

CLASSIFICATION OF EPILEPTIC AND HEALTHY INDIVIDUALS WITH RECURRENCE PARAMETERS

Funda Kutlu Onay¹, Cemal Kose²

¹ Amasya University, Amasya, Turkey

² Karadeniz Technical University, Trabzon, Turkey

Epilepsy is a brain activity disorder that manifests itself with epileptic seizures. Although the reasons of epilepsy are not fully known, the diversity and variability of the individual make it difficult to diagnose epilepsy. For this, the diagnosis of epilepsy with computerized systems is one of the most popular research topics in recent years. Although many techniques and methods have been developed for this purpose, Electroencephalogram (EEG) signals are one of the most preferred and most basic ways to diagnose epileptic seizures or epilepsies because of their practicality and easy application. However, interpretation of EEG signals is not easy due to nonlinear and variable signal characteristics. This has led to the preference of nonlinear methods as well as traditional methods in the study of EEG signals. It can be seen from the literature that non-linear methods give very successful results in previous studies. In this study; EEG signals from healthy and epileptic subjects were represented by recurrence parameters obtained by recurrence plot. The features extracted from the recurrence plot are applied to multi-layered artificial neural networks, k-nearest neighbors, and support vector machines classifiers after feature selection process. Accordingly, the highest classification accuracy was achieved at around 97.05% when the multi-layered artificial neural network was used.

Index Terms—recurrence parameters, recurrence quantification analysis, EEG, epilepsy, non-linear classification

I. INTRODUCTION

COMPUTER-ASSISTED disease diagnosis is one of the most studied areas on the present day. Thanks to the developed software, misdiagnosis of diseases is reduced and the time for diagnosis is shortened. One of the most common areas of use of this technology is the diagnosis of epilepsy.

Epilepsy is a disorder of brain dysfunction that occurs in the form of recurrent seizures and affects about 1% of the world's population [1]. Epilepsy affects human life seriously because it causes depression, anxiety disorders, psychosis in adults and deficiency, hyperactivity, stuttering in children. [2]. As a result, epilepsy has an important position in biomedical researches for many years. The reasons of epilepsy are not known exactly, but the human brain has the potential to have epileptic seizures, even with neurological damages such as head trauma, high fever. Today, approximately 30,000 new epilepsy patients are attending annually, and about 30% of them are children under the age of 18. It is known that approximately 720,000 individuals in Turkey are suffering from epilepsy [3].

Epilepsy is commonly seen in mammalian species due to their complex brain structures [1]. More than 40 types of seizures have been described in the literature. For this reason, epileptic seizures are difficult to perceive and it takes a long time to find the source of the problem. In fact, since symptoms from other ailments are sometimes interpreted as epileptic

seizures, BCI systems have begun to be developed to prevent the misdiagnosis of diseases [2].

The human brain continuously generates an electrical current at low voltage. These signals, which are mostly transmitted in a computer environment as EEG or ECoG signals, are used for automatic diagnosis and computer analysis for epilepsy or other purposes. Due to the non-invasive, practical availability and ease of application, EEG signals are the most preferred signals in BCI applications. Due to non-periodic and non-stationary characteristics and unstable amplitude, phase, and frequency values, interpretation of EEG signals are difficult. As such, there is a great need for the development of efficient automated methods and systems for EEG analysis [4].

In this study, the EEG data set was used which is recorded by Bonn University for the detection of epilepsy and perception of epileptic seizures. These data were recorded from 100 channels for 23.6 sec, with 5 healthy and 5 epileptic patients. Each trial has three periods: normal, pre-seizure, and seizure [5]. There are successful studies in the literature where this data set is used. Gautama et al. utilized nonlinear methods to represent normal and epileptic EEG signals and attained 86.2% classification accuracy [6]. Kannathal et al. achieved 90% classification accuracy using entropy and adaptive fuzzy logic interference system [7]. Tzallas et al. studied using time-frequency analysis and k-EYK classifier and obtained 97.71-100% success rate [8]. Polat and Güneş acquired 98% classification accuracy with decision trees and fast Fourier transformation based systems [9]. Acharya achieved 97.7%

Manuscript received May 5, 2017; accepted May 25, 2017.
Corresponding author: Funda Kutlu Onay (E-mail:
funda.kutlu@amasya.edu.tr; ORCID® ISNI: 0000-0002-8531-4054).
Digital Object Identifier:

and 94.7% classification accuracy using RQA with SVM and Fuzzy classifiers, respectively [10].

