# Detection and Classification of Alcoholics Using Electroencephalogram Signal and Support Vector Machine

[1]Shaymaa Adnan Abdulrahman , Rafah Amer Jaafar [1]

[1]Department of computer Engineering, Imam Ja'afar Al-Sadiq University, Baghdad, Iraq
Shaymaaa416@gmail.com; rafah_amer@ yahoo.com

**Abstract**

Alcoholism may be recognized with the use of (EEG) analyzing signals. None-the-less, the analysis of the multi-channel signals of EEG is a complicated issue that usually needs performing complex computation operations and takes quite a long time to execute. The presented research will propose 13 optimal channel to feature extraction. In this research, an innovative horizontal visibility graph entropy (HVGE) method has been proposed for evaluating signals of EEG from controlled drinkers and alcoholic subjects and comparing against an approach of sample entropy (SaE). Values of HVGE and SaE have been obtained from 1200 records of bio-medical signals. While in classification step using SVM as classifier.

**Keywords:** Alcoholics, Support Vector Machine, Using Electroencephalogram Signal,Sample Entropy, Classification.

## 1.Introduction

Alcoholism can be defined as one of the common neurological disorders that result from the combined impact of environmental and genetic factors. It results in prominent harms to the brain system and in addition to that, results in impairments in mobility and cognition [1]. Electro-encephalogram (EEG) is one off the very efficient tools for researching the complicated brain activity dynamics. It is capable of visualizing complicated activities of the brain in the form of dynamic outputs. Which is why, it may be utilized for distinguishing normal subjects from alcoholics according to differences in signals, and that is helpful in detecting and diagnosing alcoholic subjects. Time-domain approaches and Frequency-domain analysis have been commonly utilized for the assessment of alcoholic EEG or EOG signals. Hayden et al. [2] have stated that the signals of the brain may be obtained with the use of a variety of methods. It may be categorized to invasive and non-invasive techniques. Invasive methods require surgical interference for

Received: March 15, 2020 Revised: May 10, 2020 Accepted: June 29, 2020

DOI: 10.5281/zenodo.3943272

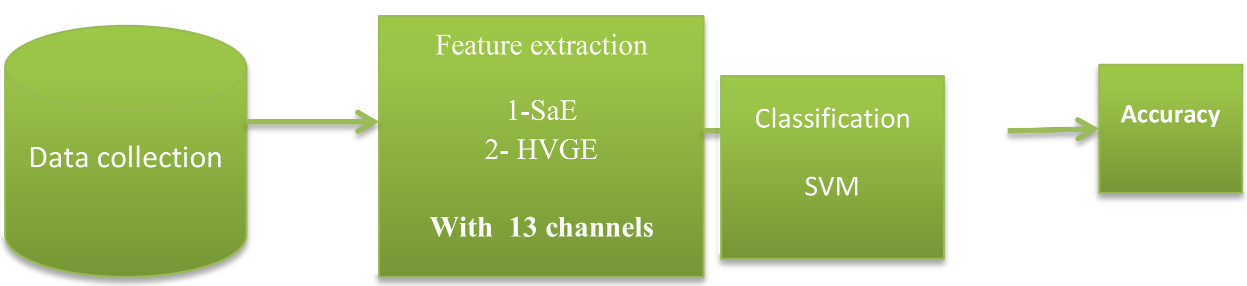
putting the electrodes under scalp. As a result of the medical risks, the experts prefer avoiding invasive method. While when using noninvasive method the electrodes are placed on the scalp of human [3].

Alcoholism the important of BCI application for medical lies detection/ prevention in the probable defeat of job and decrease of attentiveness level resulting from smoking and alcohol drinking, researcher try to exposed the upper reacting brain part to alcoholism [4]. Alcoholism can be defined as one of the common neurological disorders that result from the combined impact of environmental and genetic factors. It results in prominent harms to the brain system and in addition to that, results in impairments in mobility and cognition [5]. Signals of EEG are considerably beneficial to diagnose the brain and for analyzing anomalies and mental disorders. Some prominent researches in which detecting and classifying alcohol from data of EEG is performed have been explained as follows: classifying signals of EEG according to the drunk and non-drunk categories of people has been carried out by Ziya et al [6]. According to (Lin and Xiong) [7] Support Vector Machines (SVMs) and Independent Component Analysis (ICA) have been utilized to classify alcoholism . EEG signals have been analyzed for alcoholic and epileptic samples using methods of Auto Regressive (AR) modelling that have been advanced by( Faust et al) [8]. Automated diagnosis and detection of normal EEG signals vs. alcoholic EEG signals has been performed by( Acharya et al) [9]. EEG alcoholic and control subjects have been categorized using machine learning ensemble approaches by( Hussain et al) [10] .

