Lal Hussain*, Wajid Aziz, Sharjil Saeed, Saeed Arif Shah, Malik Sajjad A. Nadeem, Imtiaz Ahmed Awan, Ali Abbas, Abdul Majid and Syed Zaki Hassan Kazmi

Quantifying the dynamics of electroencephalographic (EEG) signals to distinguish alcoholic and non-alcoholic subjects using an MSE based K-d tree algorithm

DOI 10.1515/bmt-2017-0041
Received March 31, 2017; accepted June 21, 2017

Abstract: In this paper, we have employed K-d tree algorithmic based multiscale entropy analysis (MSE) to distinguish alcoholic subjects from non-alcoholic ones. Traditional MSE techniques have been used in many applications to quantify the dynamics of physiological time series at multiple temporal scales. However, this algorithm requires $O(N^2)$, i.e. exponential time and space complexity which is inefficient for long-term correlations and online application purposes. In the current study, we have employed a recently developed K-d tree approach to compute the entropy at multiple temporal scales. The probability function in the entropy term was converted into an orthogonal range. This study aims to quantify the dynamics of the electroencephalogram (EEG) signals to distinguish the alcoholic subjects from control subjects, by inspecting various coarse grained sequences formed at different time scales, using traditional MSE and comparing the results with fast MSE (fMSE). The performance was also measured in terms of specificity, sensitivity, total accuracy and receiver operating characteristics (ROC). Our findings show that fMSE, with a K-d tree algorithmic approach, improves the reliability of the entropy estimation in comparison with the traditional MSE. Moreover, this new technique is more promising to characterize the physiological changes having an affect at multiple time scales.

Keywords: complexity analysis; electroencephalogram (EEG); fast multiscale sample entropy (fMSE); multiscale sample entropy (MSE).

***Corresponding author: Dr. Lal Hussain,** Assistant Director, Quality Enhancement Cell, The University of Azad Jammu and Kashmir, City Campus, Muzaffarabad, 13100, Azad Kashmir, Pakistan, E-mail: Lall_hussain2008@live.com; Lal.hussain@ajku.edu.pk. http://orcid.org/0000-0003-1103-4938
Wajid Aziz: Faculty of Computing and IT, University of Jeddah, Kingdom of Saudi Arabia; and Department of Computer Sciences and Information Technology, The University of Azad Jammu and Kashmir, City Campus, Muzaffarabad, 13100, Azad Kashmir, Pakistan
Sharjil Saeed, Saeed Arif Shah, Malik Sajjad A. Nadeem, Imtiaz Ahmed Awan, Ali Abbas, Abdul Majid and Syed Zaki Hassan Kazmi: Department of Computer Sciences and Information Technology, The University of Azad Jammu and Kashmir, City Campus, Muzaffarabad, 13100, Azad Kashmir, Pakistan

Introduction

In 1929 Hans Berger developed a non-invasive system to record the electroencephalography (EEG) to measure and analyse the performance of neural activity which remains a challenge in neuroscience. Physicians have used EEG to compute the neurological activities of the brain such as brain disorder involving epilepsy and many other brain disorders [10]. Moreover, the non-linear EEG time series analysis and chaotic dynamical systems analysis [5] has always remained a great source of attraction for researchers. Richman and Moorman used entropy based methods for the measurement and estimation of regularity in time series signals [2].

In a previous study [3], an imbalance was observed in the central nervous system (CNS) between excitation inhibition states in the presence of alcoholic and offspring subjects resulting in the high risks of biological and other related disorders also increasing the β activity. EEG coherence due to alcoholism was observed to be significantly increased between the interhemispheric region in high θ (6–7 Hz). Porjesz and Rangaswamy [32] reported that alcoholics manifest significant increases in resting EEG that affect particularly posteriorly at the centroparietal and parietal-occipital regions. Moreover, alcoholic subjects' activity was observed as being high which is an indication of transformed thalamo and cortical functional connectivity. Alcoholic subjects may also suffer from some other fundamental psychiatric disorders, e.g. alcohol dependence, impulsivity, oppositional disorder, conduct disorder, drug dependence and attention deficit hyperactivity

disorder (ADHD) which also cause environmental and genetic influences on internalizing and externalizing disorders. The frontal brain region activation is reduced due to alcoholism and affect impulsivity.

The neurophysiological signals are very complex and require methods from the theory of nonlinear dynamics to fully understand its dynamics. The complexity of a system or signals can be computed with several nonlinear dynamical measures such as entropy-based computational techniques and fractal dimensions. These techniques have been employed to distinguish the healthy and pathological subjects to measure the complexity of the physiological activities [15, 16]. The most emerging theoretical framework is to examine the complexity of biological systems including behavior, physiology associated with aging and disease [14, 26, 34, 44]. The complexity concept is linked with several concepts in biology and physics such as entropy, information theory and randomness [22, 35, 39, 40]. Multiple factors affect the behavior and physiological signals and accordingly the systems' output. Complexity is examined using multiple methods from the theory of nonlinear dynamics [24] together with information theory [35].

Recently, the physiological and behavioral complexities were examined using correlation dimension [17], spectral analysis [20], detrended fluctuation analysis [19], recurrence plot analysis [11, 42, 43] and false nearest neighbors [25]. The entropy based methods have been widely used to analyze the complexity of the non-linear signals. In the previous studies, a number of variants were used to achieve an efficiency in computation, varying the degree of flexibility and relevance to problems. The information processing in the brain with respect to the processing point of view manifests itself through its global activity measured by EEG that is a multidimensional and nonlinear time series. Besides, variants of EEG have been employed to discriminate healthy aging and pathological aging using complexity based techniques [29].

Multiscale entropy (MSE) analysis was used by [7, 8] to compute the complexity of a time series having a finite length. Using MSE, the complexity is computed at multiple temporal scales. However, the computational methodology is important to quantify the dynamics of a time series. MSE requires an order of $O(N^2)$ which may be unrealistic for long-term data sets and online monitoring. The rolling window method was employed by [18] to reduce the execution time useful to analyze the local data. The data is portioned into different windows and MSE is computed for each window. Here the window size selection is very important for signals with long range correlation to cover the largest time scale in the signal. In this regard, we have employed an efficient way to compute the MSE using the K-d tree algorithm approach. In the field of computational geometry, the MSE was computed by considering the probability function as an orthogonal range search problem. A new K-d tree algorithm as developed by [31] was used to reduce the execution time of order $O(N^2)$.