II. DATASET DESCRIPTION AND PROPERTIES

The dataset used in the study is an EEG database recorded in the Epileptology Department of the University of Bonn in Germany and open to public access. This data was recorded with a 128-channel 12-bit EEG recording system and sampled with a sampling frequency of 173.61. Therefore, each experiment is 23.6 seconds long. There are a total of 500 trials in this way and these were taken from 5 groups (A, B, C, D, E). The experimental groups, explanations, and trial numbers are given in Table I. All experiments are free of physical artifacts caused by eye and muscle movements [5].

Table I. Information of experimental groups

Groups	Explanation	Trial numbers
A	Eyes open from healthy individuals	100
B	Eyes closed from healthy individuals	100
C	Seizure-free from hippocampal formation of the opposite hemisphere of the brain	100
D	Seizure-free from epileptic zone	100
E	Seizure activity	100

EEG signals were recorded from volunteers according to the international 10-20 electrode placement system. A sample test of each of the experimental groups is given in figure 1, as A-B-C-D-E, respectively.

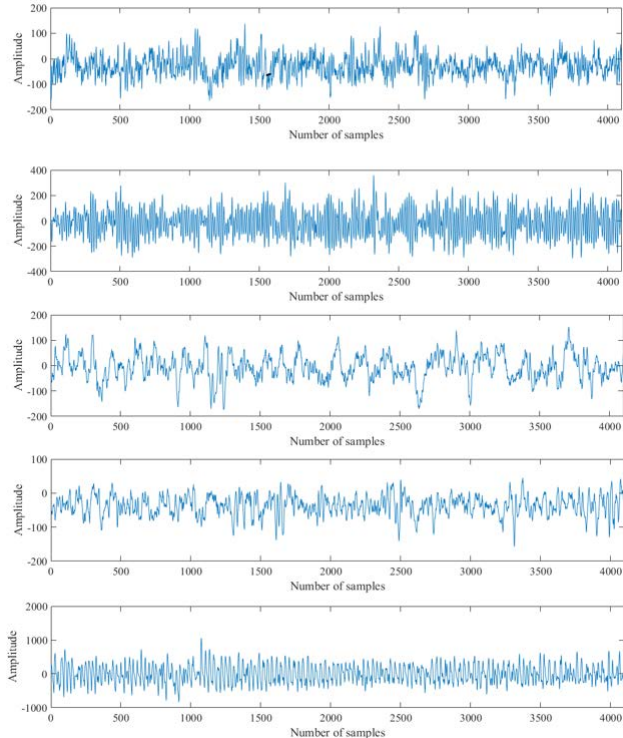


Fig. 1. Sample EEG signals for A, B, C, D, and E, respectively.

III. FEATURE EXTRACTION AND SELECTION

Recurrence quantification analysis (RQA) is a method used by Eckmann [11] to analyze nonlinear data [11]. It is possible to pick out the dynamic areas and to detect the nonlinear states on the signal with RQA parameters extracted from the recurrence plot (RP). Since EEG signals also have these characteristic features, RQA has become a frequently used method in the analysis of EEG signals. The mathematical equation used in RP calculation is given as

$$R_{i,j} = \Theta (\varepsilon_i - \|x_i - x_j\|), \quad x_i \in \mathfrak{R}^m, \quad i, j = 1 \dots N, \quad (1)$$

The RQA metrics obtained from RP are explained in Table II. Here, $\|\cdot\|$ corresponds to Euclidean distance; x_i , points to the considered point and N is the total number of points to be considered. ε_i is the predetermined threshold distance and, θ_i Heaviside step function and is defined as

$$\Theta = \begin{cases} 1 & \text{if } s \geq 0 \\ 0 & \text{if } s < 0 \end{cases} \quad (2)$$

