy. The fast correlation-based feature (FCBF) algorithm is used to determine the features that produce the highest effect. The next phase aims to identify the algorithm that provides the highest level of classification accuracy with the selected features. The study aims to identify the best model that automatically classifies epileptic EEG signals.

2. Data source

The data used is publicly available [20]. In this section, only a short description is provided and further details are provided later on [20]. The database contains 5 different EEG data sets, entitled A, B, C, D, and E. Data set A consists of EEG data obtained from healthy individuals recorded with surface electrodes while their eyes were open. Data set B consists of EEG data obtained from healthy individuals recorded with surface electrodes while their eyes were closed. Data sets C and D carry EEG data recorded from ill patients while no seizures were present. Data set E consists of EEG data recorded from ill patients during seizures.

All of the EEG signals were recorded with the same 128-channel amplifier system, using an average common reference. The data were digitized at 173.61 samples per second using 12-bit resolution. The band-pass filter settings were 0.53–40 Hz (12 dB/oct). In this study, 2 data sets (A and E) were used and the typical EEGs are depicted in Figure 1.

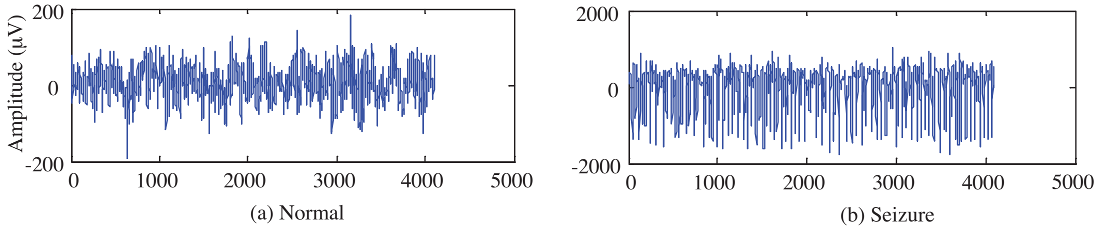

Figure 1. Sample EEG segments from sets A and E.

3. Method

3.1. Feature extraction

Initially, 5 categories, *statistical*, *nonlinear*, *energy*, *time frequency*, *and entropy*, are identified in the feature extraction phase. In these 5 categories, 16 attribute algorithms are used, and 36 feature parameters are obtained from these algorithms. These feature values are obtained with the help of software developed by the MATLAB programming language. The obtained features are presented in Table 1.

3.1.1. Statistical features

In this phase, the statistical attributes of the EEG signals are obtained. Short explanations of the attributes are provided in Table 2.

Table 1. Attribute table.

N	Feature	N	Feature	N	Feature	N	Feature	N	Feature
1	AM	9	MTE	17	D3-1	25	D5-1	33	MCL
2	MaxV	10	PFD	18	D3-2	26	D5-2	34	Hjorth activity
3	MinV	11	RE	19	D3-3	27	D5-3	35	Hjorth mobility
4	SD	12	SpEn	20	D3-4	28	D5-4	36	Hjorth complexity
5	V	13	WV coefficient 1	21	D4-1	29	A5-1		
6	MN	14	WV coefficient 2	22	D4-2	30	A5-2		
7	ZCs	15	WV coefficient 3	23	D4-3	31	A5-3		
8	Mean energy	16	WV coefficient 4	24	D4-4	32	A5-4		

Table 2. Short explanations of the statistical attributes.

Feature name		Formula		Explanation
MinV	:	$MinV = min[x_n]$	(1)	Where $x_n = 1, 2, 3 \ldots n$ is a time series, N is the number of data points, AM is the mean of the sample.
MaxV	:	$MaxV = max[x_n]$	(2)	
SD	:	$SD = \sqrt{\frac{\sum_{n=1}^{N}(x_n - AM)^2}{N-1}}$	(3)	
AM	:	$AM = \frac{1}{N}\sum_{n=1}^{N} x_n$	(4)	
V	:	$V = \frac{\sum_{n=1}^{N}(x_n - AM)^2}{N-1}$	(5)	
		$MN = \left(\frac{N+1}{2}\right)^{th}$	(6)	If the number of values is odd, then
MN	:	$MN = \frac{\left(\frac{N}{2}\right)^{th} value + \left(\frac{N}{2}+1\right)^{th} value}{2}$	(7)	If number of values is even, then (where N = number of items)

3.1.2. Nonlinear-based features

PFD: Petrosian's algorithm is one of the simplest and fastest methods for the estimation of fractal dimension [21]. It is based on the change of a sign in the signal's derivation. However, in discrete signals, the derivation can be defined as the subtraction of 2 consecutive signals [21]. It can be estimated by the following expression:

$$PFD = \frac{\log_{10} n}{\log_{10} n + \log_{10}\left(\frac{n}{n+0.4N_\delta}\right)}, \tag{8}$$

where n is the number of signal samples and N_δ is the number of sign changes in the signal derivative.

MTE: The MTE, first proposed in [22], is defined as:

$$MTE[n] = \frac{1}{N}\sum_{m=n-N+3}^{n}\left(x\,[m-1]^2 - x\,[m]\,x\,[m-2]\right), \tag{9}$$

where $x[m]$ is an EEG time series, N is the window length, and n is the last sample in the epoch.

MCL: The MCL [23] is an estimate of Katz's fractal dimension [24] and is used as an effective measure for seizure detection from the EEG. The MCL is defined as:

$$CL\,[n] = \frac{1}{N}\sum_{m=n-N+1}^{n}\left|x\,[m] - x[m-1]\right|, \tag{10}$$

where $x[m]$ is an EEG time series, N is the window length, and n is the last sample in the epoch.

3.1.3. Time frequency-based features

WV distribution: A WV distribution is a distribution that displays the time and frequency information on the same plane. The WV can be successfully used in many applications where the signal used is nonstationary [25]. It is calculated as shown in Eq. (11):

$$WV\,(t,\omega) = \sum_{t=-\infty}^{\infty} f\left(t+\frac{t_0}{2}\right) f^*(t-\frac{t_0}{2})e^{-jt_0\omega}d\omega_0, \tag{11}$$

where t corresponds to the time of space, f^* represents the complex conjugate of f, and ω is the frequency. The features are deducted from the time-frequency plane obtained after the application of the WV transformation to the signal. The highest frequency values corresponding to the time values in the plane are used for this process. In this study, the function consists of the frequency values modeled by a third-order polynomial and the coefficients of this polynomial are used as the features [26]. Hence, 4 features (WV-1, WV-2, WV-3, and WV-4) consisting of the coefficients of the mentioned polynomial are obtained.

