Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 1, 2024, pp. 8-19

DOI: https://doi.org/10.25195/ijci.v50i1.438
Print ISSN: 2313-190X, Online ISSN: 2520-4912

Research Article

Computer Based Detection of Normal and Alcoholic Signals Using Discrete Fourier Transform

[1] shaymaa adnan abdulrahman
University of Information Technology & Communications
College of Business Informatics Technology,
Baghdad, Iraq
dr.shaymaa.adnan@Uoitc.edu.iq

[2]Mohamed Roushdy
Ain Shams University,
Faculty of Computer & Information Science
Egypt, Cairo
mroushdy@cis.asu.edu.eg

[3]Abdel-Badeeh M.Salem
Ain Shans University
Dept of computer Science
Egypt, Cairo
absalem@cis.asu.edu.eg

ARTICLEINFO
Article History
Received: 28/11/2023
Accepted: 01/02/2024
Published: 01/06/2024

ABSTRACT

Alcoholism is a severe, disorder that affects; the functionality of neurons in the central nervous system and leads to the loss of. health and wealth. The suggested technique applies statistical and fractal dimension (FD) features to classify alcoholic and normal subjects using eight channels under an SF-based machine learning architecture. Electroencephalogram (EEG) signals are placed in a framework and separated into different EEG bands using an orthogonal wavelet filter. The following three classification approaches are used to classify the alcoholic and normal patterns of EEG data: least-square support vector machine, vector machine (SVM), and Naïve Bayesian. Results showed that the best classification method was SVM with a sensitivity of 0.9267%, an accuracy of 0.9892%, and a specificity of 0.9916%.

Keywords: *alcoholism, electroencephalogram, naïve Bayesian, SVM.*

1. INTRODUCTION

Alcoholism is one of the most typical addiction problems in the world [1], accounting for at least 5% of the global disease burden. According to the WHO's global status report, the use of alcohol has risen in emerging nations such as India. According to a survey, at least 3 million individuals have died as a result of alcohol poisoning/overdose, 75% of which are men. Excessive alcohol use has negative impacts on human organs, such as the brain [2]. In extreme cases, individuals become physically reliant on alcohol. If not treated and diagnosed early, alcoholism can result in physical and mental problems and mortality. Early identification of alcoholism could aid individuals in being aware and preventing long-term damages. Conventional approaches for measuring the impact of alcohol on a certain subject, such as blood tests, questionnaires, and physiological testing, have been utilized to detect alcoholism. Each of this method has benefits and drawbacks. Owing to the secretive nature of people regarding alcoholism, questionnaire-based analysis provides less accuracy. By contrast, a blood test is not just intrusive and uncomfortable but also unreliable [4]. Electroencephalogram (EEG) is a noninvasive diagnostic test used to examine the brain's electrical activity and assess a person's neuropsychiatric state. Several electrodes are placed on the subject's scalp to record electrical activities in the brain. However, EEG signals have a nonlinear and nonstationary nature. Hence, manual signal screening is time consuming and inaccurate. The two different forms of messages in our brains are chemical and electrical. Our emotions, bodily responses, and mood are all controlled by neurotransmitters, chemical messengers that communicate between neurons. The chemical signal is.

The two different forms of messages in our brains are chemical and electrical. Our emotions, bodily responses, and mood are all controlled by neurotransmitters, chemical messengers that communicate between neurons. The chemical signal is balanced under typical circumstances. Alcohol consumption causes a chemical signal imbalance, slowing

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 1, 2024, pp. 8-19

DOI: https://doi.org/10.25195/ijci.v50i1.438
Print ISSN: 2313-190X, Online ISSN: 2520-4912

down our response. The electrical signal measured by electrodes on the scalp changes with the chemical signal. Electrical activity over a short period of time can be measured by brain signals or EEG signals. Given its potential to represent the condition of the complete body, electrical activity encourages the employment of digital signal processing techniques. Current flows within the tips of dendrites and axons and between distinct dendrites, generating an electrical signal (measured in mV).

Machine learning and digital signal processing methods are employed to determine whether a signal is indicative of alcoholism. Digital signal processing uses recorded brain activity data (EEG), which are split into subbands (SBs) for signal reconstruction. EEG signals offer a waveform that can be observed and are processed and described using computer-assisted digital signal processing methods and machine learning methods, respectively. Microvolt (v) measurements of EEG signals are presented as a timed series. EEG signals have a number of frequency components, including theta, delta, beta, alpha, and gamma.

Figure: 1 gives a brief overview of EEG signal SBs. The voltage fluctuations created via electrodes in EEG signals are relatively small (in v). Manual EEG signal screening is challenging and sometimes leads to incorrect results. The number of delta and theta SBs is higher in an alcoholic's EEG signals than in normal, nonalcoholic EEG signals. Bauer [3] found an increase in beta amplitude in alcoholic individuals ranging from 19.5 Hz to 39.8 Hz. Examining the power spectrum of nonalcoholic and alcoholic EEG data aids neurophysiological evaluation. However, frequency or time domain analysis is extremely challenging because of the nonlinear structure of EEG signals. The following time–frequency domain signal processing methods are used to interpret EEG signals: short-time Fourier transform, Fourier transform (FFT) [2], Wigner–Ville distribution, wavelet transform (WT) [14], and autoregressive method, Hilbert–Huang transform, time–frequency distribution, and eigenvector methods. Different machine learning (ML) and feature extraction approaches are used to automate EEG signal screening. The collected features are fed into different classifiers as input data [5]. Many ML and feature extraction approaches for automated EEG signal processing [6] alcoholism diagnosis [1], [3] have been explored. ML and digital signal processing are employed to determine if a signal is alcoholic or not. The brain's recorded data (EEG) are subjected to digital signal processing, which divides the data into SBs to reconstruct the signals. EEG signals give a visible waveform display and are characterized using ML and computer-aided digital signal processing. The ability of a medical expert to perform realistic automatic EEG data analysis is critical in the diagnosis of neurobiological diseases. We can avoid the danger of information misinterpretation by adopting computer-aided diagnosis (CAD). However, the automatic classification of recorded signals is a complicated task because their temporal and morphological properties vary significantly for various patients in different physical and temporal conditions. Increased quantization levels and high sampling rate are needed for accurate signal analysis.

