

---

# Wavelet analysis of EEG for the identification of alcoholics using probabilistic classifiers and neural networks

---

E. Malar\*

Department of ECE,  
PSG Institute of Technology and Applied Research, India  
Email: emjournal@gmail.com  
\*Corresponding author

M. Gauthaam

CrowdANALYTIX Solutions Private Limited,  
2621, 2nd Floor, 27th Main Rd,  
Sector 1, HSR Layout,  
Bangalore 560102, India  
Email: gauthaam@crowdanalytix.com

**Abstract:** Electroencephalography (EEG) is the process of recording the complex activity of the brain in the form of signals. EEG primarily has delta, theta, alpha, beta and gamma frequency bands whose presence and strength describes changes in brain under different kinds of activities. On the other hand alcohol consumption leads to depression and confusion which reduces the activity of the nervous system thereby affecting the brain. Alcoholics are identified from normal persons by multi-resolution and multi-scale analysis of EEG. In our research, EEG is decomposed into sub frequency bands using wavelet. The effect of alcohol on each of these wave bands is identified using power spectral density analysis. These evident variations in EEG are manifested due to depression in brain activity caused by intake of alcohol. The first order and second order statistical measures of the EEG signal are selected as features. Classifiers such as Bayes, Naive Bayes, radial basis function network (RBFN), multilayer perceptron (MLP) and extreme learning machine (ELM) are used for classification. Results show that our proposed EEG analysis acts as an effective bio-marker for differentiating alcoholics from non-alcoholics and extreme learning machine provides higher classification efficiency (87.6%) compared to other classifiers used.

**Keywords:** electroencephalography; EEG; frequency bands; power spectral density; wavelet decomposition; extreme learning machine; ELM.

**Reference** to this paper should be made as follows: Malar, E. and Gauthaam, M. (2020) 'Wavelet analysis of EEG for the identification of alcoholics using probabilistic classifiers and neural networks', *Int. J. Intelligence and Sustainable Computing*, Vol. 1, No. 1, pp.3–18.

**Biographical notes:** E. Malar received her BE in Electrical and Electronics Engineering from the Bharathiar University, India in 1998 and ME degree from the Anna University, Chennai, India in 2008 and PhD degree from the Anna University, Chennai, India in 2014. She is working as a Professor and Head, Department of ECE, PSG Institute of Technology and Applied Research,

Coimbatore. India. Her main research areas are biomedical instrumentation and medical image processing. She is a life member of Indian Society of Technical Education and Biomedical Engineer Society of India.

M. Gauthaam received his BE in Biomedical Engineering from the PSG College of Technology, India in 2012. He is currently working as a Senior Data Scientist (R&D – product development) at CrowdANALYTIX. His main research interests include innovations in deep learning algorithms for converting unstructured data into structured information.

---

## 1 Introduction

Brain acts as a central control and data processing unit for the biological medium. It produces our every thought, action, memory, feeling and experience of the world. Brain consists of staggering one hundred billion nerve cells or neurons. The neuronal elements (dendrites, somata and axons) of each of these billions of neurons generate a small amount of electrical charge as a result of excitation of the brain for an activity. The process of mapping this electrical activity of the brain onto a graph is called electroencephalography (EEG). It records the potential fluctuations generated due to neuronal current generators in the brain. The field potentials recorded are the aggregates of neuronal elements with complex interconnections. The intensity and the patterns of the recorded potentials depend on the overall excitation of the brain.

EEG consists of four major brain waves at different frequency ranges which are prominent under diverse activities. They are alpha (8–12 Hz), beta (12–32 Hz), theta (4–8 Hz) delta (0–4 Hz) which is intense at occipital, parietal, frontal and cortex regions respectively. The alpha waves are prominent in normal awake persons in quiet and pleasant environment. Beta waves occur during intense mental activity. Theta waves occur prominently in children and during emotional stress in adults. Delta waves occur in deep sleep.

EEG as an effective diagnostic tool finds wide applications in studying the response of the brain, assessing alertness and detecting disease like epilepsy and seizures. Alcohol, the most widely consumed drug worldwide greatly decreases the brain activity. Also alcohol dependants face numerous health consequences due to increase in the level of norepinephrine, the neurotransmitter responsible for heightened excitement. The regions of the brain that suffers maximum decrease in activity are the prefrontal cortex and temporal cortex. These are the regions responsible for decision making and forming new memories. EEG is best suited to visualise this decrease in brain activity (Begleiter and Platz, 1972; Hommer et al., 2001).