The maim objective of this study is analysis alcoholic EEG and its impact on the diagnosis of ahuman brain . Two methods used to extract data such as sample entropy and HVG and to find which one is the best way to get hogher accuracy if using SVM as classifier. The structure of the paper is organized as follows. Section 2 Materials and Methods , Data Acquisition process, Feature Extraction, Sample entropy, HVG, Classification Process 3. Section Result & Analysis , Section 4 Conclusion.

**2. Materials and Methods**

Many stage was suggested in this study . The Figure 1 refer to proposed research methodology of this work .




DOI: 10.5281/zenodo.3943272

Figure 1: Diagram of Proposed Search

**2.1 Data Acquisition process**

In this paper ,the data set were obtained from the University of California, Irvine Knowledge Discovery in Data-bases Archive UCI KDD [11]. The dataset has been gathered from 122 individuals. Every one of those individuals completed 120 trials with 3 stimuli types [12]. Records from each one of the subjects includes 61 channel signals of EEG, 2 channels of EOG and a single reference electrode. The rate of sampling of all of the channel data is 256Hz, and each trial's duration is 1 sec. There are respectively 3 data-sets that are SMNI_CMI_TRAIN, SMNI_CMI_TEST and FULL. In the present research, only the first 2 data-bases are utilized due to the fact that FULL data-sets include a few all-0 recordings [13] In this paper using 13 channel like(FPZ,APZ,FZ,FCz,Cz, C1,C2,C3, C4,FC1,FC2,FC3,FC4) as feature extraction . These channels are distributed according to electrode placement in the standard 10 -20 system.

**2.2 Feature extraction**

**2.2.1 Sample entropy**

Sample entropy, due to the fact that it is related to the dynamic systems, is a reflection of information production rate. The Approximate entropy (ApEn) which has been presented by Pincus, is a collection of measurements of the complexity of the system that is applied easily on clinical cardio-vascular as well as other time series [14]. Sample entropy has been suggested by Richman and Moorman [15] It was utilized for measuring alcoholic EEG complexity [16]. The Sample entropy algorithm has 3 parameters of the input (i) m, which represents embedded dimension, (ii) r, which represents the criterion of similarity, (iii) n which represents the time series length. The sample entropy can be define in the equation 1 .

$$\text{Sample Entropy (m,r)}= \ln\left[ \frac{B^m(r)}{A^m(r)} \right] \ldots\ldots\ldots (1) \text{ where}$$

$$A^m(r) = \frac{1}{N-m}\sum_{k=1}^{N-m} A_k^m(r) \quad \ldots\ldots\ldots\ldots (2)$$

$$B^m(r) = \frac{1}{N-m}\sum_{k=1}^{N-m} B_k^m(r) \quad \ldots\ldots\ldots\ldots (3)$$

$B^m$ (r) represents the possibility of the 2 streams matching for m points

$A^m$ (r) represents the possibility for the 2 streams in matching for the m+1 points [17] and

$$B_k{}^m(r) = \frac{1}{N-m-1} B_K$$

Whe re considering N points of data from a time series $\{x(n)\} = x(1), x(2), \ldots .\, x(N)$.

Considering m vectors

$$x_m(1), \ldots .\, x_m(N - M + 1) \text{defind as } (i) = [x(k), x(k+1) \ldots x(k+m-1)] \text{ For} 1 \leq k \leq N - m + 1. \text{ at the ith sample [17].}$$