The remainder of this research is structured as follows: Section 2 introduces the problem statement. Section 3 shows the materials and methods. Section 4 presents the experimental data. Section 5 contains the results and discussion, and Section 6 offers the conclusion.

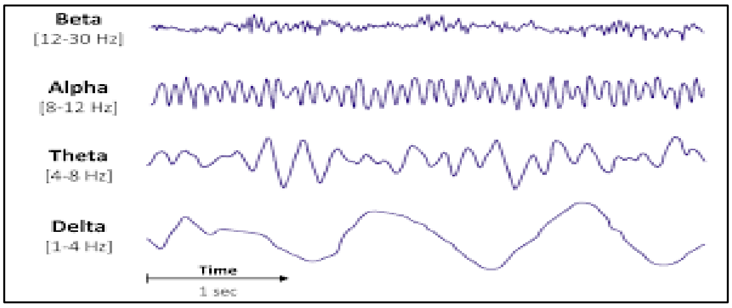

Fig1. Frequency bands in EEG signals.

2. PROBLEM STATEMENT

Alcoholism is a socioeconomic syndrome that can cause a person to lose their health and money. A WHO study reports 2 billion drinkers worldwide, 76.3 million of whom have alcohol-dependent syndrome. A total of 58.3 million individuals (4.8% of the population) live with a disability, and drinking is responsible for 1.8 million deaths (3.2% of all fatalities). Alcoholism has a negative impact on a person's brain functions. Given that the human brain is complicated, determining alcoholism effects on it is difficult [15]. In light of the severity of this illness, a cost-effective, precise, and reliable system to distinguish between alcoholics and nonalcoholic is urgently needed. Here, a strategy offering the best channels that are essential and sufficient to detect and predict alcoholism is suggested [11].

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 1, 2024, pp. 8-19

DOI: https://doi.org/10.25195/ijci.v50i1.438
Print ISSN: 2313-190X, Online ISSN: 2520-4912

- Which aspects are the most potent, and which need extraction and which don't?
- Which feature combination produces the greatest outcomes across all datasets that are currently available?
- Which method is the most effective for differentiating between normal and alcoholic EEG data patterns?

3. MATERIALS AND METHODS

Figure 2 depicts the proposed scheme's operational flow. The UCI dataset contains two types of EEG signals recorded from patients: alcoholics and healthy individuals [11]. Preprocessing is the first step in filtering that removes unwanted noises from the recorded EEG signal. The resulting noise-free EEG data are then split into several SBs using an orthogonal wavelet filter bank. For each SB, the LEs are calculated and employed as isolation features. Least square support vector machine (LS-SVM), support vector machine (SVM), and Naïve Bayesian (NB) algorithm are used for categorization.

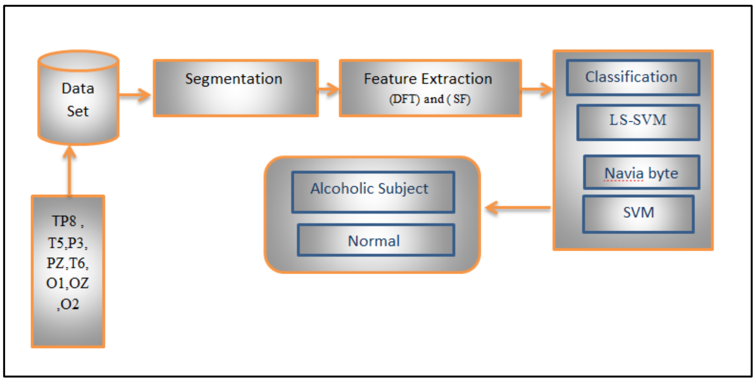

Fig. 2. Block diagram of the proposed method for alcoholism detection using EEG signals

3.1 EEG recordings

The publicly available dataset of Knowledge Discovery in Databases (kdd.ics.uci.edu/databases/eeg/eeg.data.html) from the University of California, Irvine (UCI) is used in the experiments [7]. This dataset is the result of a long study of genetic predispositions among alcoholic participants and is divided into two groups: normal and alcoholic subjects. It has three different versions. The first version is small data with just two subjects: one for all the classes and one for all the stimuli. In this small dataset, two stimulus types are utilized: single ($S_i 1$) and dual stimuli ($S_i 1$ match S_i 2 and $S_i 1$ do not match $S_i 2$). The second dataset is large and includes 10 subjects from each class, each with their own testing and training data. The third dataset represents a complete dataset with 122 people and 120 trials. Data are recorded for every person utilizing 90 images of diverse stimuli. Given that the publicly accessible dataset (with 122 participants) is not complete and a few trials contain empty files or labeled as "err," the current study uses the small dataset. The 10/20 international montage is used for fixing each electrode's position. Cz is employed as a reference as indicated in Fig. 3, and grounding is carried out by a nose electrode with impedance that is less than 5 k [9] [10]. The vertical and horizontal bipolar electrooculography (EOG) signals are recorded using the Y and X electrodes, respectively. The EEG signals are captured for 32 seconds (approximately 16,400 samples) at a sampling rate of 256 Hz and a resolution of 12 bits. The baseline filter is used to remove artifacts such as eye and muscle movements (>73.3v) from the short dataset prior to the experiments. After preprocessing, only 30 EEG recordings remain from the two classes. The large 32-second EEG recordings are divided into 8-second windows with four equal 2048-sample samples. Figures 4 and 5 depict one trail of alcoholic and normal subjects. The amplitude of a normal subject is greater than that of an alcoholic subject because the mind of the normal subject is awakened and agitated. Meanwhile, alcoholic patients are considerably less conscious. Figures 4 and 5 provide the examples of data segments for nonalcoholic and alcoholics, respectively.