Many researchers have conducted studies to determine the effect of alcohol on human brain activity using EEG. Wu Di et al. (2010) analysed EEG using power spectral density. He found out that with the increase in the amount of alcohol intake, the power of the EEG signal decreases in frontal region and increases in central, occipital region. Yazdani and Kamaledin (2007) further went on to classify EEG of alcoholics from non-alcoholics. He used different statistical classifiers such as Bayes classifier, k-nearest neighbour classifier and minimum mean distance classifier. While Nazari Kousarrizi et al. (2009) classified the EEG of alcoholics and non-alcoholics using support vector machine and multilayer perceptron network.

In this paper, we propose a novel technique to classify alcoholics and non-alcoholics using EEG. We perform wavelet decomposition of EEG to separate information section from noise and extract different wave bands. Power spectral analysis of each wave bands of EEG is done to validate the effect induced by alcohol on human brain. Then the first order and second order statistical features are extracted from the wave bands. These features are used to train and test the classifiers.

We use set of supervised classifiers namely Bayes network, Naive Bayes network, RBFN, MLP and ELM to test the effectiveness of classification. In this study we find extracted wavelet features that are classified using ELM produces relatively better classification accuracy with a significant reduction in training time. It also avoids problems like local minima, improper learning rate and over fitting which the other networks experience.

The forthcoming part of the article is organised as follows: Section 2 presents the materials and methods of the proposed system. Section 3 discusses the details of classification of EEG. Section 4 explains the performance of classifiers required to be evaluated in order to compare their sensitivity, specificity along with overall accuracy. Section 5 shows performance evaluation and results and Section 6 provides the conclusion.

## 2 Materials and methods

### 2.1 Data collection

The EEG database is acquired from standard EEG database of State University of New York health centre. This database consists of measurements of 64 electrodes placed on various location on the scalp of the subject and is sampled at 256 Hz. There were two groups of subjects: alcoholic and control. The alcoholic subjects refer to persons who have genetic predisposition to alcoholism. Most of the alcoholics had been drinking heavily for a minimum of 15 years and started drinking at approximately 20 years of age. The alcoholics were non-amnesic and also not substance abusers. They were also matched for socioeconomic status. The control subjects had no personal and no family history of alcohol and/or drug abuse or any history of neurological or psychiatric disease. Totally 2,560 datasets are used; out of which 50% of them are alcoholic data while other 50% being non-alcoholic data (Malar et al., 2011). This data is decomposed using wavelet decomposition and first and second order statistical features are extracted for further processing. The distribution of the data collected is presented in Tables 1 and 2.

**Table 1** Distribution of dataset

<i>Type of data</i>	<i>Training dataset</i>	<i>Testing dataset</i>
Alcoholic data	800	480
Non-alcoholic data	800	480
Total	1,600	960