Number of ZCs: ZC is a term commonly used in electronics, mathematics, and image processing. The number of 'ZCs' in the EEG is thought to change during seizure activity [27]. The ZC is calculated by counting the number of times that the time domain signal crosses zero within a given window. ZC is defined as $(x_{n-1} < 0\,and\,x_n > 0)$ or $(x_{n-1} > 0\,and\,x_n < 0)$, or $(x_{n-1} \neq 0\,and\,x_n = 0)$.

Hjorth parameters: Three quantitative descriptors of the EEG are presented: the activity, mobility, and complexity [28]. The variance of the signal and its 1st and 2nd derivatives in a segment can be related to these parameters. If the variance of the original signal is σ_0, then the variance of the ith derivative can be defined as σ_i. The simulated and calculated Hjorth parameters are displayed in Table 3.

Table 3. Hjorth parameters.

Feature name	Activity	Mobility	Complexity
Formula	$HA = \sigma_0^2$	$HM = \sigma_1/\sigma_0$	$HC = \sqrt{(\sigma_2/\sigma_1)^2 - (\sigma_1/\sigma_0)^2}$

Discrete wavelet transform (DWT) and wavelet coefficients: DWT is a spectral analysis technique used in the analysis of nonstationary signals [7,29]. It provides time-frequency representations of the signals. DWT uses long time windows at low frequencies, whereas short time windows are used for high frequencies. This leads to good "time frequency" localization.

The DWT of a signal, $x(t)$, is the integral of the signal multiplied by the scaled and shifted versions of a wavelet function ψ and is defined by:

$$DWT(j, k) = \frac{1}{\sqrt{|2^j|}} \int_{-\infty}^{\infty} x(t)\,\psi\left(\frac{t - 2^j k}{2^j}\right) dt. \tag{12}$$

DWT analyzes the signal in different resolutions by separating it into approximate detail coefficients [30]. The sample outputs that belong to the first high-pass filter (g[.]) and the low-pass filter (h[.]) form the detailed D1 and approximate A1 subbands, respectively. The A1 approximate band separates again and the process is continued, as seen in Figure 2.

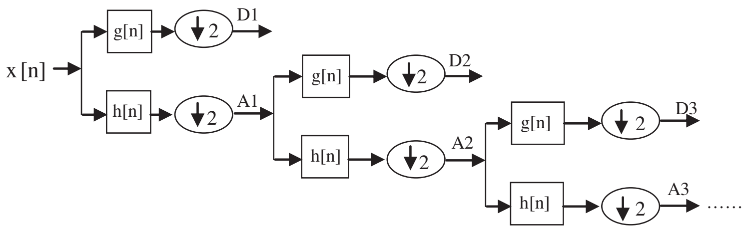

Figure 2. Subband decomposition of the DWT implementation, where h[n] is the high-pass filter and g[n] is the low-pass filter.

EEG signals are decomposed into subbands using DWT with the Daubechies wavelet of order 4 (db-4) because it yields good results in the classification of the EEG segments [7]. In this study, the EEG signals are separated until level 5 using the db-4 wavelet. At the end of the process, 5 detailed coefficient signals (D1–D5) and 1 approximate coefficient signal (A5) are extracted. These feature vectors, which are calculated for the frequency bands A5 and D3–D5, are used for the classification of the EEG signals. Figure 3 shows the approximation (A5) and details (D1–D5) of an epileptic EEG signal. Figure 4 shows the approximation (A5) and details (D1–D5) of a normal EEG signal.

A compact representation is provided by these extracted wavelet coefficients that shows the EEG signal's energy distribution in time and frequency. Further processes involve using statistics over the set of wavelet coefficients in order to decrease the dimensionality of the extracted feature vectors [7]. In order to represent the time-frequency distribution of the EEG signals, the following statistical features are employed:

- The mean of the absolute values of the coefficients in each subband (D3-1, D4-1, D5-1, and A5-1).
- The average power of the wavelet coefficients in each subband (D3-2, D4-2, D5-2, and A5-2).

- The SD of the coefficients in each subband (D3-3, D4-3, D5-3, and A5-3).
- The ratio of the absolute mean values of the adjacent subbands (D3-4, D4-4, D5-4, and A5-4).

Features 1 and 2 represent the signal's frequency distribution, whereas features 3 and 4 show the quantity of changes in the frequency distribution. In this manner, 16 features are obtained.

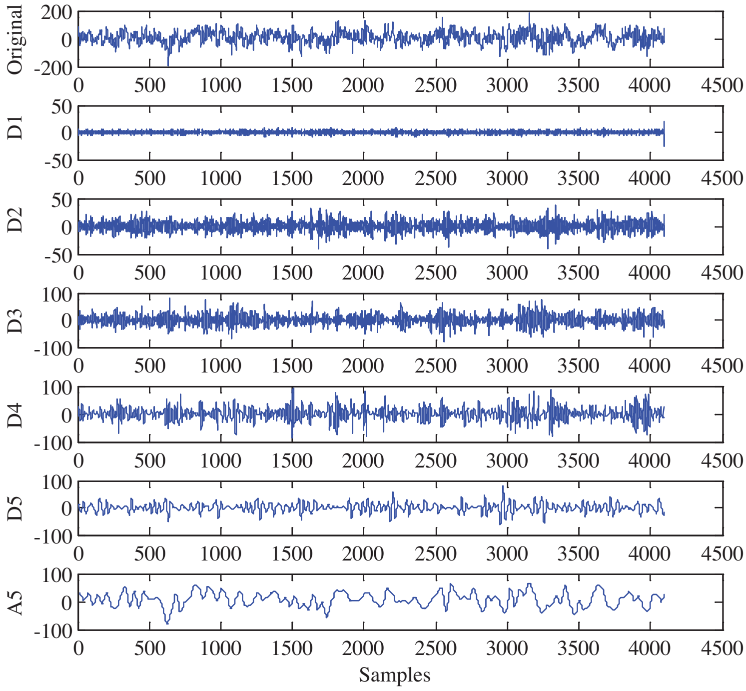

Figure 3. Approximate and detailed coeficients of the EEG signal taken from a healthy subject.