The aim of this study is to apply the nonlinear complexity methods using sample entropy with a K-d tree algorithm approach at multiple time scales and comparing the results with MSE as approximate entropy depends on the length of data, but sample entropy is independent of data length. From the literature, it was observed that complexity decreases for pathological subjects such as AD, dementia and epilepsy, etc. Alcoholic subjects may also suffer different pathological disorders such as impulsivity, conduct disorder, oppositional disorder and drug dependence. Thus, the present study will focus on investigating, if there is possibility, that using nonlinear entropy-based techniques the complexity of alcoholic subjects can be decreased more than in control subjects. The research reported in this article is aimed to investigate the dynamics of EEG signals in alcoholics and to distinguish them from non-alcoholic subjects using a robust entropy measure with the K-d tree algorithmic approach. The research also reveals which electrode and brain regions are more dominant to distinguish these conditions.

Proposed methods

EEG recordings

The datasets are taken from an online database available at the University of California, Irvine Knowledge Discovery in Databases (UCI KDD) Machine Learning Repository archive (http://archive.ics.uci.edu/ml/datasets/EEG+Database visited on March 11, 2014). For the current analysis, we have used 29 subjects for both the alcoholic (mean age 35.83, SD 5.33, range 22.3–49.8 years) and control groups (mean 25.81, SD 3.38, range 19.4–38.6 years) and were used previously in alcohol-related studies [38, 45, 46]. The alcoholic dependence or abuse were confirmed by the psychiatrist at the Addictive Disease Hospital in Brooklyn where the participants were recruited from. Moreover, the alcoholic participants were detoxified fully and most of them were also abstained from alcohol for at least 28 day, to ensure that there would be no short-term effects of alcohol. The alcoholic subjects had been drinking for minimum of 15 years. The control subjects participated in the study reported no personal or family history relating to drink alcohol or drug abuse. Individuals with severe medical problems or drug dependence were excluded from the study. Moreover, these participants are right handed with normal vision. During the data acquisition the subjects were exposed to visual stimuli – pictures of these objects were chosen from the Snodgrass and Vanderwart picture set [37]. The data were recorded from 61 electrodes placed according to the 10–20 international montage (impedance of electrode was kept below 5 kΩ, implication gain

of 10,000, pre-filtered from 0.02 Hz to 50 Hz and referenced to Cz), sampled at a rate of 256 Hz. Moreover, excess of body and eye movements were rejected [45]. Likewise, for event related potentials (ERP), the magnitude was recorded below 5 μV. The standard electrode positions were used as illustrated elsewhere in the literature [1]. The EEG signals are extracted from 14 electrodes – occipital (O1, O2), central (C3, C4), frontal (F3, F4, F7, F8), frontpolar (Fp1, Fp2), parietal (P3, P4) and temporal (T7, T8).

Sample entropy

Richman and Moorman [33] proposed the sample entropy to analyze the dynamics of physiological time series. Moreover, sample entropy [27] was used to extract information in a time series and measurements of the systematic structure was carried out by testing the repeated patterns of varying length. Mathematically, it is computed as a negative average natural logarithm of conditional probability.

Consider a time series $x(1)$, $x(2)$, $x(3)$, $x(4)$, …, $x(N)$, where N is the data length. Sample entropy can be computed as follow:

(i) $X(i) = X(i), X(i+1), X(i+2), \ldots, (Xi+m-1)]$. (1)
where $i = 1, 2, 3, \ldots, N-m+1$.

(ii) The distance $d[X(i), X(j)]$ between two series $X(i)$ and $X(j)$ is computed as

$$[X(i), X(j)] = \max[|x(i+k) - x(j+k)|], \quad (2)$$

where $k = 0, 1, 2, \ldots, m-1$.
and $j = 1, 2, 3, \ldots, N-m+1$; but $j \neq i$.

(iii) Count $d[X(i), X(j)]$ which are smaller than the given threshold r, then compute the ration of this number with total $N-m$ as

$$C_i^m(r) = \frac{\{\text{number of } d[X(i), X(j)] < r\}}{N-m}, \quad (3)$$

where $i = 1, 2, 3, \ldots, N-m+1$.

(iv) Compute $C_i^m(r)$ for all I as

$$C^m(r) = \frac{\sum C_i^m(r)}{N-m+1}, \quad (4)$$

where $i = 1, 2, 3, \ldots, N-m$.

(v) To yield $m+1$ dimension, add one to number of dimension of the vector. Repeat the steps i to iv until yield $C^{m+1}(r)$.

(vi) Finally compute the sample entropy as

$$\text{Samp}En(m, r, N) = -\ln\left[\frac{C^{m+1}(r)}{C^m(r)}\right] = \ln\frac{n_d}{n_n}. \quad (5)$$

Multiscale sample entropy (MSE): The limitations of sample entropy restrict it to apply only over a single time scale factor that may not be suitable for a complex time series whose characteristics are computed at multiple temporal scales. To address this, [7] developed a multiscale sample entropy that is flexible in terms of its applications. MSE is computed at various time scales.

Consider an one-dimensional time series

$$x = \{x_1, x_2, x_3, \ldots x_N\}.$$

The following procedure is used to compute the coarse-grained time series $y^{(\tau)}$ at scale τ

$$y_j^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i, \; 1 \le j \le \frac{N}{\tau}, \quad (6)$$

where coarse-grained is developed in non-overlapping window of length τ and data points are averaged in each window.

K-d tree algorithm approach

Manis [28] proposed a fast K-d tree algorithmic approach based on approximate entropy. Silpa-Anan and Hartley [36] proposed an improved version of the K-d tree algorithm that can create the multiple randomized K-d trees. Pan et al. [31] developed a new K-d tree algorithm based on sample entropy to compute the dynamics of physiological time series from RR intervals and ECG and EEG signals. Using the K-d tree algorithm to compute the sample entropy, the distance vectors (probability) n_n and n_d are computed for each scale.