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 1, 2024, pp. 8-19

DOI: https://doi.org/10.25195/ijci.v50i1.438
Print ISSN: 2313-190X, Online ISSN: 2520-4912

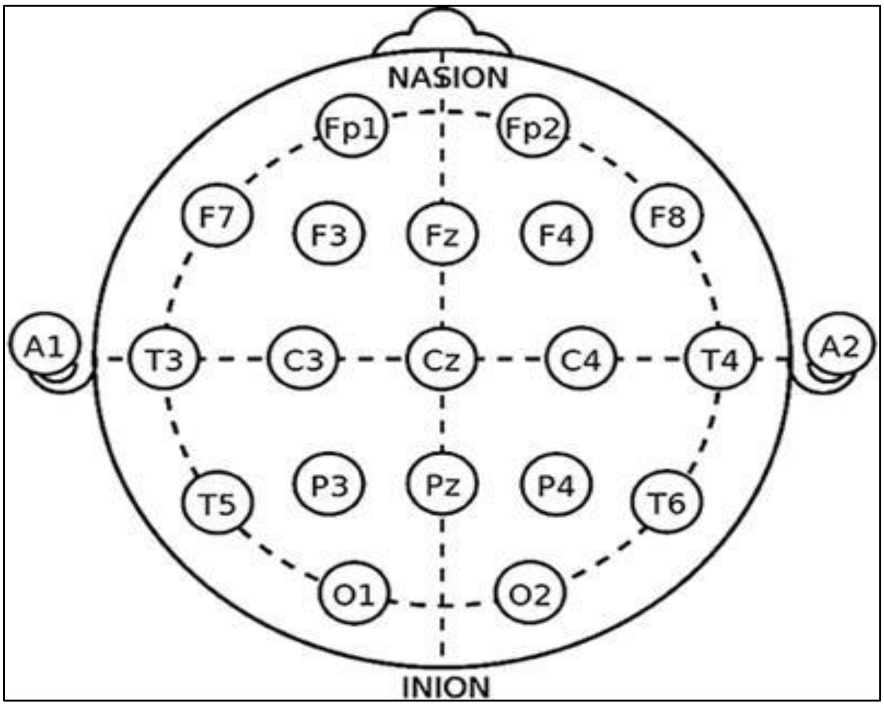

Fig. 3. Each electrode using standard [10/20]. Scheme for record EEGs [11] [12].

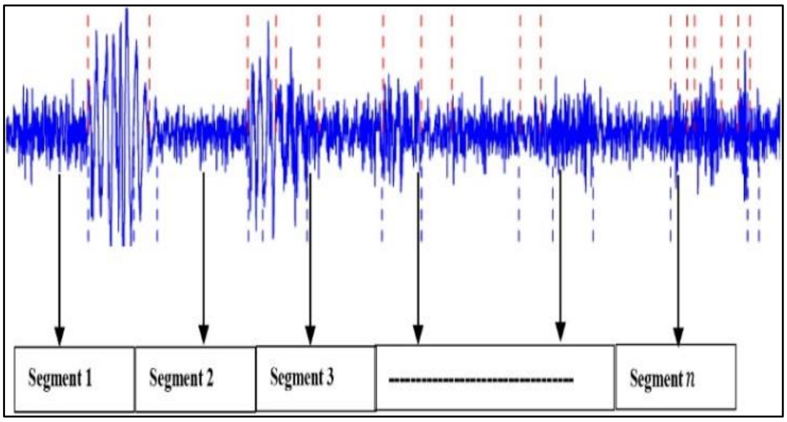

Fig. 4. Alcoholic EEG signal segment [11].

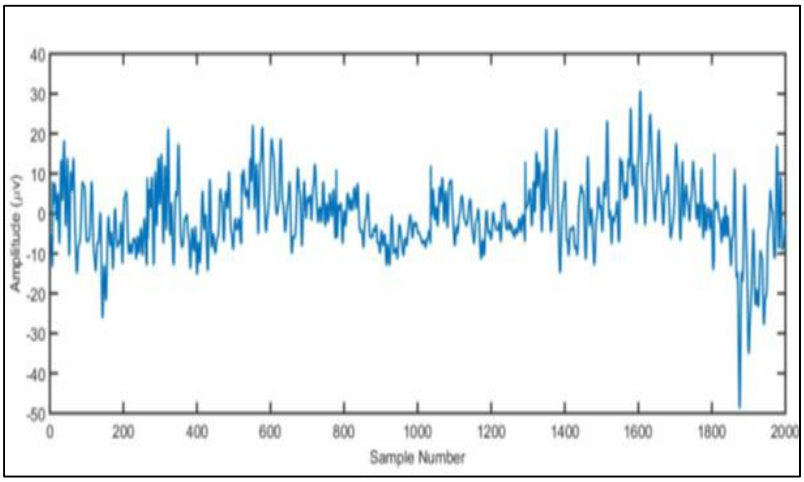

Fig. 5. Normal EEG signal segment [11].