Table II. Description of RQA parameters

RQA parameters	Equation	Description
Recurrence Rate (RR)	$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}$	It gives the density of repetitive points in RP. Corresponds to the sum of the correlations.
Determinism (DET)	$DET = \frac{\sum_{l=\min}^N IP(l)}{\sum_{i,j}^N R_{i,j}}$	Corresponds to the percentage of repetition points located on the diagonals.
Laminarity (LAM)	$LAM = \frac{\sum_{v=\min}^N vP(v)}{\sum_{v=1}^N vP(v)}$	Corresponds to the percentage of repeat points located on the vertical line.
L	$L = \frac{\sum_{l=\min}^N IP(l)}{\sum_{l=\min}^N P(l)}$	Average diagonal line length
Trapping time (TT)	$TT = \frac{\sum_{v=\min}^N vP(v)}{\sum_{v=\min}^N P(v)}$	Average vertical line length
Selective Entropy (ENTR)	$ENTR = - \sum_{l=\min}^N p(l) \ln p(l)$	

There are three important parameters for the creation of the correct and most selective RP: embedding dimension, time delay and threshold. False Nearest Neighbor (FNN) was used in determining the best embedding dimension parameter in the study and was found as 7 for healthy individuals and as 6 for epileptic individuals. The Mutual Information (MI) algorithm was applied to find the best time delay parameter and was obtained as 4 for healthy individuals and as 3 for epileptic individuals. The healthy and epileptic RPs generated

by these parameters are given in figure 2-a and figure 2-b, respectively.

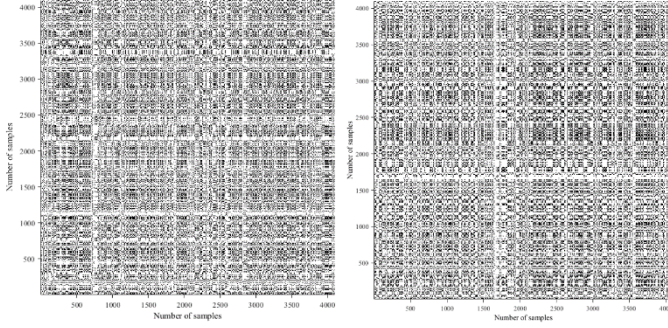


Fig. 2. Samples RPs for a) healthy b) epileptic

Feature selection is a process performed to determine the properties that best represent the class within a feature set. In order to increase classification performance, the features that do not contain enough information about the class are eliminated by feature selection algorithms. In this study, the feature selection process was performed by Chi-square method. Chi-square is a hypothesis testing method that tests the dependency and independence between two variables. The feature selection based chi-square testing consists of two steps. In the first part of the method, the Chi-square statistics of the attributes are calculated according to the class characteristic. In the second phase, the attributes are decomposed until the inconsistent features in the dataset are determined by looking at the Chi-square values in the Chi-Merge principle, depending on the degree of freedom and the level of significance determined. The Chi-square value for an instance in the dataset indicates the dependency of this in the class. If this value is zero, the instance is independent within that class; if it is a high value, this example is more specific for the data set [12]. The mathematical equation used to calculate the Chi-square value is given as

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^k \frac{(A_{ij} - E_{ij})^2}{E_{ij}} \quad (3)$$

According to this equation, k is the number of classes, A_{ij} (i , row, j , column) is the measured frequency value and E_{ij} is the expected frequency value.

The feature vector obtained after the feature selection process is applied to the classifiers. The classifiers used in the next section will be mentioned.

IV. CLASSIFICATION

A. MULTILAYER ARTIFICIAL NEURAL NETWORKS (MLANN)

Artificial neural network is a classification technique inspired by the work of the human brain and is based on learning. The human brain has the ability to learn events and make inferences from these events, and ANN studies are simulated by models of neural networks in the brain. The task of an ANN is to specify the set of outputs that correspond to a

given set of inputs [13]. MLANN is a neural network model consisting of one or more layers. A MLANN consists of three basic layers: an input layer that serves data, a hidden layer (or more) that best separation between classes and an output layer that reveals class information about the input value. An example of a MLANN structure is given in figure 3.

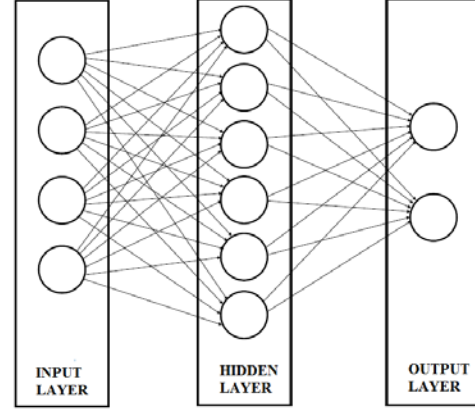


Fig. 3. Illustration of a MLANN architecture

The neurons in the hidden layer(s) detect features. For this reason, the number of layers and neurons in the hidden layer(s) are important parameters for the best separation. Apart from this, learning coefficient and momentum coefficient are two other important parameters. The momentum coefficient allows the weight change value to be added to the next change in a certain amount in the learning process [14].