The time series $x = \{x_1, x_2, x_3, \ldots x_N\}$ can be transformed into "d" dimensional points where $d = m+1$ by setting

$$p_i = x_i, \; q_i = x_{i+1}, \; r_i = x_{i+2}. \quad (7)$$

The terms in equation (5) satisfying the following constraints are equivalent to the number of points:

$$x_i - \epsilon < x_i + \epsilon; \; x_{i+1} - \epsilon < x_{i+1} < x_{i+1} + \epsilon;$$

$$x_{i+2} - \epsilon < x_{i+2} < x_{i+2} + \epsilon \quad (8)$$

Define

$$p_1 = x_j - \epsilon, \; p_2 = x_j + \epsilon;$$

$$q_1 = x_{j+1} - \epsilon, \; p_2 = x_{j+1} + \epsilon;$$

$$r_1 = x_{j+2} - \epsilon, \; r_2 = x_{j+2} + \epsilon. \quad (9)$$

From equations (7) and (9), n_n corresponds to the number of points that satisfy the following constraints:

$$p_{i1} < p_j < p_{i2};$$

$$q_{i1} < q_j < q_{i2};$$

$$r_{i1} < r_j < r_{i2}. \quad (10)$$

In other words, for each point

$$X_i = (P_i, Q_i, R_i) \quad 1 \le i \le N.$$

The bounding box number of points can be calculated as:

$$[p_1 : p_2] \times [q_1 : q_2] \times [r_1 : r_2]. \quad (11)$$

In computational field geometry, this approach is employed for orthogonal range search problems. Where the computation of n_d to m counting problems and n_n are equivalent to $m-1$ dimensional orthogonal range counting problems. When n_n and n_d are calculated then the sample entropy can be computed directly.

K-d tree algorithm: The K-d tree algorithm is a binary tree algorithm which was proposed by Bentley in [4], its each node "v" is associated with a rectangle Bv. "v" will be the leaf node, if Bv does not contain any point in its interior. Otherwise, Bv is partitioned into two rectangles by drawing a vertical and horizontal line such that each rectangle contains at most half of the points. The following steps

as proposed by [28, 31, 36] are used to compute the sample entropy using the K-d tree algorithm:

(i) Compute the coarse scale of discrete time series by using Eq. (6) for each scale.
(ii) Discrete time series is transformed into discrete space point sets by Eq. (7).
(iii) Set $K = m - 1$; build the K-d tree using the space points.
(iv) The query box is calculated using Eq. (9) to each space point.
(v) Using the K-d tree query algorithm, the number of points inside the box (n_i^m) are queried.
(vi) n_n is computed using Eq. (5).
(vii) Set $k = m$, repeat step (ii) to (vi) to compute n_d.
(viii) Finally, SE is computed using Eq. (5).

The time complexity of the original MSE algorithm was $O(N^2)$ for each scale as two loops (i, j) is required and thus total execution time is:

$$\sum_s n_s^2 = \sum_s \left(\frac{n}{s}\right)^2 = O(N^2). \tag{12}$$

To overcome the time and complexity, the following steps are employed as used by [31]:

Step 1. Original discrete time series was transformed to a special point set from $x = \{x_1, x_2, x_3, \ldots x_N\}$.

Step 2. The d-dimensional K-d tree is constructed using $N - m$ points for which the total cost is $O(N \log N)$ and the memory is $O(N)$.

Step 3. Range query; for, d-dimensional K-d tree search the time cost is $N\,O\left(N^{1-\frac{1}{d}}\right)$ of for N queries and memory cost is $O(N)$.

Fast MSE (fMSE) technique

The sample entropy algorithm used in the Section "Sample entropy" require two loops (i, j) for each scale, thus total execution time for all the scale require the time complexity of $O(N^2)$ which is too slow in many applications. For computation of sample entropy (S_E), the values for $m = 2$ and $r = 0.15 \times \text{SD}$ of the original time series were taken. Computing the sample entropy directly from the Section "Sample entropy" one needs to count the number of its matched pairs thus it requires execution time proportional to the square of length of the input time series. Thus, for long time series the K-d tree approach is highly desirable. Manis [28] used the bucket assisted technique to reduce the execution time for approximate entropy that improved it, however, the algorithm still requires a $O(N^2)$ execution time where N represent the length of the time series. To overcome the complexity of MSE, the fMSE was used that employs the K-d tree algorithmic approach to accelerate the counting of the number of matched pairs of the pattern emplaces in a time series. Bentley, in 1975 [4] developed the K-d tree algorithm based on the well-utilized nearest neighbor (NN) method. This algorithm works as a binary search tree in which each node is representing a partition of K-dimensional space. The entire space is represented by the root node, whereas, the subspace is represented by the leaf nodes that contain the mutually exclusive small subsets. The computational time is accelerated by speeding up the counting of the number of matched pairs of the input time series and subsequences derived from the input time series are organized in the K-d tree data structure. The fMSE algorithm uses $O\left(N^{2-\frac{1}{m+1}}\right)$ execution time, where m denotes the number of pattern template of the time series. Moreover, space complexity is $O(N \log N)$. Time complexity is measured in terms of total elementary operations including addition/subtraction, division/multiplication and comparison of two numbers. The space complexity is measured in terms of number of the elementary objects required to be stored during the execution.

Performance measures based on confusion matrix parameters

To distinguish the alcoholic and non-alcoholic subjects, the following measures were used to compute the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), false discovery rate (FDR), false omission rate (FOR) and total accuracy with one concrete example at electrode C3 from 29 alcoholic subjects and 29 from that of non-alcoholic subjects:

Confusion matrix:

	Predicted		
Actual	True positive (TP) = 29	False positive (FP) = 0	PPV = TP/(TP + FP) =29/(29 + 0) =100%
	False negative (FN) = 1	True negative (TN) = 28	NPV = TN/(TN + FN) =28/(28 + 1) =96.55%
	Sensitivity = TP/(TP + FN) =29/(29 + 1) =96.67%	Specificity = TN/(TN + FP) =28/(28 + 0) =100%	

Sensitivity: The sensitivity measure is used to test the proportion of people who test positive for the disease among those who have the disease. Mathematically, it is expressed as:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{13}$$

i.e. the probability of positive test given that the patient has the disease.