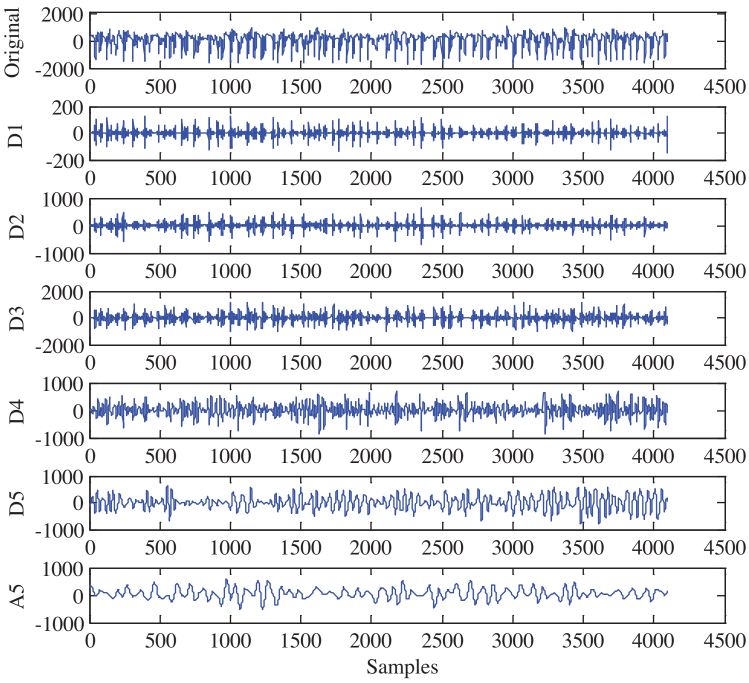

Figure 4. Approximate and detailed coefficients of the EEG signal taken from an unhealthy subject (epileptic patient).

3.1.4. Entropy-based features

SpEn: SpEn is a measure of the regularity of the signal. A pure sine wave has entropy of 0 and uncorrelated white noise has entropy of 1. In contrast to normal entropy, SpEn is calculated using the probabilities of the power spectra constituents of the signal. The data with a uniform probability distribution will have higher entropy [31].

SpEn is based on the power spectrum of EEG waves and describes the irregularity of the signal spectrum. The entropy of the power spectrum is denoted by H_{sp} and defined as:

$$H_{sp} = -\sum_{i=f_i}^{f_k} P_i \log(\frac{1}{P_i}), \tag{13}$$

where P is the power density over a defined frequency band of the signal, f_i and f_k are the lower and upper frequencies, and the power is normalized such that $\sum P_n = 1$ [31]. H_{sp} is also used in the normalized form as:

$$SpEn = \frac{H_{sp}}{\log N_f}, \tag{14}$$

where N_f is the number of frequencies within the defined band $[f_i, f_k]$. In this work, the frequency band is specified as [0, 50] Hz.

RE: RE is a special case of SpEn. The concept of generalized entropy of a probability distribution was introduced by Alfred Rényi [32]. The definition of the RE is shown in Eq. (15):

$$RE\,(X) = \frac{1}{1-\alpha} \sum_{k=i}^{n} \log P_k^{\alpha} \qquad (\alpha \neq 1\, and\, X = \alpha_i)\,, \tag{15}$$

where α is the order of the entropy and P_k are the probabilities of $\{X_1, X_2, \ldots, X_n\}$. RE is calculated as $\alpha = 2$ in this study.

3.2. FCBF selection algorithm

Feature selection is used to eliminate irrelevant and redundant features. This improves the prediction accuracy and reduces the computational overheads in terms of classification. The FCBF selection algorithm [33] is used in the study. The algorithm is an efficient feature selection algorithm based on the relevance among features and redundancy values (Figure 5). The FCBF algorithm is a multivariate feature selection method starting with a full set of features, using symmetrical uncertainty (SU) to calculate the dependences of features and arriving at the best subset, using a backward selection technique with a sequential search strategy [34].

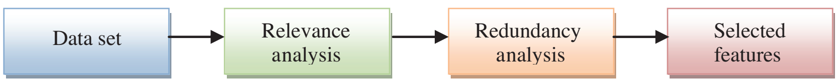

Figure 5. FCBF selection algorithm.

Relevance analysis: Irrelevant features are removed from the original features and correlation is widely used to analyze the relevance. It is the linear correlation coefficient for the X and Y chance variables:

$$r = \frac{\sum_i (x_i - \overline{x_i})\,(y_i - \overline{y_i})}{\sqrt{\sum_i (x_i - \overline{x_i})^2}\,\sqrt{\sum_i (y_i - \overline{y_i})^2}}, \tag{16}$$

where $\overline{x_i}$ is the mean of X and $\overline{y_i}$ is the mean of Y. The value of r lies between –1 and 1. If X and Y are completely correlated, r takes the value of 1 or –1, and if X and Y are totally independent, r is 0 [33]. The SU values are used instead of linear correlation in the FCBF selection algorithm since many systems are nonlinear in the real world. Linear correlation computations may provide inaccurate results in situations where the correlations among the features or the correlations between a feature and a class label are not linear [34]. The SU is defined as:

$$SU\left(X,Y\right) = 2\left[\frac{IG\left(X\,|\,Y\right)}{H\left(X\right)+H(Y)}\right]. \tag{17}$$

In Eq. (17), X and Y represents a vector pair composed of any 2 features or a feature and a class label.

$$\begin{aligned} &IG\left(X|Y\right) = H\left(X\right) - H(X/Y) \\ &H\left(X\right) = -\sum_i P\left(x_i\right)\log_2\left(P\left(x_i\right)\right) \end{aligned} \tag{18}$$

Here, $IG\left(X\,|\,Y\right)$ is the information gain of X after observing variable Y [33]. The entropies of variable X and Y are $H\left(X\right)$ and $H\left(Y\right)$, respectively. $P\left(x_i\right)$ is the probability of variable x and the entropy $\left(H\left(X \mid Y\right)\right)$ of X after observing the values of another variable Y is defined as:

$$H\left(X\,|\,Y\right) = -\sum_j P(y_j)\sum_i P(x_i|y_j)\log_2\left(P\left(x_i y_j\right)\right), \tag{19}$$

where $P\left(x_i\right)$ is the prior probabilities for all of the values of X and $P\left(x_i\,|\,y_i\right)$ is the posterior probabilities of X given the values of Y [33].

SU is the modified version of the information gain with a range between 0 and 1 [35]. The FCBF algorithm removes irrelevant features by ranking correlations $(SU_{i,c})$ between the feature and the class. If the SU between the feature and the class equals 1, this feature is definitely related to that class. However, if the SU is equal to 0, the features are believed to be irrelevant to this class [35].

The FCBF algorithm calculates the SU correlation between any feature F_i and class C, generating a list in descending order, and decides heuristically that feature F_i is relevant if it is highly correlated with class C, i.e. if $SU_{i,c} > \delta$, where δ is a relevance threshold that can be determined by users [36]. Since the calculated nonlinear correlation value is always positive, a situation where the threshold value is 0 indicates that all of the features are ranked with the class labels, starting from the most relevant to the least relevant. Figure 6 displays the list of $SU_{i,c}$ values obtained in this study. Since the use of threshold values bigger than 0 may cause the elimination of some features in the first phase, it is actually effective in accelerating the algorithm [34]. However, since this study is not related to the analysis of the speed but to the performance, the threshold value of 0 is accepted. The selected relevant features are then subjected to redundancy analysis.