B. SUPPORT VECTOR MACHINES (SVM)

The support vector machine (SVM) is a classification technique that can be used for both linear and non-linear problems. The basic approach in SVMs is based on finding the best decision boundary that separates the two classes, namely hyperplane [15]. The SVM initially applied for linear classifications was then applied to nonlinear problems with the addition of kernel functions. Because real-time problems are largely unsuitable for linear decomposition, which makes the use of kernel functions a requirement. The data cannot be separated linearly, it is moved from one dimension to another using kernel functions, and they can be linearly separated in this new space [16].

Choosing the appropriate kernel function is very important for correct classification. The most commonly used kernel functions in the literature are polynomials, normalized polynomials, Pearson VII and radial basis kernel functions, and radial basis function (RBF) is the most successful kernel function among them. RBF is used in this study and the mathematical expression is calculated as

$$K(x, y) = e^{-\gamma \|x - x_i\|^2} \quad (4)$$

Here, γ refers to the kernel dimension.

C. K-NEAREST NEIGHBORS (K-NN)

k-nearest neighbors (k-NN) is a successful classification technique that is often used to classify data in pattern recognition and machine learning areas [17]. The label value of the pattern is marked according to the k neighbor(s) which is closest to itself. If the majority of the neighbors belong to which class, the pattern to be recognized belongs to that class. The key parameter in the performance of the k-NN algorithm is to determine the best k value. One of the most commonly used methods for determining the best k value is cross-validation. Another important parameter is the distance metric used to calculate the closeness between patterns. There are studies on Euclid, Hamming, Mahalanobis, Manhattan, Minkowski, Cosine distances in the literature. As a result, the Euclidean distance was preferred in this study as it gave more positive results in the previous studies. Euclidean distance between two points is calculated as

$$d = \sqrt{(A_1 - B_1)^2 + \dots + (A_N - B_N)^2} \quad (5)$$

V. RESULTS AND DISCUSSION

In this study, RQA parameters were calculated to classify healthy and epileptic individuals. As a result, the Chi-square test was applied to determine the distinctive RQA parameters. According to this, 6 of the 8 RQA parameters were found: RR, DET, L, ENTR, LAM, TT.

For k-NN, 10-Fold Cross Validation (10-FCV) was used for the analysis of k value. The k values and the classification accuracy determined for each group are given in Table III.

Table 3. Classification accuracy of k-NN classifier

Groups	The best k value	Sensitivity (%)	Specificity (%)	Accuracy (%)
AB-E	1	100	92.8	96.4
AB-CD	3	87	85.9	86.4
AB-CDE	3	87.4	89.1	88.4
CD-E	1	91	82.2	86.6

The kernel-based SVM classifier is used to solve non-linear problems. An important parameter in the SVM classification which performed by Gauss RBF is the σ parameter representing the radial diameter. For the best σ value, 10-FCV was used as in k-NN. Classification results are given in Table IV.

Table IV. Classification accuracy of SVM classifier

Groups	The best σ value	Sensitivity (%)	Specificity (%)	Accuracy (%)
AB-E	0.7	98.4	92.4	95.4
AB-CD	0.5	87.9	83.7	85.8
AB-CDE	0.5	86	86.2	86.1
CD-E	0.3	94.2	82.6	88.4

Finally, the MLANN system is designed to consist of input layer with 6 neurons, hidden layer with 40 neurons and output layer with 1 neuron. Sigmoid activation function was used and

the learning and momentum coefficients were determined as 0.2 and 0.3, respectively. The classification results obtained are given in table V.