Specificity: Specificity measures the proportion of negatives that are correctly identified. Mathematically, it is expressed as:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{14}$$

i.e. probability of a negative test given that patient is well.

Positive predictive value (PPV): PPV is mathematically expressed as:

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{15}$$

where TP denote that the test makes a positive prediction and subject has a positive result under the gold standard while FP is the event where the test makes a positive prediction and subject has a negative result.

Negative predictive value (NPV): NPV can be computed as:

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}, \tag{16}$$

where TN indicates that test makes a negative prediction and the subject also has negative result, while FN indicates that test makes a negative prediction and the subject has a positive result.

False discovery rate (FDR): It is the compliment of PPV, i.e. FDR = 1 – PPV. It can also be mathematically computed as:

$$\mathrm{FDR} = \frac{\mathrm{FP}}{\mathrm{TP}+\mathrm{FP}}. \tag{17}$$

False omission rate (FOR): It is the compliment of NPV, i.e. FOR = 1 – NPV. It can also be mathematically computed as:

$$\mathrm{FOR} = \frac{\mathrm{FN}}{\mathrm{TN}+\mathrm{FN}}. \tag{18}$$

Total accuracy (TA): The total accuracy is computed as:

$$\mathrm{TA} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}+\mathrm{TN}}. \tag{19}$$

Results

Table 1 shows the results distinguishing the healthy and alcoholic subjects, using MSE and fMSE at multiple temporal scales in electrodes O1, O2, C3, C4, F3, F4, F7, F8, Fp1, Fp2, T7, T8, P3 and P4 using the rank sum test. The mean ranks of control subjects are greater than that of alcoholic subjects at each electrode, which shows that control subjects exhibit higher complexity than the alcoholic subjects. The maximum separation was obtained at smaller scales, however; few electrodes show highest separation at higher scales. The highest separation was obtained at central electrodes followed by parietal, occipital, temporal and frontal electrodes using both MSE and fMSE. Moreover, fMSE exhibits higher separation than MSE to distinguish alcoholic and control groups in most of selected 14 electrodes except in few electrodes such as F3, F8, Fp1 and P4 showed maximum separation using MSE.

Cao et al. [6] recently used synchronization analysis to investigate synchronization difference between 28 alcoholics and 28 non-alcoholics during certain cognitive tasks on the scale EEG. The results reveal that the alcoholic group exhibits lower synchronization than the control group when performing the same cognitive tasks. Moreover, synchronization for the control groups shows the complexity levels of cognitive tasks whereas the alcoholic groups shows only erratic changes. Tcheslavski and Gonen [41] employed spectrum coherence and phase synchrony and observed that alcoholic group have lower phase synchrony than the control group because prolonged excessive alcohol consumption results in damage of individuals and societies both physically and psychologically. de Bruin et al. [9] used a synchronization measure and observed that there is a loss of lateralization in α and β synchronization in both male and female heavy drinkers. These measures are based on correlation and coherence. In the literature, entropy-based complexity measures are most widely used to quantify the dynamics of highly complex and nonlinear signals. Recently, [21]

Table 1: Comparison of results using MSE and fMSE profiles for time scales ($t \leq 20$) to distinguish alcoholic and control subjects for all channels with $m = 2$ and $r = 0.15$ times the standard deviation of the original data sequence.

Electrode	Mean ranks		MSE		fMSE		ROC	
	Control	Alcoholic	p-Value	Scale	p-Value	Scale	MSE	fMSE
C3	41.14	18.63	2.01E-10	1	6.51E-11	1	0.976	0.998
C4	40.46	19.27	6.04E-10	1	6.51E-11	1	0.951	0.968
F3	36.43	23.03	1.68E-03	1	4.22E-03	2	0.746	0.720
F4	31.43	27.70	3.05E-03	18	2.45E-02	20	0.718	0.680
F7	45.54	23.87	1.12E-02	14	5.72E-04	2	0.690	0.764
F8	37.86	21.70	5.27E-05	1	1.30E-04	1	0.809	0.798
Fp1	30.89	28.20	2.88E-04	16	1.91E-02	10	0.768	0.695
Fp2	31.36	27.77	3.05E-02	17	8.63E-03	10	0.680	0.692
O1	38.36	21.23	1.46E-02	1	7.08E-05	2	0.690	0.791
O2	34.61	24.73	1.02E-01	1	9.40E-03	1	0.626	0.708
P3	38.86	20.77	4.80E-08	1	1.81E-08	1	0.912	0.930
P4	32.50	26.70	6.03E-06	2	1.15E-03	2	0.842	0.743
T7	39.43	20.23	1.81E-02	1	2.85E-06	2	0.688	0.877
T8	32.13	26.68	3.43E-01	4	7.48E-03	2	0.553	0.702

Significant group differences are calculated using the rank sum nonparametric test.

employed the symbolic time series analysis to compute the dynamics of epileptic seizures and EEG during resting states and results are compared with the MSE. Both MSE and symbolic entropy give highly significant difference to distinguish the control group from epileptic (ictal-seizures interval), focal and non-focal (interictal interval-seizure free) and EEG signals eye closed from that of eye open during the resting state. In the present study, the mean ranks for the 14 selected electrodes were computed and results revealed that the control group exhibit higher ranks than that of the alcoholic group which indicate that the control group shows higher complexity than the alcoholic group consistent with previous studies. However, MSE and fMSE at multiple temporal scales give higher significant results (p-value) and performance in terms of specificity, sensitivity, PPR, NPR, total accuracy and AUC then the existing techniques employed. The findings reveal that entropy base measures are a very helpful tool to quantify the dynamics of highly complex time series data. Moreover, MSE with the K-d tree algorithm gives higher separation and significance results than MSE in most of the electrodes and regions. Moreover, from the existing literature, fMSE is also more robust in terms of time and space complexity in addition to the performance measure.