Redundancy analysis: In the FCBF algorithm, the SU correlation between the individual features for redundancy analysis is evaluated based on an approximate Markov blanket concept [36]. For 2 relevant features, F_i and F_j $(i \neq j)$, F_i can be eliminated if $SU_{j,i} \geq SU_{i,c}$ (where $SU_{i,c}$ is a correlation between any features and the class, and $SU_{i,j}$ is a correlation between any pair of features). The iteration starts from the first element in the ranking and continues as follows [36]. For all of the remaining features, if F_i happens to form an approximate Markov blanket for F_j, F_j will be taken out of the list. After one cycle based on F_i, the algorithm will take the remaining features that are right next to F_i as the new reference to repeat the filtering process. The algorithm comes to a stop when no more features can be eliminated. The FCBF algorithm is presented in Figure 7 in light of the information provided.

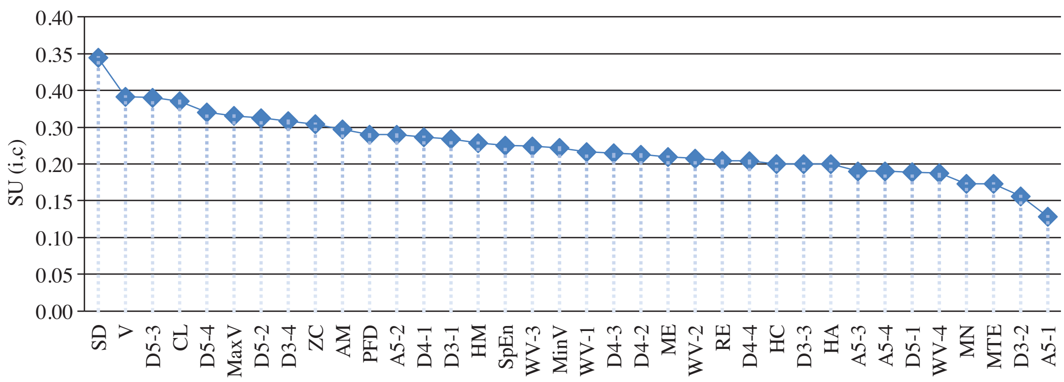

Figure 6. The $SU_{i,c}$ values of features in the data set among the class label vectors.

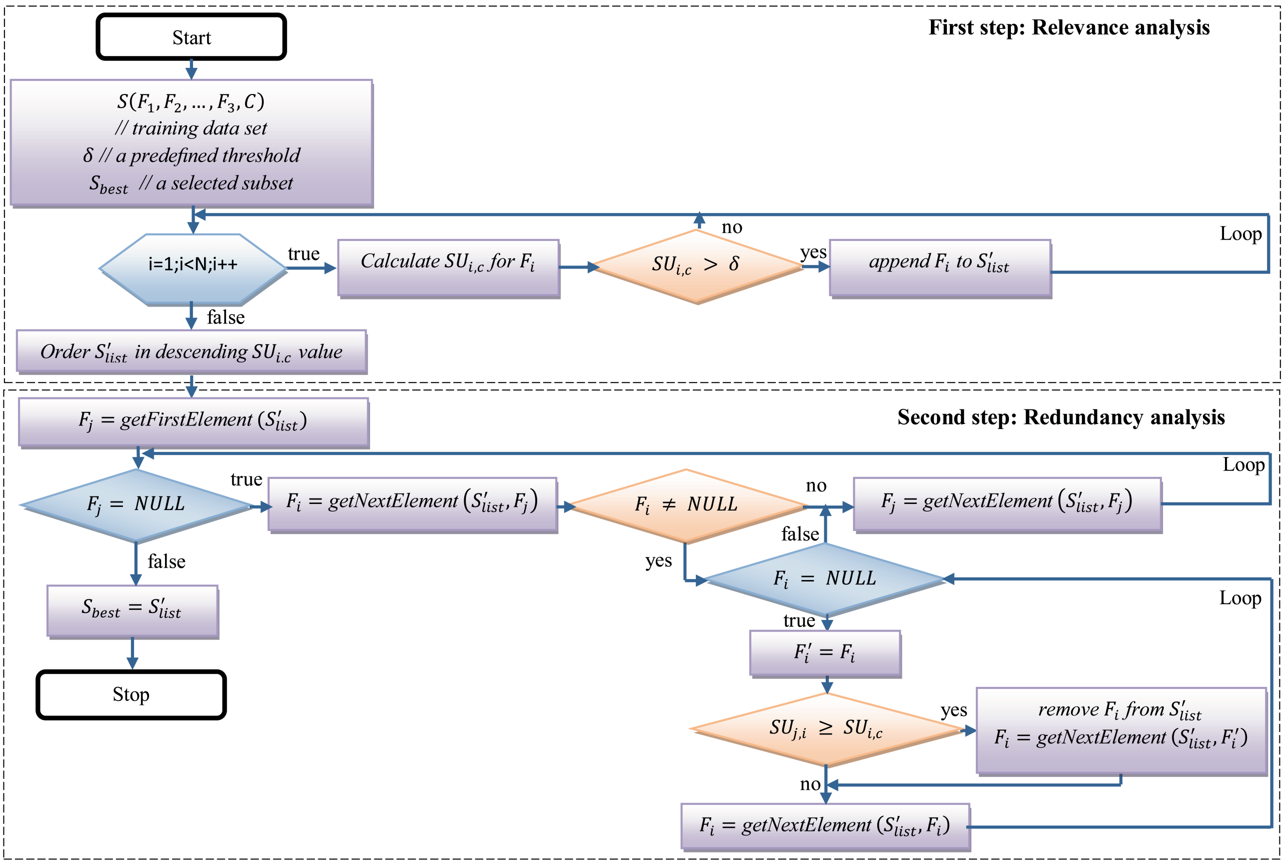

Figure 7. Flow chart of the FCBF algorithm.

One of the advantages of the FCBF algorithm over other feature selection algorithms is its speed. The FCBF algorithm, a filtering method, can identify the value of a feature in a classification by calculating the correlation values between features and class labels or correlation values between 2 features. By doing so, it does not need the results of classification algorithms [34]. This characteristic provides a great advantage over spiral feature selection algorithms in terms of speed. In this study, 7 attribute sets are obtained with the implementation of the FCBF algorithm (Table 4).

Table 4. Selected features with the FCBF.