There are many, types of approaches for l graph entropy computation according to either, the edges or, the vertex [18] the present paper, know the graph entropy (GE) with Shannon's formula of entropy,[19]

$$h = -\sum_{i=1}^{n} p(k) \log(p(k)) \quad \ldots\ldots\ldots \quad (4)$$

### 2.2.2 Horizontal visibility graphs (HVG)

This approach is utilized for the construction of a matrix of synchronization of the network of the brain according to the real signals of EEG for disabled and normal subjects [20]. The graph of visibility can be represented in the following way, $\{x_i\}$ i=1. .N be an N data time series. What is referred to as the algorithm of horizontal visibility Luque et al.[21]. Can assign every element of the series of data to one of the nodes in the horizontal graph of visibility (HVG from now on). Nodes i and j in the graph are linked in the case where one is capable of drawing a horizontal line in the time series, which joins $x_i$ with $x_j$ that doesn't intersect any of the intermediate data heights. As a result, i and j are 2 linked nodes, in the case where the following geometric condition is satisfied in the time series:

$$x_i \;,\; x_j > x_n \text{ for each n in a way that, } i < n < j \ \ldots\ldots.(5)$$

The needed processed for determining the synchronization based on HVG have been illustrated in Figure 2

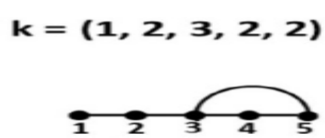



Figure 2: Represent the signal

### 2.3 Classification Process

SVM is a widely used classifier utilizing the method of supervised machine learning [22]. SVMs can be defined as one of the discriminative classifiers which is defined formally with a separating hyper-plane that may be utilized effectively for classifying a feature set of high dimensions. The hyperplane has to be in a way that it can maximize inter-class separations. It is also one of the supervised models of learning in which predetermined values of the class are utilized for the training[23]. The data of EEG might as well be turned into a mapping of higher dimensions. The function of mapping may be obtained with the use of the functions of Kernel. The most common one of the functions of kernel is the kernel of RBF [24]. The obtained width

DOI: 10.5281/zenodo.3943272

is helpful in deciding the margin of the SVM. In the case of considering data non-linearity, SVMs impose a parameter of trade-off for the separation of this type of data, as a result, allows margin flexibility [25].

## 3. Results and Analysis

The calculation feature are then nourished to the classifier for characterization between different states of the human brain. In this study had many objective 1) collect data t from by scraping a website and extracting information like UCI KDD data set 2) Preparing the input data 3) Analyzing the input data by using algorithms such as sample entropy and Horizontal visibility graphs 4) Train and testing the algorithm and after that Evaluating and the result generated by the proposed approach. We have utilized SVM classifiers in our work. Using thirteen channels with SaE as feature extraction shown figure 3 . While figure 4 refers to analysis of alcoholic EEG based on Horizontal visibility graphs and represent time of alcoholic EEG shown figure 5. The accuracy with SaE of 13 channels are 80.6% and with HVG are 91.2% . The experimental results in this study , HVG is higher reliability than SaE in the analysis of the EEG signals. Table 1 refer to accuracy when applied two different methods to extract the information when used SVM as classification .

Table 1 comparison the accuracy of the different in two methods

| Classification | Feature extraction | Accuracy |
|---|---|---|
| SUPPORT VECTOR MACHINE (SVM) | SaE | 80.6% |
| | HVGE | 91.2% |