The highest separation using fMSE was obtained at the central probe such as C3 with p-value (6.51E-11), C4 with p-value (6.51E-11) followed by the parietal probe P3 with p-value (1.81E-08), temporal probe T7 with p-value (2.85E-06), occipital probe O1 with a p-value (7.08E-05), and frontal probes F8 with p-value (1.30E-04) and F7 with p-value (5.72E-04). The other electrodes such as F3, F4, Fp1, Fp2, O2 and T8 also give significant results to distinguish these groups. Moreover, the highest separation using MSE was obtained at central electrodes such as C3 with a p-value (2.01E-10), C4 with p-value (6.04E-10) followed by the parietal probe P3 with p-value (4.80E-08), P4 with p-value (6.03E-06) and frontal probes F8 with p-value (5.27E-05) and Fp1 with p-value (2.88E-04). The other electrodes such as F3, F4, Fp1, F7, Fp2 and T7 also give significant results to distinguish these groups. Only electrodes O2 and T8 did not show the significant results to distinguish these groups.

Receiver operator curve (ROC)

The ROC is plotted against sensitivity and specificity values of alcoholic and control subjects. Using MSE, the scale values for both subjects are chosen again the maximum separation p-values and are classified as 1 for alcoholic and 0 for control subjects. The vector is then passed to the ROC function developed in Matlab R2013a which plots each sample's values against specificity and sensitivity values. An important parameter from the ROC plot is area under the curve (AUC). In the past, [13] employed ROC to analyze and visualize the behavior of the diagnostic system. It is a two-dimensional graph in which the sensitivity, i.e. true positive rate (Tpr) is plotted on the y-axis while specificity, i.e. false positive rate (Fpr) is plotted on the x-axis. The value of the AUC shows the portion of the area of a unit square. The value of the AUC lies between 0 and 1. AUC values >0.5 indicate good separation. In general, a larger the AUC shows the higher separation, i.e. the better the diagnostic test. Similarly, the AUC using fMSE was greater than the AUC using MSE accordingly as the significant level was obtained. Figures 1 and 2 show the highest separation at parietal (P3, P4) and central (C3, C4) probes to distinguish the healthy and alcoholic subjects. Using fMSE the highest AUC was obtained at central electrodes such as C3 (AUC = 0.998), C4 (0.968) followed by the P3 (AUC = 0.930), T7 (AUC = 0.877), F8 (AUC = 0.798), O1 (AUC = 0.971), F7 (AUC = 0.764), P4

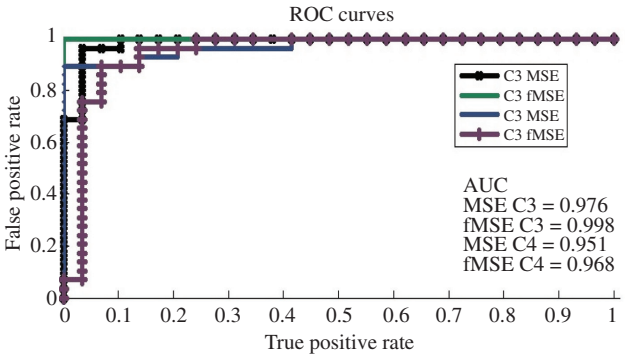

Figure 1: Receiver operator curve (ROC) at the central probe.

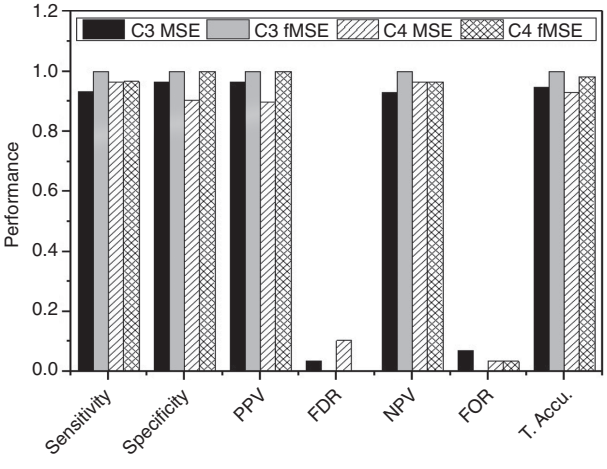

Figure 2: Performance measure at the central electrodes.

(AUC = 0.743), F3 (AUC = 0.726), O2 (AUC = 0.708), T8 (AUC = 0.702), Fp1 (AUC = 0.695), Fp2 (AUC = 0.692) and F4 (AUC = 0680). Moreover, using MSE the highest AUC was obtained at central electrodes such as C3 (AUC = 0.976), C4 (0.951) followed by the P3 (AUC = 0.912), P4 (AUC = 0.842), F8 (AUC = 0.809), Fp1 (AUC = 0768), F3 (AUC = 0.746), F4 (AUC = 0.718), F7 (AUC = 0.690), O1 (AUC = 0.690), T7 (AUC = 0.688), Fp2 (AUC = 0.680), O2 (AUC = 0.626) and T8 (AUC = 0.533).

Moreover, the performance was measured for MSE and fMSE using multilayer perceptron and 10-fold cross validation by computing sensitivity, specificity, PPV (positive predictive value), NPV (negative predictive value) and total accuracy. The highest accuracy was obtained in electrode C3 (100%), C4 (98.28%), P3 (86.21%) higher using fMSE than MSE. Other electrodes F8 (79.31%), F7 (74.14%), O1 (71.41%), T7 (75.86%) also showed good performance in terms of total accuracy using fMSE.

The central and parietal electrodes show highest separation using these performance measure evaluators followed by few frontal electrodes.

The highest sensitivity values were obtained at electrode C3 (100%), followed by C4 (96.67%), F7 (88.89%), F8 (86.96%), P3 (86.21%). The other electrodes F3, O1, O2, T7 and T8 also show good separation. The higher sensitivity values were obtained at electrodes C3 (100%), followed by P3 (86.21%). Likewise, higher PPV was found in electrode C3 and C4 (100%) followed by P3 (86.21%). Higher NPV was obtained at electrode C3 (100%) followed by C4 (96.55%), F7 (93.10%), F8 (89.66%), and P3 (86.21%). The corresponding FDR and FOR values are obtained accordingly.