Feature number	Features
4	SD
10	PFD
12	SpEn
19	D3-3
21	D4-1
33	MCL
35	Hjorth mobility

3.3. Classifier algorithms

NB: NB is a very basic classifier based on the theorem of Bayes [37]. NB assumes that all of the feature nodes are independent from each other given the class and, typically, the feature variables are assumed to have Gaussian distribution if they are continuous. Despite its naive design and apparently oversimplified assumptions, NB has worked rather well in various complex real-world situations [37]. Compared to similar complex graphical models, it requires a smaller amount of training data to accurately estimate the parameters required for classification.

The classification results are determined by the posterior probability $P\left(C|X_1, X_2, \ldots, X_n\right)$, which can be transformed using the chain rule and Bayes' theorem into Eq. (20), where α is a normalization constant:

$$P\left(C|X_1, X_2, \ldots, X_n\right) = \alpha P\left(X_1, X_2, \ldots, X_n|C\right) P\left(C\right) = \alpha \prod_{i=1}^{n} P\left(X_i|C\right) P(C), \tag{20}$$

where the C node represents different classes and $\left(X_1, X_2, \ldots, X_n\right)$ represent different components or features of a sample [37].

Given a test sample $\left(X_1 X_2, \ldots, X_n\right)$, the class is determined by Eq. (21):

$$C^{*} = argmax_C \prod_{i=1}^{n} P\left(X_i \,|\, C\right) P(C). \tag{21}$$

In our case, the class node represents 2 conditions (epilepsy, normal) and the feature nodes $\left(X_1 X_2, \ldots, X_n\right)$ represent the EEG signal features.

Decision trees (C4.5): These have been successfully used in solving problems related to machine learning and classifier systems. It is the most commonly used data-mining technique. Decision trees work by developing a series of 'if-then' rules. An observation is assigned to one segment of the tree by each rule and at that point another 'if-then' rule is applied. The initial segment contains the entire data set and forms the root node for the decision tree [38]. Unlike neural networks and regression, decision trees do not work with interval data. Decision trees work with nominal outcomes that have more than 2 possible results and with ordinal outcome variables [38].

The C4.5 decision tree learning is a method used for discrete-valued functions classifying, in which a C4.5 decision tree depicts the learned function [39]. The objective of C4.5 decision tree learning is to partition the recursive data into subgroups (see [39] for more information on C4.5 decision tree learning).

Logistic regression analysis: Logistic regression is a generalization of linear regression. Although it is very similar to linear regression, the most important difference is the existence of discrete or categorical (discontinuous) dependent variables in logistic regression. Regression analysis is a regression method used for classification and nomination. It is a technique that can be used when the dependent variable is discrete and the independent variables are both discrete and continuous. This technique does not carry prerequisites such as normal distribution and the assumption of continuity [40]. Logistic regression is used with dependent variables to calculate the probability of an expected situation for a dependent variable with double outputs. In order to provide the regression, the dependent variable is transformed into a continuous value, which is the probability of the realization of the expected event [40].

Multilayer perceptron neural network (MLPNN): In this study, 3-layer multilayer perceptron feed-forward neural network architecture is used and trained with the error back-propagation algorithm. The MLPNN is a nonparametric ANN technique that provides various identifications and prediction processes [41–43]. A multilayer neural network includes an input layer, an output layer, and one or multiple hidden layers [44]. The hidden layer neurons and the output layer neurons use nonlinear sigmoid activation functions. The classification accuracy of the applied MLPNN model is examined according to hidden neuron numbers and the minimum number of neurons that give the best results is identified as 5. In this system, 7 of the inputs (Table 4) are features and 2 of the outputs are the indices of 2 classes (epilepsy and normal).

Radial basis network (RBF): This is a different approach that presents the curve-fitting problem in multidimensional space. Radial-based functions have been used in the solutions of multivariable problems in the numerical analysis and in the design of ANNs as well as in the development of ANNs. RBFs are composed of an input layer, a hidden layer, and an output layer. However, transformation from the input layer to the hidden layer is a nonlinear constant transformation with radial-based activation functions. The transformation from the hidden layer to the output layer is a linear one. The free parameters that can be applied in the RBF are central vectors, the width of the radial functions, and the output layer weight. Detailed information about the realization of the RBF and MLPNN structures can be found in the neural network toolbox part of the MATLAB documentation [45].

KMC algorithm: KMC is one of the simplest and most popular unsupervised learning algorithms to solve the clustering problem. The working of the KMC can be summarized as follows [46]:

- Step 1: Choose K initial cluster centers $z_1 z_2, \ldots, z_K$ randomly from the n points $\{X_1, X_2, \ldots, X_n\}$
- Step 2: Assign point X_i, $i = 1, 2, \ldots, n$ to the cluster $C_j, j \in \{1, 2, \ldots, K$.

  If $\|X_i - z_j\| < \|X_i - z_p\|$, $p = 1, 2, \ldots, K$ and $j \neq p$.
- Step 3: Compute new cluster centers as follows:

$$z_i^{new} = \frac{1}{n_i} \sum_{X_j \in C_i} X_j i = 1, 2, \ldots, K, \tag{22}$$

  where n_i is the number of elements belonging to the cluster C_i.
- Step 4: If $\|z_i^{new} - z_i\| < \varepsilon$, $i = 1, 2, \ldots, K$, then terminate.

  Otherwise, continue from Step 2.

In this study, the number of clusters is selected as 2. One of them is for the healthy subjects and the other is for the epilepsy patients.

4. The experimental results

Two different vector matrices with dimensions of 100 $\times$ 4096 are created for both data set A (healthy) and data set E (epileptic activity situation) in MATLAB R2010a in order to extract the features of the data in the study. From these vectors, 36 separate feature parameters are obtained for each column in data sets A and E, with dimensions 100 $\times$ 4096, by the software developed for the study. After an FCBF selection algorithm is applied to this feature set, 7 feature sets are obtained out of the original 36 features. The feature parameters in the table are used as the input parameters for the classification algorithms. The MLPNN, decision tree, NB, KMC, RBF, and logistic regression techniques are implemented using the MATLAB software package (MATLAB version 7.10 with neural network toolbox [45]). Figure 8 presents the structure of the system developed for comparison.