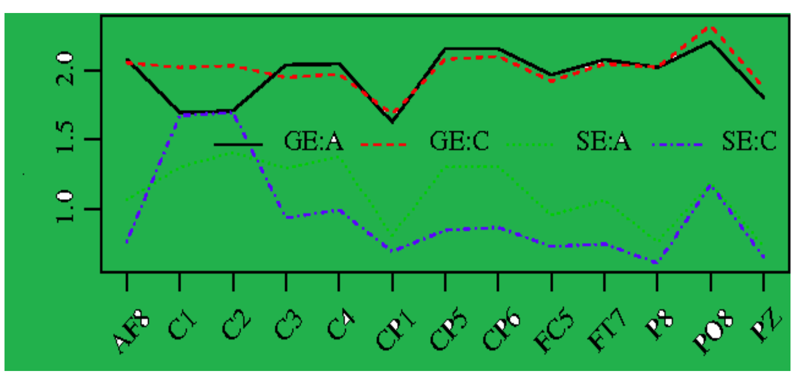



Figure 3 analysis of alcoholic EEG based on sample entropy and graphic entropy

DOI: 10.5281/zenodo.3943272

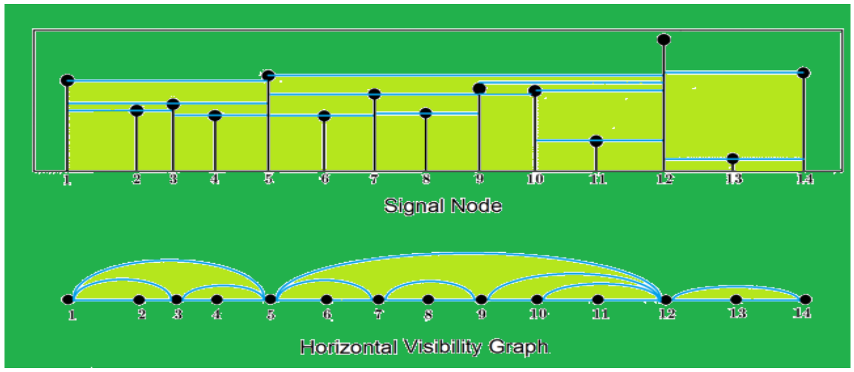



Figure 4 represent of alcoholic EEG based on Horizontal visibility graphs

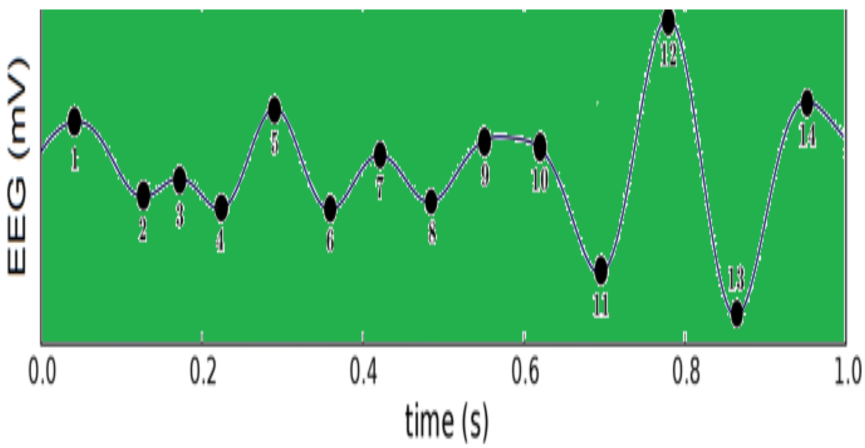



Figure 5 time(s) of alcoholic EEG

**4- Conclusion**

In this study , we investigated the EEG signal and functional human brain networks features for the automatic classification of alcoholic EEG signals. Thirteen optimal channels was applied taken from UCI KDD data set . there are many stages was applied such as (collected data , feature extraction and classification). Sample entropy and Horizontal visibility graphs used as feature extraction. While support vector machine applied as classifier . through this study the classification with Horizontal visibility graphs is more reliability than sample entropy . To increase more accuracy , we can use more EEG channels or use different signals such as EMG ,BCG , to analyze alcoholic EEG

**References**

[1]Abdulkader, S.N., Atia, A., Mostafa, M.S.M.: Brain computer interfacing: applications and challenges. Egypt. Inf. J. 16, 213–230 ,2015

DOI: 10.5281/zenodo.3943272

[2]Acharya UR, Subburam VS, Chattopadhya S, Suri J, ‘ Automated diagnosis of normal and alcoholic EEG signals’, International Journal of Neural systems, June, pp;1-9, 2012

[3]Bache K, Lichman M (2013) UCI machine learning repository. University of California, Irvine, School of Information and Computer. http://archive.ics.uci.edu/ml

[4]Faust O, Acharya UR, Alen A and Lim CM, Analysis of EEG signals during epileptic and alcoholic states using AR modeling techniques, Innovations and Technology in Biology and Medicine (ITBM-RBM) 29(1) 44-52, 2008

[5]Hayden EP, Wiegand RE, Meyer ET, Bauer LO, O'Connor SJ, Nurnberger JI, Chorlian DB, Porjesz B, Begleiter H Patterns of regional brain activity in alcohol-dependent subjects. Alcohol: Clin Experiment Res 30(12):1986–.2006, 1991

[6]Hussain L, Aziz W, Khan AS, Abbasi AQ, Hassan SZ, Abbasi MM, ‘Classification of EEG Alcoholic and Control subjects using machine learning ensemble methods’, Journal of Multidisciplinary Engineering Science and Technology, vol 2, issue 1, , pp;126- 131.2015

[7]Lin C, Xiong S, ‘ Alcoholism classification based on ICA and SVM Method, Project Report, Rutgers University, pp;1-7.