The Figure 3 shows the performance evaluation at the frontal, occipital, partial and temporal electrodes to discriminate the alcoholic and non-alcoholic subjects. The performance was evaluated using specificity, sensitivity, PPV, NPV, FDR, FOR and total accuracy using features calculated using MSE and fMSE. In the above electrodes, fMSE gives higher performance such as T7 (75.86%), F7 (74.14%), O1 (72.41%) than MSE as T7 (55.17%), F7 (67.24%) and O1 (53.4%). While P3 (89.66%) gives higher total accuracy using MSE than P3 (86.21%) total accuracy using fMSE. The sensitivity values obtained using fMSE are F7 (88.89%), O1 (74.07%), P3 (86.21%) and F7 (81.25%), O1 (71.43%), P3 (92.59%), T7 (54.84%) using MSE. The specificity values using fMSE are obtained as F7 (67.50%), O1 (70.97%), P3 (86.21%), T7 (75.86%) while specificity values using MSE are obtained as F7 (61.90), O1 (70.97%), P3 (87.10%), T7 (55.56%). Whereas PPV and NPV using fMSE at these electrodes are found to be higher than MSE as reflected in the figures above. The other detailed performance measures values in term of specificity, sensitivity, PPV, NPV, FDR, FOR and total accuracy using fMSE and MSE are depicted in the Tables 2 and 3.

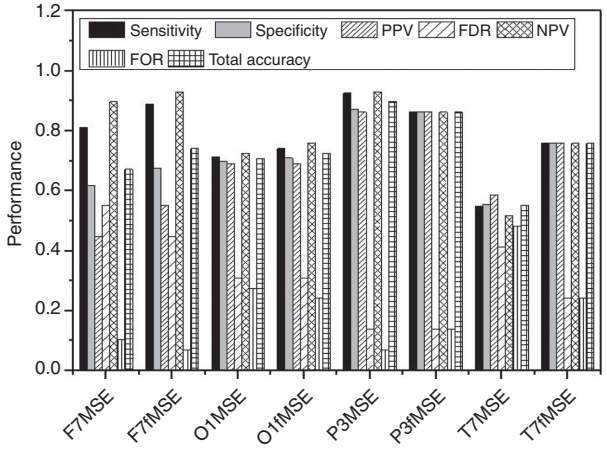

Figure 3: Comparison of performance measure at different electrodes.

Discussion

The main objective of this study is to distinguish alcoholic subjects from that of non-alcoholic subjects using the entropy based K-d tree algorithmic approach. The AUC results and mean ranks are computed for 14 selected central, frontal, occipital, parietal and temporal electrodes. The results are compared using MSE and MSE with the K-d tree algorithmic approach. The mean ranks at all the selected electrodes for non-alcoholic subjects were found to be larger than for the alcoholic subjects. This implies that the complexity of alcoholic subjects decreases due to degrading in the functional and structural components. Due to alcoholism, the impaired memory and other cognitive functions decrease [30, 47]. Intoxicated alcoholic subjects have reduced BOLD responsiveness in visual and auditory cortices. The findings suggest that alcohol can reduce the functional activation of the cortical regions. The larger differences in the mean ranks are observed at central, occipital, parietal and temporal regions which indicates that these regions may be greatly affected due to alcoholism. Using fMSE, the highest separations to distinguish alcoholic from non-alcoholic subjects were obtained at central electrodes C3 and C4 (p-value 6.5E-11) with AUC (0.998) followed by P3 (p-value 1.81E-08) with AUC (0.936), T7 (2.85 E-06) with AUC (0.877), F8 (p-value 1.30E-04) with AUC (0.789) and F7 (p-value 5.72E-04) with AUC (0.764). Other electrodes also show only significant

Table 2: Performance measure using MSE.

Electrode	Sensitivity	Specificity	PPV	FDR	NPV	FOR	Total accuracy
C3	0.9333	0.9643	0.9655	0.0345	0.9310	0.069	0.9483
C4	0.9630	0.9032	0.8966	0.1034	0.9655	0.0345	0.9310
F3	0.8500	0.6842	0.5862	0.4138	0.8966	0.1034	0.7414
F4	0.6538	0.6250	0.5862	0.4138	0.6897	0.3103	0.6379
F7	0.8125	0.6190	0.4483	0.5517	0.8966	0.1034	0.6724
F8	0.9474	0.7179	0.6207	0.3793	0.9655	0.0345	0.7931
Fp1	0.6786	0.6667	0.6552	0.3448	0.6897	0.3103	0.6724
Fp2	0.5667	0.5714	0.5862	0.4138	0.5517	0.4483	0.5690
O1	0.7143	0.7000	0.6897	0.3103	0.7241	0.2759	0.7069
O2	0.5333	0.5357	0.5517	0.4483	0.5172	0.4828	0.5345
P3	0.9259	0.8710	0.8621	0.1379	0.9310	0.069	0.8966
P4	0.7600	0.6970	0.6552	0.3448	0.7931	0.2069	0.7241
T7	0.5484	0.5556	0.5862	0.4138	0.5172	0.4828	0.5517
T8	0.5172	0.5172	0.5172	0.4828	0.5172	0.4828	0.5172

Table 3: Performance measure using fMSE.

Electrode	Sensitivity	Specificity	PPV	FDR	NPV	FOR	Total accuracy
C3	1.000	1	1	0	1	0	1
C4	0.9667	1	1	0	0.9655	0.0345	0.9828
F3	0.7391	0.6571	0.5862	0.4138	0.7931	0.2069	0.6897
F4	0.5161	0.5185	0.5517	0.4483	0.4828	0.5172	0.5172
F7	0.8889	0.675	0.5517	0.4483	0.931	0.069	0.7414
F8	0.8696	0.7429	0.6897	0.3103	0.8966	0.1034	0.7931
Fp1	0.500	0.500	0.4828	0.5172	0.5172	0.4828	0.500
Fp2	0.625	0.5476	0.3448	0.6552	0.7931	0.2069	0.569
O1	0.7407	0.7097	0.6897	0.3103	0.7586	0.2414	0.7241
O2	0.7143	0.6216	0.5172	0.4828	0.7931	0.2069	0.6552
P3	0.8621	0.8621	0.8621	0.1379	0.8621	0.1379	0.8621
P4	0.6471	0.7083	0.7586	0.2414	0.5862	0.4138	0.6724
T7	0.7586	0.7586	0.7586	0.2414	0.7586	0.2414	0.7586
T8	0.7037	0.6774	0.6552	0.3448	0.7241	0.2759	0.6897

results. The higher significance results obtained using fMSE at most of electrodes than MSE.