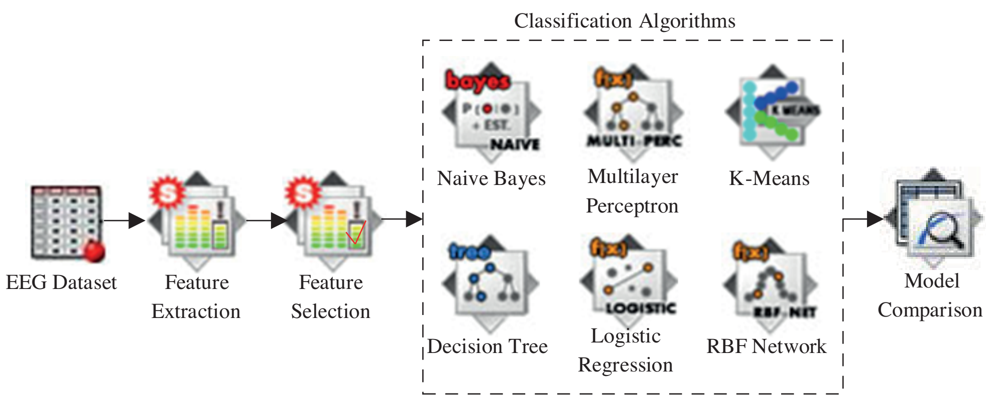

Figure 8. The applied methods for epileptic seizure detection.

4.1. Performance evaluation methods

Five methods, the classification accuracy, sensitivity and specificity analysis, confusion matrix, and k-fold cross-validation, explained in the future sections, are used for the performance evaluation of epileptic seizure detection.

4.1.1. Classification accuracy

The classification accuracies for the data sets are measured according to Eq. (23) in this study:

$$\begin{gathered} classification\,accuracy\,(K) = \frac{\sum_{i=1}^{|K|} assess(k_i)}{|K|} k_i \in K \\ assess\,(k) = \begin{cases} 1, \; if\,classify\,(k) = k.c, \\ 0, \; otherwise \end{cases}, \end{gathered} \tag{23}$$

where K is the set of data items to be classified (the test set), $k \in K$, kc is the class of the item k, and classify(k) returns the classification of k by the classification algorithms.

4.1.2. Sensitivity and specificity

The following expressions for the sensitivity and specificity analyses are used:

$$sensitivity = \frac{TP}{TP + FN}\,(\%) \tag{24}$$

$$specificity = \frac{TN}{FP + TN}\,(\%) \tag{25}$$

Here, $TPTNFP$, and FN denote the true positives, true negatives, false positives, and false negatives, respectively.

4.1.3. Confusion matrix

The confusion matrix shows the classification of the cases in the test data set. In the confusion matrix, the columns denote the actual cases and the rows denote the predicted cases. Table 5 shows the confusion matrix for a 2-class classifier. The entries of our confusion matrix can be interpreted as follows in Table 5:

Table 5. Representation of the confusion matrix.

	Predicted	
Actual	Negative	Positive
Negative	a	b
Positive	a	b

where a is the number of correct predictions that an instance is negative, b is the number of incorrect predictions that an instance is positive, c is the number of incorrect predictions that an instance is negative, and d is the number of correct predictions that an instance is positive.

4.1.4. *k*-Fold cross-validation

k-Fold cross validation is one way to improve the holdout method. The data set is divided into k subsets and the holdout method is repeated $k$times [16]. Each time, one of the k subsets is used as the test set and the other (k – *1*) subsets are put together to form a training set. The average error across all of the k trials is then computed. The advantage of this method is that it is not important how the data are divided. Every data point appears in a test set exactly once and appears in a training set (k – *1*) times. The variance of the resulting estimate is reduced as k is increased. The disadvantage of this method is the necessity of rerunning the training algorithm from scratch k times, which means it takes k times as much computation for the evaluation. A variant of this method is to randomly divide the data into a test and training set k different times. The advantage of this method is the ability of independently selecting the size of each test and number of trials [16]. A 10-fold cross-validation is utilized in this study.

4.2. Results and discussion

The classification of the EEG signals obtained from the healthy individuals and epilepsy patients (data sets A and E) are taken into consideration in 3 phases: feature extraction, feature selection, and classification. Table 6 displays the results collected according to the performance evaluation criteria. The best results for average classification accuracy are obtained using the MLPNN and NB algorithm for the epileptic seizure detection. A

100% classification accuracy is obtained using these algorithms. The lowest classification accuracy is obtained using the KMC algorithm. The time parameter is also taken into consideration in the study and it is observed that the algorithm that classified the problem in the shortest time is the logistic regression method. In the MLPNN algorithm, the classification period may take a bit longer due to the feedback steps.

Table 6. The classification accuracies and sensitivity and specificity values for the 10-fold cross validation.

Method	Fold no.	Confusion matrix		Accuracy	Sensitivity	Specificity	Time (min)
MLPNN	10	100	0	100%	100%	100%	2.1
		0	100				
Decision tree (C4.5)	10	99	1	99%	99%	99%	0.7
		1	99				
NB algorithm	10	100	0	100%	100%	100%	0.6
		0	100				
KMC algorithm	10	98	1	98.5%	98.02%	98.9%	0.9
		2	99				
RBF	10	99	0	99.5%	99.01%	100%	1.1
		1	100				
Logistic regression	10	99	1	99%	99%	99%	0.5
		1	99				

The performance analysis is performed in 2 dimensions. In the first phase, all of the features are used as input parameters, whereas in the second phase, features obtained after the implementation of the FCBF algorithm are taken as input parameters. Figure 9 shows the performance of the classifier before and after feature selection. The experimental results show that the accuracy of the classifier has improved with the removal of the irrelevant and redundant features.

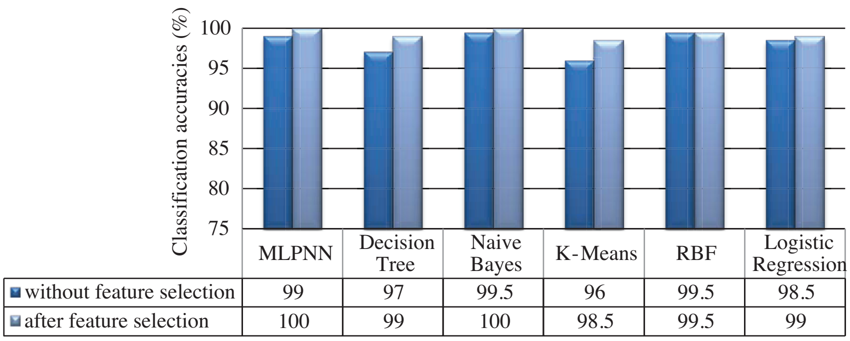

Figure 9. The performance of the classifier before and after feature selection.

The accuracy rates obtained in this study and in previous studies on the same data set are compared. A comparative analysis of the classification accuracies is presented in Table 7. Table 7 shows that the MLPNN [13] and the decision tree algorithm [16] are employed as classification algorithms, which our study also utilizes. Although the same classification algorithms are employed, a higher classification accuracy is found in the present study. This difference is believed to have been caused by the selected feature selection and extraction algorithms.