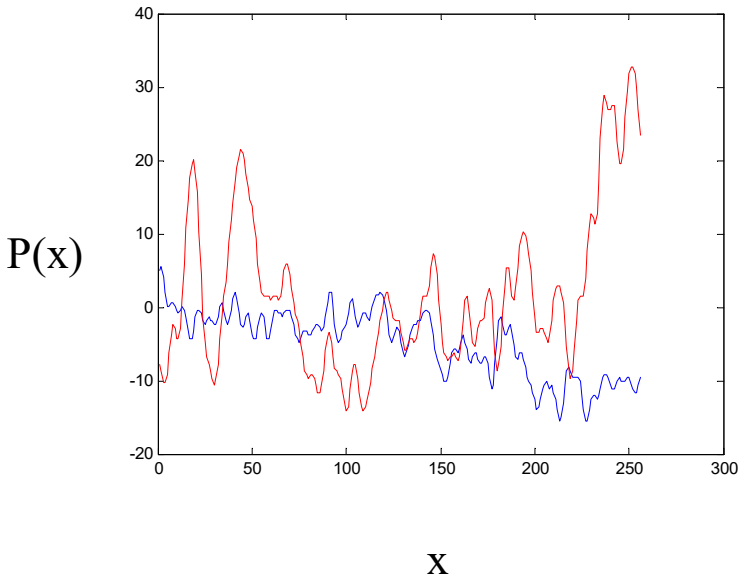
**Table 2** List of features extracted

<i>S. no.</i>	<i>Feature</i>	<i>Formula</i>
1	Mean absolute value	$\frac{1}{2} \sum_{i=1}^N  x_i $
2	Variance	$\sigma^2 = E\{x^2\} - E^2\{x\}$
3	Zero crossing	$X_i > 0$ and $X_{i+1} < 0$ or $X_i < 0$ and $X_{i+1} > 0$
4	Slope sign change	$X_i > X_{i-1}$ and $X_{i+1} < X_i$ or $X_i < X_{i-1}$ and $X_{i+1} > X_i$
5	Entropy	$-\sum(h * \log_2(h))$ , where $h$ is histogram count
6	Energy	$\sum[x(i)]^2$
7	Norm	maximum(sum(absolute(x)))

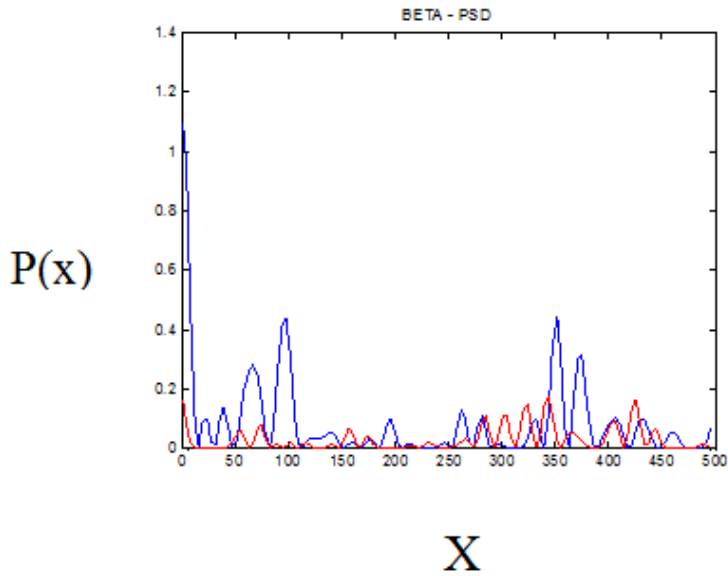
**2.2 Power spectral analysis of EEG**

In order to understand the extent of effect brought about by the intake of alcohol, the power spectral density of the EEG pertaining to alcoholic person is analysed. Power spectral density is the analysis of power of the signal that gives measure of strength of the signal. This analysis is extended to individual wave bands of EEG to illustrate the effect on it. Figures 1–4 give a comparison of power spectral density of EEG of alcoholic and non- alcoholic person.

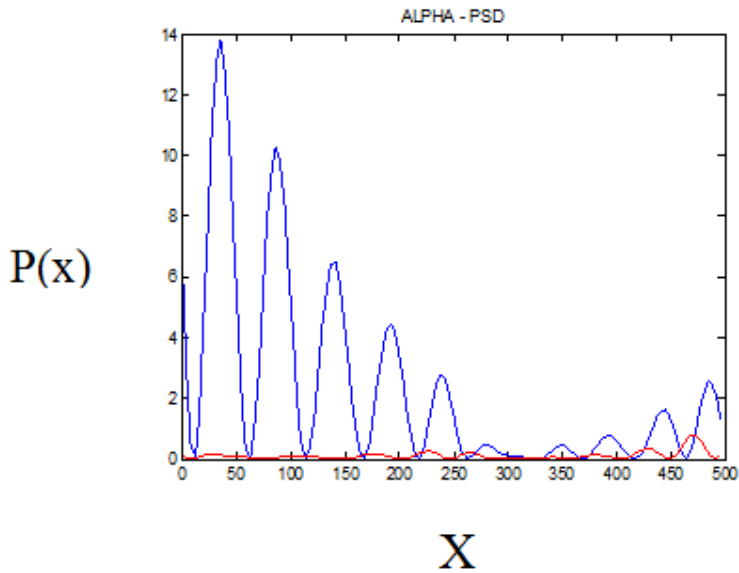
**Figure 1** EEG of normal person (blue) vs. alcoholic person (red) (see online version for colours)