Table V. Classification accuracy of MLANN classifier

Groups	Sensitivity (%)	Specificity (%)	Accuracy (%)
AB-E	96	89.4	92.7
AB-CD	97.8	98.08	97.05
AB-CDE	97.5	93.5	95.5
CD-E	95.7	85.5	90.6

According to this, the highest classification rate is seen in the case of using MLANN, in the non-seizure period when the patients and healthy individuals are distinguished. The most important advantage of this method is that it can be thought of as processing in time space without moving the sign to another space. However, since the creation of the RP takes a long time, when the signal size increases, the size of the RP can be reduced by performing a size reduction on the signal during the preprocessing step. Otherwise, in the previous work [18] the default values were used to generate the RP. In this study, the parameters that are effective in the creation of the RP are obtained by optimizing with the existing algorithms. As a result, it is also seen that the classification accuracy achieved in the previous study is over.

ACKNOWLEDGMENT

The study is selected from International Engineering Research Symposium-UMAS 2017 (Duzce University).

REFERENCES

- [1] H. Reynolds, "Epilepsy: The Disorder," *Epilepsy Atlas*, pp. 15-27, 2005.
- [2] "Epilepsy Foundation", received from the address: <http://www.epilepsy.com/learn/epilepsy-101/what-epilepsy>.
- [3] G. D. Cascino, "When drugs and surgery don't work", *Epilepsia*, vol.9, pp. 79-84, 2008.
- [4] R. M. Rangayyan, *Biomedical Signal Analysis*, Canada: John Wiley & Sons, Inc., 2002.
- [5] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state", *Physical Review E*, vol. 64, no.6, 2001.
- [6] T. Gautama, D. P. Mandic and M. M. V. Hulle., "Indications of nonlinear structures in brain electrical activity." *Physical Review E*, vol. 67, no. 4, 2003.
- [7] N. Kannathal, M. L. Choo, U. R. Acharya and P. K. Sadasivan, "Entropies for detection of epilepsy in EEG", *Computer methods and programs in biomedicine*, vol. 80, no. 3, pp. 187-194, 2005.
- [8] T. Tzallas, M. G. Tsipouras and D. I. Fotiadis, "Automatic seizure detection based on time-frequency analysis and artificial neural networks", *Computational Intelligence and Neuroscience*, 2007.
- [9] K. Polat and 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, no. 2, pp. 1017-1026, 2007.
- [10] U. R. Acharya, S. V. Sree, S. Chattopadhyay, W. Yu and P. C. A. Ang, "Application of recurrence quantification analysis for the automated identification of epileptic EEG signals", *International journal of neural systems*, vol. 21, no. 3, pp. 199-211, 2011.

- [11] J. P. Eckmann, S. O. Kamphorst and D. Ruelle, "Recurrence plots of dynamical systems", *Europhysics Letters*, vol.4, no.9, 973, 1987.
- [12] H. Liu and R. Setiono, "Chi2: Feature selection and discretization of numeric attributes", *Tools with artificial intelligence*, pp. 388-391, 1995.
- [13] E. Alpaydın, *Yapay Öğrenme*, Boğaziçi Üniversitesi Yayınevi, İstanbul, Mart 2011.
- [14] E. Öztemel, *Yapay Sinir Ağları*, Papatya Yayıncılık, İstanbul, 2006.
- [15] S. Abe, *Support Vector Machines for Pattern Classification*, Springer, Newyork, 2010.
- [16] C. Cortes and V. Vapnik, "Support Vector Networks", *Machine Learning*, vol. 20, pp. 273-297, 1995.
- [17] D. T. Larose, *Discovering Knowledge in Data*, A John Wiley & Sons Inc. Publication, Newyork, 2005.
- [18] Kutlu and C. Kose, "Detection of epileptic seizure from EEG signals by using recurrence quantification analysis" presented at the 22nd IEEE Signal Processing and Communications Applications Conference (SIU), pp. 1387-1390, Trabzon, 2014.

BIOGRAPHIES

Funda Kutlu Onay received the BSc and MSc degrees from the Karadeniz Technical University (KTU), Computer Engineering Department, in 2011 and 2013 respectively. She is currently PhD student at the same university and research assistant in Computer Engineering Department of Amasya University. Her research interests are biomedical signal processing, pattern recognition and human-computer interfaces.

Cemal Kose received the BSc and MSc degrees from the Karadeniz Technical University, Turkey, in 1986 and 1990 respectively. He received a PhD degree from the University of Bristol, UK, in 1997. He became associate professor in 2009 and full professor in 2014 at the Karadeniz Technical University (KTU). He is currently professor in the Department of computer Engineering at KTU. His research interests are medical image processing, pattern recognition and information extraction.