[8]Oscar-Berman M, Marinkovi K Alcohol: effects on neurobehavioral functions and the brain. Neuropsychol Rev 17(3):239–257, 2007

[9]Richman JS, Moorman JR (2000) Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol-Heart Circulat Physiol 278(6):H2039–H2049

[10]Shannon CE (2001) A mathematical theory of communication. ACM SIGMOBILE Mob Computer Commun Rev 5(1):3–55

[11]Shaymaa Adnan Abdulrahman, Wael Khalifa, Mohamed Roushdy, Abdel-Badeeh M. Salem " A survey of biometrics using electroencephalogram EEG " International Journal "Information Content and Processing", Volume 6, Number 1, 2019

[12]Shaymaa Adnan Abdulrahman, Mohamed Roushdy, Abdel-Badeeh M. Salem " Human Identification based on electroencephalography Signals using Sample Entropy and Horizontal Visibility Graphs " WSEAS TRANSACTIONS on SIGNAL PROCESSING , ISSN: 2224-3488 Volume 15, 2019

[13]Shaymaa Adnan Abdulrahman, Mohamed Roushdy, Abdel-Badeeh M. Salem "support vector machine approach for human identification based on EEG signals , journal of mechanics of continua and mathematical sciences , ISSN (Online) : 2454 -7190 Vol.-15, No.-2, pp 270-280 ISSN (Print) 0973-8975, February 2020

[14]Shaymaa Adnan Abdulrahman, Wael Khalifa, Mohamed Roushdy, Abdel-Badeeh M. Salem" Comparative Study for 8 Computational Intelligence Algorithms for Human Identification" Computer Science Review Journal, Vol 36 , 2020

[15] Mohammed, M.A., Al-Khateeb, B., Rashid, A.N., Ibrahim, D.A., Abd Ghani, M.K. and Mostafa, S.A., 2018. Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. Computers & Electrical Engineering, 70, pp.871-882.

[16] Mohammed, M.A., Abd Ghani, M.K., Arunkumar, N.A., Hamed, R.I., Abdullah, M.K. and Burhanuddin, M.A., 2018. A real time computer aided object detection of nasopharyngeal carcinoma using genetic algorithm and artificial neural network based on Haar feature fear. Future Generation Computer Systems, 89, pp.539-547.

[17] Mohammed, M.A., Abd Ghani, M.K., Hamed, R.I., Ibrahim, D.A. and Abdullah, M.K., 2017. Artificial neural networks for automatic segmentation and identification of nasopharyngeal carcinoma. Journal of computational science, 21, pp.263-274.

[18]Zhu G, Li Y, Wen P (2011) Evaluating functional connectivity in alcoholics based on maximal weight matching. J Adv Computat Intell Intell Inform 15(9):1221–1227, 2011

[19]Ziya E, Akif A, Mehmet RB, 'The Classification of EEG signals recorded in drunk and non-drunk people', International Journal of Computer Applications, vol.68, no.10, pp;40-44, 2013

[20]Zou Y, Miao D, Wang D (2010) Research on sample entropy of alcoholic and normal people. Chin J Biomed Eng 29:939–942

[21] Mohammed, M.A., Abd Ghani, M.K., Arunkumar, N.A., Mostafa, S.A., Abdullah, M.K. and Burhanuddin, M.A., 2018. Trainable model for segmenting and identifying Nasopharyngeal carcinoma. Computers & Electrical Engineering, 71, pp.372-387.

[22] Abd Ghani, M.K., Mohammed, M.A., Arunkumar, N., Mostafa, S.A., Ibrahim, D.A., Abdullah, M.K., Jaber, M.M., Abdulhay, E., Ramirez-Gonzalez, G. and Burhanuddin, M.A., 2020. Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques. Neural Computing and Applications, 32(3), pp.625-638.

[23] Obaid, O.I., Mohammed, M.A., Ghani, M.K.A., Mostafa, A. and Taha, F., 2018. Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer. International Journal of Engineering & Technology, 7(4.36), pp.160-166.