It can be concluded from the above results that the hybrid system combining the FCBF selection-MLPNN and the FCBF selection-NB, obtains highly accurate results in classifying the possible epileptic seizure of patients. It is thought that the proposed system can be very helpful to physicians in terms of their final decision with regard to their patients.

Table 7. Comparison of the classification accuracies (in percentages) obtained by other researchers for the detection of epileptic seizures.

Authors	Method	Data set	Accuracy
Fathima et al. [4]	Statistical features-discriminant analysis	A, E	96.90%
Subaşı [7]	DWT-mixture of expert model	A, E	95.00%
Kannathal et al. [10]	Entropy measures-ANFIS	A, E	92.22%
Srinivasan et al. [11]	Time and frequency domain feature-recurrent neural network	A, E	99.60%
Guo et al. [13]	DWT-relative wavelet energy-MLPNN	A, E	95.00%
Nigam et al. [14]	Nonlinear preprocessing filter-diagnostic neural network	A, E	97.20%
Tzallas et al. [15]	Time frequency analysis-ANN	A, E	100%
Polat et al. [16]	Fast Fourier transform-decision tree (10-fold cross-validation)	A, E	98.72%
Kannathal et al. [47]	Chaotic measures-surrogate data analysis	A, E	90.00%
Our work	**Different features-FCBF selection-MLPNN**	**A, E**	**100%**
Our work	Different features-FCBF selection-decision tree (C4.5)	A, E	99.00%
Our work	**Different features-FCBF selection-NB**	**A, E**	**100%**
Our work	Different features-FCBF selection-KMC	A, E	98.5%
Our work	Different features-FCBF selection-RBF	A, E	99.50%
Our work	Different features-FCBF selection-logistic regression	A, E	99.00%

5. Conclusion

It is a difficult task to detect epilepsy, which requires the observation of the patient, an EEG, and the collection of additional clinical information. This study proposes a new hybrid model for neurologists to help them to analyze EEG signals in a short time with high accuracy rates. This can be used in the automatic diagnosis of epileptic activity. The prominent parts of the study are listed below:

- The effect of the features selected by the FCBF algorithm on the performance is found to be more positive and higher compared to the use of all of the features. This algorithm can be used in the classification of other medical signals. In this manner, the features that can be more effective in the provision of better performances can be selected.

- This study employs 6 classification algorithms that have been used in the solution of different problems in the literature. Some of the hybrid models presented in the study have proven to be more effective compared to the results of previous studies.

- When the literature on EEG classification is examined, it is seen that many feature extraction algorithms have been used. However, no studies have been undertaken in order to determine the most effective features. The current study presents a new approach to the use of the algorithm in this sense.

- The study presents a different analysis, both in the identification of the most effective features from among the features used for the representation of the problem and also in the identification of the most efficient algorithm in the classification of the problem from among the 6 most popular classification algorithms. The identification of effective features and effective classification algorithms has resulted in high levels of classification accuracy. The same method can be followed for the other biomedical signals in order to provide high accuracy rates. These types of studies may be instrumental in finding effective solutions to the question of which feature algorithm should be used to acquire the feature that can best represent the data.

- The method proposed in this study can easily be applied to other medical data and it is not only for the diagnosis of epilepsy. Researchers can identify the effective features after establishing the feature algorithms that have previously been used for the problems they are studying. It is not difficult to create the format that the model calls for. After feeding the data, the system automatically handles the implementation of the feature algorithms, feature selection algorithm, and classification algorithms in turn. The results of analysis are also provided graphically. When evaluation results are examined, we see that obtaining results in a shorter time is an important benefit. We plan to prepare a visual interface for the model in our future studies. The application of a visual interface may increase the applicability.

Consequently, these structures can be helpful as a learning-based decision support system for aiding doctors in making diagnostic decisions.

References

[1] S. Aydın, "Determination of autoregressive model orders for seizure detection", Turkish Journal of Electrical Engineering & Computer Science, Vol. 18, pp. 23–30, 2010.

[2] World Health Organization, Epilepsy, Geneva, WHO, at http://www.who.int/mediacentre/factsheets/fs999/en/index.html, 2012.

[3] R.B. Artan, E. Yazgan, "Epileptic seizure detection from SEEG data by using higher order statistics and spectra", İTÜ Dergisi, Vol. 7, pp. 102–111, 2008.

[4] T. Fathima, M. Bedeeuzzaman, O. Farooq, U.K. Yusuf, "Wavelet based features for epileptic seizure detection", MES Journal of Technology and Management, pp. 108–112, 2010.

[5] M. Albayrak, E. Köklükaya, "The detection of an epileptiform activity on EEG signals by using data mining process", e-Journal of New World Sciences Academy, 4(1), article number 2A0001, 2009.

[6] C.T. Yuen, W.S. San, M. Rizoni, T.C. Seong, "Classification of human emotions from EEG signals using statistical features and neural network", International Journal of Integrated Engineering, Vol. 1, pp. 71–79, 2009.

[7] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model", Expert Systems with Applications, Vol. 32, pp. 1084–1093, 2007.

[8] H.R. Mohseni, A. Maghsoudi, M.B. Shamsollahi, "Seizure detection in EEG signals: A comparison of different approaches", Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6724–6727, 2006.

[9] M. D'Alessandro, G. Vachtsevanos, A. Hinson, R. Esteller, J. Echauz, B. Litt, "A genetic approach to selecting the optimal feature for epileptic seizure prediction", Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 2, pp. 1703–1706, 2001.

[10] N. Kannathal, M. Choo, U. Acharya, P. Sadasivan, "Entropies for detection of epilepsy in EEG", Computer Methods and Programs in Biomedicine, Vol. 80, pp. 187–94, 2005.

[11] V. Srinivasan, C. Eswaran, N. Sriraam, "Artificial neural network based epileptic detection using time domain and frequency domain features", Journal of Medical Systems, Vol. 29, pp. 647–660, 2005.

[12] F.S. Bao, D.Y. Lie, Y. Zhang, "A new approach to automated epileptic diagnosis using EEG and probabilistic neural network", Proceedings of the 2008 20th IEEE International Conference on Tools with Artificial Intelligence, Vol. 2, pp. 482–486, 2008.

[13] L. Guo, D. Rivero, J.A. Seoane, A. Pazos, "Classification of EEG signals using relative wavelet energy and artificial neural networks", Proceedings of the First ACM/SIGEVO Summit on Genetic and Evolutionary Computation, pp. 177–184, 2009.