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 1, 2024, pp. 8-19

DOI: https://doi.org/10.25195/ijci.v50i1.438
Print ISSN: 2313-190X, Online ISSN: 2520-4912

3.2 EEG data segmentation

The EEG data must be segmented because EEG signals change throughout time. Each EEG segment is partitioned to semistationary divisions by dividing it to small parts using the sliding window. Wen and Li (2011) [13] used the same approach to identify EEG signals for tracing the depth regarding anesthesia in EEG signals. Al-Salman etal. (2019) [14] evaluated various sliding window sizes for tracing and detecting the spindles of sleep in brain EEG signals. Lee and Mehmood (2015) [15] used various sliding window sizes for tracing the main waves in EEG signals and showed that this approach improves classification accuracy. Therefore, the sliding window method is utilized in the present work. Y is assumed to be EEG signal with X data points, with Y= {y.1, y.2, y.3... yx}. Every 30-second EEG portion is segmented to small T subsegments with t data points (T= {m1, m2, m3.,.. mt}). Throughout the training phase, the sliding window size is established empirically depending on a large number of experiments. The entire EEG signal is divided to small interval portions. A 0.5-second window is used, with a 0.4-second overlap. Figure 6 depicts the partitioning of an EEG signal to overlapping segments.

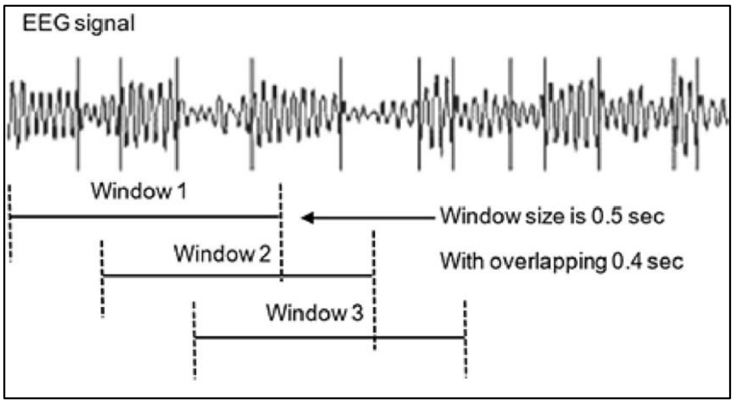

Fig. 6. The sliding window method [11].

3.2.1. Feature Extraction Methods

Raw EEG signals are noisy due to the presence of artifacts. Prior to feature extraction and analysis, the signals must be preprocessed. In this research, statistical model, fractal dimension (FD), nonlinear features, and frequency are used to extract representative features from each 0.5-second EEG segment using a multi domain technique depending on the EEG characteristics' analysis. Each segment yields 22 statistical features. Decreasing the features' dimensionality is a crucial step in reducing the technique's complexity and improving the results' performance. The relevant features representing each 0.5-second EEG segment are analyzed and selected using the selection features obtained from the statistical analyses in this work. The features of extraction are briefly demonstrated in the next subsection:

3.2.2. FD features

The FD approach works well for evaluating the features of EEG signals. In this work, Katz's approach is used to estimate and evaluate the FD of EEG data from each segment. In 1988, Katz made the algorithm's initial suggestion. This method can be used to calculate the maximal distance between the curve's initial point and any other point, given by the symbol d, and the curve's length. The total of the Euclidean distances between all of the points that make up curve L is used to determine its length. The ratio of the whole curve length to the line for the maximization Euclidean distance is applied to compute the maximization Euclidean distance from the curve's first point line. Katz's technique is used to estimate the FD of signals (Dutt & Raghavendera, 2009; Katz, 1988; Ali et al., 2016) [10]. In this study, FDs are computed for each epoch using Katz's method. The FD value ranges from 1 to 2.

$$FD = \frac{Log_{10}(n)}{Log_{10}(n) + Log_{10}\left({}^{d}/_{L}\right)} \ldots\ldots (1)$$

where n_i is the number of curve steps calculated as

$$n_i = \frac{L}{ai} \ldots\ldots\ldots\ldots\ldots (2)$$

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 1, 2024, pp. 8-19
DOI: https://doi.org/10.25195/ijci.v50i1.438
Print ISSN: 2313-190X, Online ISSN: 2520-4912

where ai symbolizes the average distance between two following locations, L is the sum of the Euclidean distances between consecutive points or the overall length of the curve, and d is the radius of the curve. The following is a definition of these terms (Dutt & Raghavendera, 2009) [16]):

$$L = \sum_{j=1}^{n}[(x_{1j+1} - x_{1j})^2 + (y_{1j+1} - y_{1j})^2]^{1/2} \ldots\ldots\ldots..(3)$$

$$d = \max([x_{j+1} - x_{1j})^2 + (y_{j+1} - y_{1j})^2]^{1/2}$$

$$where\ j = 1,2,3,\ldots\ldots.(4)$$

3.2.3. Statistical features (SFs)

The two statistical metrics mean and standard deviation are computed as follows:

1 -Mean (M) eliminates random errors in the signals and helps obtain reliable results every time the signal is used in the feature's computation. It is computed as (5)

$$M = \frac{1}{N}\sum_{i=1}^{N} x_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(5)$$

where xi, N, and M stand for the sample size, overall number of time samples, and mean.