In this study, the dynamics of EEG signals obtained from alcoholic and non-alcoholic subjects were quantified using the K-d tree algorithm at multiple values and results were compared with MSE and fMSE. fMSE is a suitable method for the analysis of the physiological signals that can be applied to comparatively noisy and short time series regardless of knowing the origin of the signals whether deterministic or stochastic [7]. MSE requires $O(N^2)$ computational time which may be inefficient for online applications and long-term correlations in the data. A new approach using an efficient fMSE method as developed by [31] was used to give more significant results and reduced execution time as well.

Entropy analysis shows that alcoholic subjects usually have lower sample entropy values than the control subjects at smaller time scales ($\tau \leq 5$) on most of the electrodes with significant p-values as shown in the Table 1. Thus, it may be clearly inferred that the brain activity in alcoholic subjects are lesser complex than that of control subjects. These results also reveal same resemblance with the previous studies that reports the less complexity of pathological subjects than the average and healthy subjects [12]. Thus, low complexity in the case of pathological subjects may be due to the extensive death of neurons and the decrease in connectivity of local neural networks that affect overall neuron cell death [23].

The findings indicate that we can discriminate the alcoholic subjects from that of control subjects by means of complexity analysis, that due to alcoholism the complexity decreases in various conditions as depicted above which is consistent with complexity analysis using MSE [8]. Based upon the findings and results obtained during

the current study, it can be inferred that fMSE profiles in most time scales ($\tau \geq 5$) reveal no significant differences which indicate that the alcoholic and control groups evolve in a similar way in these time scales, however, the significant differences are seen in smaller time scales ($\tau \leq 4$) with p-value < 0.05 at electrodes located in the central region (C3 and C4) followed by the parietal region (P3, P4), posterior brain region (T7, T8, O1, O2) and frontal region (F3, F7 and F8). While electrodes F4, frontpolar (Fp1 and Fp2) have shown significant differences at higher time scales. Likewise, the ROC curves are shown to measure the ability of the slope to help the clinical classification of alcoholic subjects against control subjects. The higher separations were obtained accordingly as the significant results were obtained using both MSE and fMSE. The performance was also measured to distinguish the alcoholic and non-alcoholic subjects in terms of sensitivity, specificity, PPV, NPV, FDR, FOR and total accuracy using features from MSE and fMSE. The values obtained at most of the electrodes such as central, frontal, occipital, parietal and temporal electrodes using fMSE give higher performance than MSE as depicted in Figures 2, 3 and Tables 2, 3.

Finally, MSE is more robust than other traditionally used non-linear techniques such as L1, D2 as it can be applied as relatively noisy and small physiological signals and independent of the model [8] as applicable in this case with time series having smaller data points for both alcoholic and control subjects but it has limitations of time and space complexity for long-term correlation data sets and online monitoring. However, fMSE using the K-d tree algorithm approach is more effective and provides significant results at all electrodes and higher separations using ROC than traditional MSE.

Conclusion

Contemporarily, the complexity based measures have been used ubiquitously to examine the behavior of physiological signals. The entropy based methods are effectively used to characterize and distinguish the physiological signals across diverse pathological and physiological conditions such as HRV abnormality, gait analysis and neural disorder. In the present study, MSE with the K-d tree algorithmic approach was used to distinguish the alcoholic subjects from that of control subjects and results are compared with traditional MSE developed by [7]. The results revealed that the complexity of healthy subjects was higher than the complexity of alcoholic subjects and it was shown that neurophysiological signals of the human body are effected at multiple temporal signals and exhibit fluctuations. The central and parietal probes show the highest significant results and higher separation followed by occipital and frontal probes which show that these brain regions can help the neurologist and clinician in curing the patients suffering from alcoholism. Moreover, it was also observed that performance measures, to distinguish the alcoholic and non-alcoholic subjects, in most of the electrodes using sensitivity, specificity, NPV, PPV, FDR, FOR and in terms of total accuracy obtain higher rankings using fMSE features than MSE.

References

[1] Acharya JN, Hani AJ, Cheek J, Thirumala P, Tsuchida TN. American Clinical Neurophysiology Society guideline 2: guidelines for standard electrode position nomenclature. Neurodiagn J 2016; 56: 245–252.
[2] Amin HU, Malik AS, Badruddin N, Chooi WT. EEG mean power and complexity analysis during complex mental task. In Complex medical engineering (CME), 2013 ICME International Conference. IEEE, 2013 648–651.
[3] Begleiter H, Porjesz B. What is inherited in the predisposition toward alcoholism? A proposed model. Alcohol Clin Exp Res 1999; 23: 1125–1135.
[4] Bentley JL. Multidimensional binary search trees used for associative searching. Commun ACM 1975; 18: 509–517.
[5] Briggs K. An improved method for estimating Liapunov exponents of chaotic time series. Phys Lett A 1990; 151: 27–32.
[6] Cao R, Deng H, Wu Z, Liu G, Guo H, Xiang J. Decreased synchronization in alcoholics using EEG. IRBM 2017; 38: 63–70.
[7] Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of complex physiologic time series. Phys Rev Lett 2002; 89: 068102.
[8] Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of biological signals. Phys Rev E 2005; 71: 021906.
[9] de Bruin EA, Stam CJ, Bijl S, Verbaten MN, Kenemans JL. Moderate-to-heavy alcohol intake is associated with differences in synchronization of brain activity during rest and mental rehearsal. Int J Psychophysiol 2006; 60: 304–314.
[10] Easwaramoorthy D, Uthayakumar R. Analysis of biomedical EEG signals using wavelet transforms and multifractal analysis. In Communication Control and Computing Technologies (ICCCCT), 2010 IEEE International Conference on (pp. 544–549).
[11] Eckmann JP, Kamphorst SO, Ruelle D. Recurrence plots of dynamical systems. Europhys Lett 1987; 4: 973.
[12] Escudero J, Abásolo D, Hornero R, Espino P, López M. Analysis of electroencephalograms in Alzheimer's disease patients with multiscale entropy. Physiol Meas 2006; 27: 1091.
[13] Fawcett T. An introduction to ROC analysis. Pattern Recogn Lett 2006; 27: 861–874.
[14] Glass L, Mackey MC. From clocks to chaos: the rhythms of life. Princeton, NJ: Princeton University Press 1988.
[15] Goldberger AL, Amaral LA, Hausdorff JM, Ivanov PC, Peng CK, Stanley HE. Fractal dynamics in physiology: alterations with