[14] V. Nigam, D. Graupe, "A neural-network-based detection of epilepsy", Neurological Research, Vol. 26, pp. 55–60, 2004.

[15] A.T. Tzallas, M.G. Tsipouras, D.I. Fotiadis, "Automatic seizure detection based on time-frequency analysis and artificial neural networks", Computational Intelligence and Neuroscience, Vol. 7, pp. 1–14, 2007.

[16] K. Polat, S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform", Applied Mathematics and Computation, Vol. 187, pp. 1017–1026, 2007.

[17] D.S.T. Pavithra, M.S. Vijaya, "Electroencephalogram wave signal analysis and epileptic seizure prediction using supervised classification approach", Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India, pp. 18–24, 2010.

[18] A., Alkan, E., Köklükaya, A., Subaşı, "Automatic seizure detection in EEG using logistic regression and artificial neural network", Journal of Neuroscience Methods, Vol. 148, pp. 167–176, 2005.

[19] U. Orhan, M. Hekim, M. Özer, "EEG signals classification using the k-means clustering and a multilayer perceptron neural network model", Expert Systems with Applications, Vol. 38, pp. 13475–13481, 2011.

[20] R.G. Andrzejak, G. Widman, K. Lehnertz, C. Rieke, P. David, C.E. Elger, "The epileptic process as nonlinear deterministic dynamics in a stochastic environment: an evaluation on mesial temporal lobe epilepsy", Epilepsy Research, Vol. 44, pp. 129–140, 2001.

[21] A. Petrosian, "Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns", Proceedings of the 8th IEEE Symposium on Computer-Based Medical Systems, pp. 212–217, 1995.

[22] J.F. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal", International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 381–384, 1990.

[23] R. Esteller, J. Echaus, T. Tcheng, B. Litt, B. Pless, "Line length: an efficient feature for seizure onset detection", Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1707–1710, 2001.

[24] M.J. Katz, "Fractals and the analysis of waveforms", Computers in Biology and Medicine, Vol. 18, pp. 145–56, 1988.

[25] J. Semmlow, Biosignal and Biomedical Image Processing, New York, Marcel Decker, 2004.

[26] E. Avşar, "Epileptic EEG signal classification using one-class support vector machines", MSc, İstanbul Technical University, 2009.

[27] M.L.V Quyen, J. Martinerie, M. Baulac, F. Varela, "Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings", Neuroreport, Vol. 10, pp. 2149–215, 1999.

[28] B. Hjorth, "Time domain descriptors and their relation to a particular model for generation of EEG activity", in: CEAN – Computerized EEG Analysis, Stuttgart, Gustav Fischer Verlag, pp. 3–8, 1975.

[29] Ö. Özgonenel, G. Önbilgin, Ç. Kocaman, "Transformer protection using the wavelet transform", Turkish Journal of Electrical Engineering & Computer Sciences, Vol. 13, pp. 119–135, 2005.

[30] H. Ocak, "Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy", Expert Systems with Applications, Vol. 36, pp. 2027–2036, 2009.

[31] M. Sabeti, S. Katebi, R. Boostani, "Entropy and complexity measures for EEG signal classification of schizophrenic and control participants", Artificial Intelligence in Medicine, Vol. 47, pp. 263–274, 2009.

[32] A. Rényi, "On a new axiomatic theory of probability", Acta Mathematica Hungarica, Vol. 6, pp. 285–33, 1995.

[33] L. Yu, H. Liu, "Feature selection for high-dimensional data: a fast correlation-based filter solution", Proceedings of the Twentieth International Conference on Machine Learning, pp. 856–863, 2003.

[34] B. Senliol, G. Gulgezen, L. Yu, Z. Cataltepe, "Fast correlation based filter (FCBF) with a different search strategy", Proceedings of the 23rd International Symposium on Computer and Information Sciences, pp. 1–4, 2009.

[35] R. Duangsoithong, T. Windeatt, "Relevance and redundancy analysis for ensemble classifiers", Proceedings of the 6th International Conference on Machine Learning and Data Mining in Pattern Recognition, pp. 206–220, 2009.

[36] J. Sánchez-Monedero, M. Cruz-Ramírez, F. Fernández-Navarro, J. C. Fenández, P. A. Gutiérrez, C. Hervás-Martínez, "On the suitability of extreme learning machine for gene classification using feature selection", Proceedings of the 10th International Conference on Intelligent Systems Design and Applications, pp. 507–512, 2010.

[37] R.M. Hope, Z. Wang, Z. Wang, Q. Ji, W.D. Gray, "Workload classification across subjects using EEG", Proceedings of the HFES Annual Meeting, Vol. 55, pp. 202–206, 2011.

[38] R. Daş, "A comparison of multiple classification methods for diagnosis of Parkinson disease", Expert Systems with Applications, Vol. 37, pp. 1568–1572, 2010.

[39] S. Güneş, K. Polat, Ş. Yosunkaya, "Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting", Expert Systems with Applications, Vol. 37, pp. 7729–7736, 2010.

[40] C. Colak, M.C. Colak, M.N. Orman, "[The comparison of logistic regression model selection methods for the prediction of coronary artery disease]", The Anatolian Journal of Cardiology, Vol. 7, pp. 6–11, 2007 (article in Turkish with English abstract).

[41] H. Göktaş, A. Çavuşoğlu, B. Şen, İ. Toktaş, "The use of artificial neural networks in simulation of mobile ground vehicles", Mathematical and Computational Applications, Vol. 12, pp. 87–96, 2007.

[42] F.V. Celebi, K. Danisman, "A different approach for the computation of refractive index change in quantum-well diode lasers for different injection levels", Proceedings of the SPIE - The International Society for Optical Engineering, Vol. 5662, pp. 384–388, 2004.

[43] F.V. Celebi, "A proposed CAD model based on amplified spontaneous emission spectroscopy", Journal of Optoelectronics and Advanced Materials, Vol. 7, pp. 1573–1579, 2005.

[44] F.V. Celebi, "Modeling of the linewidth enhancement factors of the narrow and wide GaAs well semiconductor lasers", Journal of the Faculty of Engineering and Architecture of Gazi University, Vol. 21, pp. 161–166, 2006.

[45] MATLAB®Documentation, Version 7.10, Release (R2010a), The MathWorks, Inc.

[46] H. Güldemir, A. Sengür, "Comparison of clustering algorithms for analog modulation classification", Expert Systems with Applications, Vol. 30, pp. 642–649, 2006.

[47] N. Kannathal, U.R. Acharya, C.M. Lim, P.K. Sadasivan, "Characterization of EEG-a comparative study", Computer Methods and Programs in Biomedicine, Vol. 80, pp. 17–23, 2005.