2-Standard deviation (SD) is computed using the following formula and is used to assess the degree of variance in EEG signals:

$$SD = \left(\frac{1}{n-1} \quad \sum_{i=1}^{n}(x_i - x^*)^2\right)^{1/2} \ldots\ldots\ldots\ldots(6)$$

SD= where x, x^*, and n represent the signal mean and. number of the present samples. The above features are retrieved for, feature vector, resulting in a new 28-dimensional feature vector. T-tests and p-values are utilized for calculating the statistical parameters such as SD and mean. A T-test is used to rank features depending on the results and to compute p-values that correspond to every one of the features. A low p-value suggests good feature separation

4. CLASSIFIER USED FOR AUTHENTICATION ANALYSIS

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

4.1 SVMs

Can be defined as a collection of the supervised ML techniques for analyzing data and identifying data structures for classification. They feature an excellent balance between the accuracy attained on limited training data and the generalization of the testing data [2]. For a classification model, the first few principal elements are given to the SVM classifier during the training phase with associated class labels. The kernel is a standard Gauss/radial kernel. The SVM classifier could be written as follows:

$$g(x) = w.\Phi(x) + b \ldots\ldots\ldots\ldots\ldots\ldots(7)$$

where xp represents the input vector from (from now onward subscript p refers to the input data projected on the principal components and obtained after the PCA). 'w' represents normal to separating hyperplane that is defined by Φ (x). For a group of data points $(xpi,)$ where $(y{:}= +1;-1)$, the margin between a pair of classes is given by $2\|w2\|$. The optimum margin can be estimated by solving the problem of constrained optimization by reducing it to quantization. programming optimization problem which produces [10]

$$w{=}\Sigma\alpha ini{=}1(xpi) \ldots\ldots\ldots\ldots\ldots.(8)$$

$$b{=}\Sigma\alpha inj{=}1yj\ (xpi).(xpj){+}\ yi,\forall\ I \ldots\ldots\ldots\ldots.(9)$$

The training samples xi with a nonzero Lagrange coefficient are support vectors. The substitution of values provides updated SVMs

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 1, 2024, pp. 8-19

DOI: https://doi.org/10.25195/ijci.v50i1.438
Print ISSN: 2313-190X, Online ISSN: 2520-4912

gt(x) = Σαini=1yi (xpi).Φ(xp) + b …………(10)

=Σαini=1yi K (xpi,) + b

Equation (10) represents the formal expression for the SVM classifier. K represents the radial/Gauss kernel function, which can be represented as:

(xpi ,j) = e||xpi−xpj||(2σ)^22 ………..(11)

Repeated trials proved that the data projected on the first two principal components sufficiently account for the whole data.

4.2 Naïve Bayes

Naïve Bayesian algorithm operates on the basis of the Bayesian theorem through the assumption that all features are independent for the prediction of a sample's class [2]. Naïve Bayes is probabilistic and computes the likelihood of all attributes. The maximum likelihood attribute is then utilized for the classification. Likelihoods are estimated using Bayesian theorem, expressing the likelihood of a feature based on prior knowledge. The main benefit of this classifier is that it takes considerably short duration for the testing and classification and insensitivity of independent features

4.3 LS-SVM

LS-SVM is a commonly utilized classifier in biomedical area, representing an extended SVM form. LS-SVM incorporates solutions of equality constraint that are acquired by the linear equation rather than the quadratic programming (QP) that is utilized in standard SVMs [2]. LS-SVM's objective function can be expressed mathematically by Equation (12).

$$y(t) = \text{sign} \sum_{i=1}^{N} y_{i\alpha_i} k(t, t_i + b) \ldots\ldots (12)$$

A detailed description of LS-SVM can be found in [21]. LS-SVM may be carried out using the kernels in Eq. (12) $K(t_1, t_i)$ represents the kernel function that can be replaced by other kernel functions such as the linear- $k(x1, y1) = (xi^T y + c)$. polynomial

$$k(x, y) = (\propto x^{iT} y + c)^{id} \ldots..(13).$$

5. EXPERIMENTAL DATA

The University of California, Irvine Knowledge Discovery in Data-bases Archive provided the experimental data [10]. The dataset included 122 participants and therefore is accurate. Each individual had completed 120 trials using three different stimuli [2]. In particular, 61 channel EEG signals, 1 reference electrode, and 2 EOG channels are all recorded from each patient. Every trial lasts 1 second, and the sampling rate for all channels of data is 256 Hz. The dataset has three versions—FULL, SMNI_CMI_TRAIN, and SMNI_CMI_TEST. In the present work, only the last two versions are utilized because the FULL dataset includes some all-zero recordings [13]. A total of 600 recorded files are included in SMNI_CMI_TRAIN, every one of which comprise signals from 64 electrodes caps. The features have been obtained from eight channels, namely, TP8, T5, P3, PZ, T6, O1, OZ, and O2. The sliding window is then applied for segmentation. A new method is deployed to extract features from each segment. Statistical parameters, such as SD and mean, are estimated using T-test and (p)-values. In the present work, the data are divided into 10 equal blocks. Forward training and testing are performed as follows: fold1: training, test; fold2: training, test; fold3: training [1 2 3], test [4]. The trained model undergoes validation on 1 (unknown) data block, prior to the testing on the following (unknown) block. This procedure is repeated 10 times to obtain the average results of classification. The following five performance indexes are calculated to measure the classifiers' performance: sensitivity (Sen), specificity (Spec), precision (Prec), accuracy (Acc), and F1 measure (F1):

$$S1^* = \frac{TP}{TP+FN} \qquad Sp1^* = \frac{TN}{TN+FP}$$

$$Ac1^* = \frac{TP+TN}{TP+TN+FP+FN} \qquad Pre1^* = \frac{TP}{TP+FP}$$

$$F1^* = 2 * \frac{Prec*Recall}{Prec*Recall} \qquad Re1^* = \frac{TP}{TP+FN}$$

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 1, 2024, pp. 8-19
DOI: https://doi.org/10.25195/ijci.v50i1.438
Print ISSN: 2313-190X, Online ISSN: 2520-4912

where TP stands for "true positive" occurrences, FP for "false positive" situations, TN for "true negative," and FN for "false negative." A graphical representation that depicts the trade-off between the true(+) positive rate (Sen) and the false(-) positive rate (1 - Spec) is known as the receiver operating characteristic (ROC) curve. This curve serves as an adequate mean for evaluating the effectiveness of a classifier. Meanwhile, the area under the ROC curve (AUC), the sixth performance indicator, is calculated using the following formula:

$$AUC = \int_0^1 Sen(x)dx \quad \ldots\ldots(14)$$

where $x = 1 - Spec$

6. RESULTS AND DISCUSSION

The experiments have two phases:

1. Classification of alcoholics using normal subjects

2. Choosing the best channel to look for alcoholism.

The dataset utilized in this work is outlined in Table 1. The important point in this study is to use eight channels, namely, TP8, T5, P3, PZ, T6, O1, OZ, and O2, instead of 61 channels to analyze brain signals to prove that these eight channels are optimal compared with the other channels. These channels show superiority in previous studies [11]. The current experiment uses the EEG of 100 alcoholics and 100 healthy human participants with ten trials each. Orthogonal wavelet filters are applied to separate EEG signals into several EEG bands. Absolute gamma band power is taken from each band for the control subject and the alcoholic. The preparation of EEG signal data, which comprises 32-second segments, is the initial step of the framework. Another 8-second time-series recording of duration is performed. A sampling frequency of 256 Hz is used during segmentation to create 120 trails with 2048 samples of each participant. Filters are used to eliminate artifacts that are present in the signal.

The sliding window is used to partition each EEG signal into a group of segments, with the sliding window size set at 0.5 seconds. EEG brain signals are processed using FD and statistical approaches. It aids in the decomposition of an input signal into its constituent sinusoidal signals of various frequency values. For reducing the background offset, the EEG signals are processed through high pass filter with a 0.16 Hz cut-off frequency before being applied to DFT. Throughout signal acquisition, the high-pass filter often filters away slow artefacts like movement artifacts and electro galvanic signals. A lowpass filters that has a cut off frequency that equals the maximum frequency of interest (40 Hz) is also utilized to guarantee that the signal is band-limited. DFT is applied across filtered data using the sliding window approach to eliminate the impact of data discontinuity at the two ends. The significance level of retrieved feature for the distinction of alcoholic and normal EEGs information is assessed through using statistical features. For statistical significance (p less than 0:001). P-values and t-tests are used for calculating statistical parameters such as SD and Mean. A T-test is used to rank features, and such features are ranked depending on results. The T-test is used as well for computing the p-values corresponding to every one of the features. A lower p-value suggests better feature separation. The Statistical Range (MEAN_ SD) of alcoholic and normal features is shown in Table 2. Navia byes, SVM, and LLS-SVM classifiers are used to classify alcoholic and normal features.

Table I. The Datasets Applied In This Work

Type	**ׅof electrode**	**EG-Segment**	**ׅotal Sample**	**ׅple Frequency**
Normal	8	240	2048	256Hz
Alcoholic	8	240	2048	256Hz

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 1, 2024, pp. 8-19
DOI: https://doi.org/10.25195/ijci.v50i1.438
Print ISSN: 2313-190X, Online ISSN: 2520-4912

Table Ii. The Statistical Range (Mean/Sd) Of Normal and Alcoholic Features

Electrodes	Alcoholic	Control	p. value
TP8	1.583±0.475	1.984±0.430	<0.01
T5	1.846±0.295	1.837±0.310	<0.001
P3	1.572±0.475	1.873±0.430	<0.06
PZ	1.672±0.465	1.947±0.432	<0.01
T6	1.623±0.465	1.984±0.420	<0.01
O1	1.493±0.475	1.994±0.431	<0.01
OZ	1.583±0.474	1.984±0.430	<0.01
O2	1.578±0.475	1.972±0.435	<0.001

The SVM classifier achieves the best classification results. As illustrated in Fig. 7, the confusion matrix shows the total number of occurrences that have been classified incorrectly and correctly. This tool aids in the evaluation of several indexes of the suggested classifiers, such as specificity, sensitivity, precision, accuracy, and F-measure

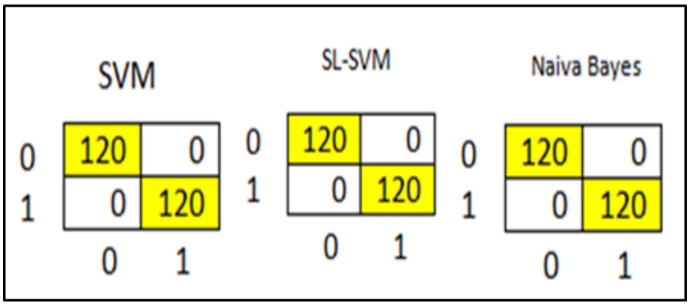

Fig. 7. Confusion matrices for SVM, LS-SVM, and naive Bayes

Table 3 shows that the SVM classifier has the best sensitivity (0.9767%), accuracy (0.9892%), specificity (0.9916%), and F-measure (0.9788%). Table 4 also indicates how the signal-to-noise (SNR) ratio affects the performance of the different classifiers. While the performance of the other classifiers falls as the SNR decreases, SVM achieves reasonable accuracy even at low SNR levels.

The dataset is randomly divided into four separate subdatasets, with (80%) used for training and the remaining (20%) for testing.