disease and aging. Proc Natl Acad Sci 2002; 99(suppl 1): 2466–2472.
[16] Goldberger AL, Peng CK, Lipsitz LA, Vaillancourt DE, Newell KM. What is physiologic complexity and how does it change with aging and disease? Authors' reply. Neurobiol Aging 2002; 23: 27–29.
[17] Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. In The theory of chaotic attractors. New York: Springer 2004: 170–189.
[18] Guzman-Vargas L, Ramírez-Rojas A, Angulo-Brown F. Multiscale entropy analysis of electroseismic time series. Nat Hazards Earth Syst Sci 2008; 8: 855–860.
[19] Hausdorff JM, Peng CK, Ladin ZVI, Wei JY, Goldberger AL. Is walking a random walk? Evidence for long-range correlations in stride interval of human gait. J Appl Physiol 1995; 78: 349–358.
[20] Hayashi Y. Space-time spectral analysis and its applications to atmospheric waves. J Meteor Soc Japan 1982; 60: 156–171.
[21] Hussain L, Aziz W, Alowibdi JS, et al. Symbolic time series analysis of electroencephalographic (EEG) epileptic seizure and brain dynamics with eye-open and eye-closed subjects during resting states. J Physiol Anthropol 2017; 36: 21.
[22] Jensen MH, Kadanoff LP, Procaccia L. Measuring the strangeness of strange attractors. Phys Rev A 1987; 36: 1409.
[23] Jeong J. EEG dynamics in patients with Alzheimer's disease. Clin Neurophysiol 2004; 115: 1490–1505.
[24] Kaplan D, Glass L. Understanding nonlinear dynamics. New York: Springer-Verlag 1995.
[25] Kennel MB, Brown R, Abarbanel HD. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A 1992; 45: 3403.
[26] Kugler PN, Turvey MT. Information, natural law, and the self-assembly of rhythmic movement. London: Routledge 1987.
[27] Lake DE, Richman JS, Griffin MP, Moorman JR. Sample entropy analysis of neonatal heart rate variability. Am J Physiol Regul Integr Comp Physiol 2002; 283: R789–R797.
[28] Manis G. Fast computation of approximate entropy. Comput Meth Prog Bio 2008; 91: 48–54.
[29] Morabito FC, Labate D, La Foresta F, Bramanti A, Morabito G, Palamara I. Multivariate multi-scale permutation entropy for complexity analysis of Alzheimer's disease EEG. Entropy 2012; 14: 1186–1202.
[30] Oscar-Berman M, Marinković K. Alcohol: effects on neurobehavioral functions and the brain. Neuropsychol Rev 2007; 17: 239–257.
[31] Pan YH, Lin WY, Wang YH, Lee KT. Computing multiscale entropy with orthogonal range search. J Mar Sci Technol 2011; 19: 107–113.
[32] Porjesz B, Rangaswamy M. Neurophysiological endophenotypes, CNS disinhibition, and risk for alcohol dependence and related disorders. ScientificWorldJournal 2007; 7: 131–141.
[33] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000; 278: H2039–H2049.
[34] Rosen R. Fundamentals of measurement and representation of natural systems vol. 1. New York: Elsevier Science Ltd 1978.
[35] Shannon CE. The mathematical Theory of Communication. Urbana 1949.
[36] Silpa-Anan C, Hartley R. Optimised KD-trees for fast image descriptor matching. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on (pp. 1–8).
[37] Snodgrass JG, Vanderwart M. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J Exp Psychol Hum Learn 1980; 6: 174.
[38] Sun Y, Ye N, Xu X. EEG analysis of alcoholics and controls based on feature extraction. In Signal Processing, 2006 8th International Conference on vol. 1, 2006.
[39] Takens F. Detecting strange attractors in turbulence. In Dynamical systems and turbulence, Warwick 1980. Berlin Heidelberg: Springer, 1981: 366–381.
[40] Tashman LJ, Lamborn KR. Ways and means of statistics. New York: Harcourt Brace Jovanovich 1979.
[41] Tcheslavski GV, Gonen FF. Alcoholism-related alterations in spectrum, coherence, and phase synchrony of topical electroencephalogram. Comput Biol Med 2012; 42: 394–401.
[42] Webber Jr, CL, Zbilut JP. Assessing deterministic structures in physiological systems using recurrence plot strategies. In: Bioengineering approaches to pulmonary physiology and medicine. New York: Plenum Press 1996: 137–148.
[43] Webber Jr, CL, Zbilut JP. Recurrence quantification analysis of nonlinear dynamical systems. Tutorials in contemporary nonlinear methods for the behavioral sciences 2005; 26–94.
[44] Yates FE. Self-organizing systems: the emergence of order. New York: Plenum Press 1987.
[45] Zhang XL, Begleiter H, Porjesz B, Wang W, Litke A. Event related potentials during object recognition tasks. Brain Res Bull 1995; 38: 531–538.
[46] Zhang XL, Begleiter H, Porjesz B, Litke A. Electrophysiological evidence of memory impairment in alcoholic patients. Biol Psychiat 1997; 42: 1157–1171.
[47] Zoethout RW, Delgado WL, Ippel AE, Dahan A, van Gerven J. Functional biomarkers for the acute effects of alcohol on the central nervous system in healthy volunteers. Br J Clin Pharmacol 2011; 71: 331–350.