Table III. Classification Performance Using Svm, Lls- Svm, And Naive Bayes.

Electrode	Classifier	Accuracy	Sensitivity	Specificity	f-Score
TP8,T5,P3,PZ,T6 ,O1,OZ,O2	SVM	0.9892	0.9767	0.9916	0.9788
	LLS-SVM	0.8721	0.8527	0.8671	0.8402
	Naiva Bayes	0.8123	0.8734	0.8711	0.8390

Table IV. Effect Of Eeg Signal To Noise (Snr) Of Input Signal On The Accuracy (*%) Of Svm, L Ls-Svm, And Naïve Bayes Classifiers

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 1, 2024, pp. 8-19

DOI: https://doi.org/10.25195/ijci.v50i1.438
Print ISSN: 2313-190X, Online ISSN: 2520-4912

SNR	SVM	LS-SVM	Naiva Bayes
-10	86.3	81.3	80.8
-5	90.2	88.2	89.2
0	93.6	90.6	91.6
5	96.4	92.6	94.4
10	99.2	93.2	95.2
15	99.4	95.4	95.4

ROC-AUC, which is often known as AUC, is used for evaluating the effectiveness of the classifiers. ROC can be defined as graphical representation regarding the false-positive rate (FPR) versus the true positive rate (TPR), and the area covered by such curve shows the optimal efficiency. Eq14 is used to calculate the AUC. Figure 8 depicts the ROC-AUC curve for each classifier and variation. With Burg's approach, the SVM is found to deliver the best performance with 0.99% AUC.

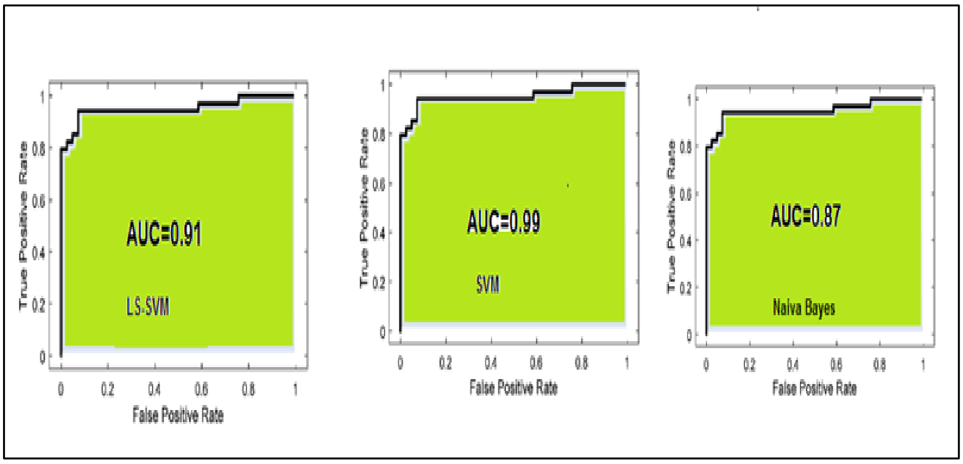

Fig. 8. (ROC-AUC) plot (AUC values) for different classifiers.

Various methods for separating alcoholics from nonalcoholics have been researched. Quantitative electroencephalography (QEEG) provided the author of [4] with two features: relative power and absolute power. The most crucial features that successfully distinguish between the two topics for classification were chosen using PCA and t-test. When 10-fold cross-validation was adopted, the logistic model tree outperformed all other classifiers with an accuracy rate of 92%.

Interhemispheric coherence and spectral powers have been the subject of an experiment. AUCs were determined for each feature, referred to as the feature selection strategy. The classifiers were then given a set of features to use in assessing the performance of the system. With a 90% accuracy rate, logistic regression (LR) classifier fared better than the other classifies [16]. The time–frequency domain, which uses multiple wavelet transforms, replaces the loss of temporal and spectral information, which is a drawback.

For classifying participants in an EEG study who were either drunk or sober, Shaymaa et al. [12] used ML ensemble approaches. Pachori et al. [17] employed a tuned-Qi wavelet transform with center corresponding entropy as a feature for signal decomposition. The feature vector's dimension was reduced by using PCA methods. The LS-SVM classifier—whose best performance was 97%—was trained and evaluated afterward. An index was also developed to keep track of the patients' medical care. For the evaluation of dual tree complex wavelet transform using L2 norms and log energy entropy features from each signal SB, LS-SVM and sequential minimal optimization support vector machine showed the best performance with 97.917% accuracy [9].

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 1, 2024, pp. 8-19

DOI: https://doi.org/10.25195/ijci.v50i1.438
Print ISSN: 2313-190X, Online ISSN: 2520-4912

Log-energy features were extracted using the 3-band orthogonal wavelet filter bank and then chosen based on the ranking technique and input to multiple supervised ML approaches. LS-SVM (RBF kernel) produced a notable result with 97.08% accuracy. Different entropy-based features such as negentropy, entropy, skewness, kurtosis, and mean were evaluated and assessed using LLS-SVM and various kernel tricks [9]. The RBF kernel showed the best result with a 97.20% accuracy. The signal was disintegrated using the empirical mode decomposition (EMD) wavelet transformation [9].

Empirical wavelets transform (EWT) represents an advanced variant of EMD that adaptively selects the decomposed signal mode based on the EEG signal's frequency spectrum, making it suitable for nonstationary signals. Works related to automating nonalcoholic and alcoholic subject/user detection with EEG utilizing various wavelet transformations are summarized in Table 5.

Table V. Summarizes Related Works for Alcoholic And Nonalcoholic Subjects Based On Eeg By Applied Different Wavelet Transforms.

# No	Reference	Features Set	Classifier	Performance of the classifier
1	Sharma et al [25]	Log energy	LS-SVM	97.08 %
2	Sharma et al.[26]	LEEs, L2Ns	SMO-SVM, LS-SVM, FSC	97.91 %
3	Priya et al.[29]	mean, kurtosis, skewness, entropy, and negentropy	LS-SVM (RBF) LS-SVM (polynomial)	97.92 % with RBF kernel
4	Pachori et al[27]	Centered Correntropy (CC)	LS-SVM	97.02 %
5	Shaymaa. et al [11]	Sample entropy	SVM	0.95%
6	**Proposed work**	**Fractal dimension features, Statistical features(SF)**	**SVM**	**0.9292**
			LS-SVM	**0.8721**
			Naiva Bayes	**0.8124**

7. CONCLUSION

Alcoholism is a disease that is diagnosed by evaluating the recorded EEG signal. If the medical expert is unable to recognize the signal fluctuations, this phenomenon may also result in significant disease in some situations. Hence, an automated diagnosis system that is economical and nonintrusive must be developed to be used with ease in all hospital centers and outlying villages. This study suggests a technique for automatically identifying participants who are alcoholics. The channels used for feature extraction are T5, TP8, PZ, P3, O1, T6, O2, and OZ. Such channels are allocated based on electrode position in the standard 10/20 scheme. Statistical features are deemed to be significant for further consideration when the extracted characteristics and features are examined using the p-test. These important properties are used to create feature vectors, which are then used to train different machine learning models. Leave-one-out cross-validation is used for training and testing to prevent overfitting. The effectiveness of the classification model is assessed using a number of metrics and is represented using ROC plots. According to the results, LLS-SVM with linear and RBF kernel has achieved 0.8721% accuracy, and SVM with polynomial kernel has performed the best with an accuracy of 0.9892. Meanwhile, naive Bayes' accuracy is 0.8123 The same pattern has been observed in the other measurements. The suggested computer-supported framework gives medical professionals the ability to spot alcoholism in a patient who is not showing symptoms of alcohol misuse. Developing a suitable EEG-based system to identify drivers who consume alcohol while operating a vehicle could help reduce traffic accidents. The strategy in this work can be improved in the future by combining various optimization algorithms and deep learning approaches to increase detection accuracy. In addition, the assertions made in this study on the subject of alcoholism can be validated in real time.

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 1, 2024, pp. 8-19

DOI: https://doi.org/10.25195/ijci.v50i1.438
Print ISSN: 2313-190X, Online ISSN: 2520-4912

References

[1] T. Rieg, J. Frick, M. Hitzler, and R. Buettner, “High-performance detection of alcoholism by unfolding the amalgamated EEG spectrausing the random forests method,” in Proc. 52nd Hawaii Int. Conf. Syst. Sci., 2019

[2] 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

[3] L. Bauer, Predicting relapse to alcohol and drug abuse viaquantitativeelectroencephalography, Neuropsychopharmacology 25 (September 3)(2001) 332–340.

[4] P. Gonzalo, S. Radenne, and S. Gonzalo, “Biomarkers of chronic alcohol misuse,” Current Biomarker Findings, vol. 4, pp. 9–22, Jan. 2014.

[5] E. Alpaydin, Introduction to Machine Learning (Adaptive Computation and Machine Learning Series). Cambridge, MA, USA: MIT Press, 2004

[6] R. Chatterjee, T. Maitra, S. H. Islam, M. M. Hassan, A. Alamri, and G. Fortino, “A novel machine learning based feature selection for motor imagery EEG signal classification in Internet of medical things environment,” Future Gener. Comput. Syst., vol. 98, pp. 419–434, Sep. 2019.

[7] Singhal, Vatsal etal “ Detection of alcoholism using EEG signals and a CNN-LSTM-ATTN network“Computers in biology and medicine , vol 138, 2021

[8] J.N. Acharya, A. Hani, J. Cheek, P. Thirumala, T.N. Tsuchida, American ClinicalNeurophysiology Society guideline 2: guidelines for standart electrodeposition nomenclature, J. Clin. Neurophysiol. 33 (4) (2016) 308–311.

[9] Salankar, etal “EEG based alcoholism detection by oscillatory modes decomposition second order difference plots and machine learning” Biocybernetics and Biomedical Engineering vol 42 , 2022

[10] Abdulrahman, Shaymaa Adnan and Alhayani, Bilal ," A comprehensive survey on the biometric systems based on physiological and behavioural characteristics" in 2020 Conference on Materials Today, 2020

[11] Li, Y., Wen, P.P., 2011. Clustering technique-based least square support vector machine for EEG signal classification. Compute. Methods Programs Biomed. 104,358–372

[12] Al-Salman, W., Li, Y., Wen, P., 2019a. Detecting sleep spindles in EEGs using wavelet Fourier analysis and statistical features. Biomed. Signal Process. Control 48,80–92.Al-Salman, W., Li, Y., Wen, P., 2019b.

[13] Mehmood, R.M., Lee, H.J., 2015. Exploration of prominent frequency wave in EEGsignals from brain sensors network. Int. J. Distrib. Sens. Netw. 11, 386057

[14] Ali, Z., Elamvazuthi, I., Alsulaiman, M., Muhammad, G., 2019. Detection of voicepathology using fractal dimension in a multiresolution analysis of normal anddisordered speech signals. J. Med. Syst. 40, 20.

[15] 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 https://doi.org/10.1016/j.cosrev.2020.100237

[16] J. Ye, T. Xiong, SVM versus least squares SVM, J. Mach. Learn. Res. Proc. Track 644–651

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