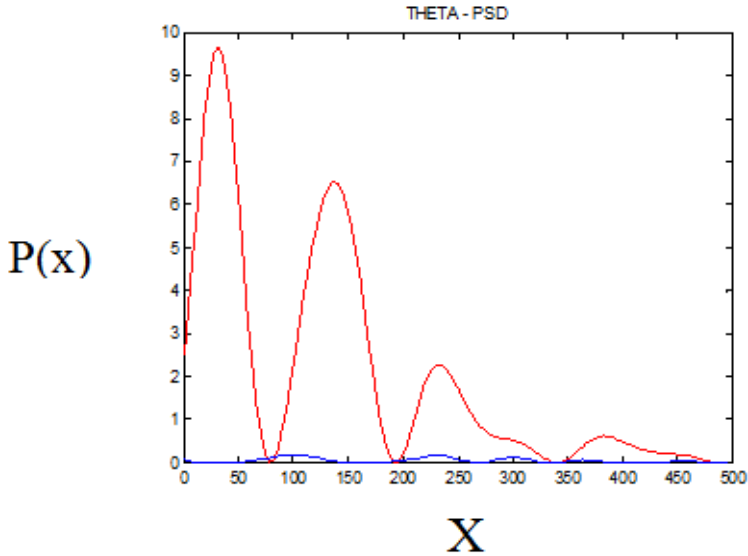
**Figure 2** Power spectrum density of beta wave of normal person (blue) and alcoholic person (red) (see online version for colours)



**Figure 3** Power spectrum density of alpha wave of normal person (blue) and alcoholic person (red) (see online version for colours)



**Figure 4** Power spectrum density of theta wave of normal person (blue) and alcoholic person (red) (see online version for colours)



It is evident from the above study that intake of alcohol has an appreciable effect on EEG. It can be noticed that after consuming alcohol: theta ( $\theta$ ) waves begin to appear and gradually enhances (as subjects enter in to a state of sleepy and also central nervous system is inhibited), alpha ( $\alpha$ ) waves gradually decreases and the region of alpha ( $\alpha$ ) waves is expanded, beta ( $\beta$ ) waves gradually enhanced and the area of beta ( $\beta$ ) waves are expanded (as cerebral cortex remains in an excitable condition) (Zhang et al., 1997).

As alcohol is susceptible to affect brain activity and subsequent brain waves, distinct changes can be mapped in EEG. These changes can be utilised to differentiate alcoholics from non-alcoholics (Ocak, 2008).

### 2.3 Wavelet decomposition

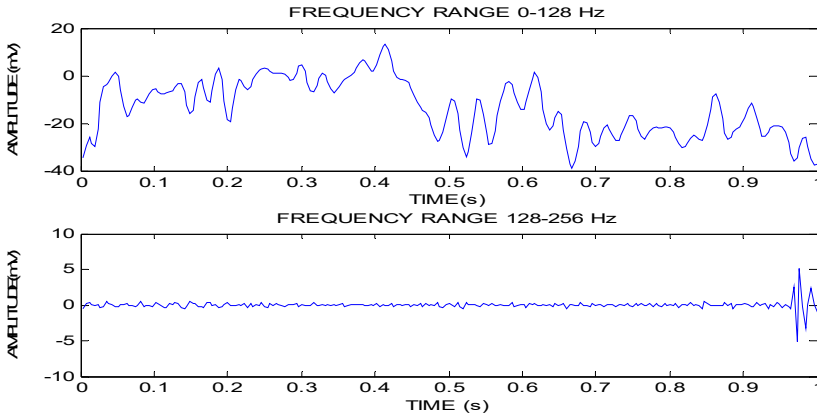
For non-stationary waves like EEG, the wavelet decomposition is preferred as it gives precise frequency information at lower frequency and precise time information at higher frequency. The EEG data taken into consideration is sampled at 256 Hz. The useful frequency range is (0–60) Hz and the higher frequency range is considered to be noise. Therefore, it is required to eliminate higher frequencies.

Two levels wavelet decomposition with Daubechies mother wavelet of order 10 is employed. After the first level of decomposition, the original EEG data with frequency range (0–256) Hz is decomposed into d1, higher resolution components (128–256) Hz and a1, lower resolution components (0–128) Hz. In the second level decomposition, a1 is decomposed to a2, lower resolution component (0–64) Hz and d2, higher resolution component (64–128) Hz. This a2 component of EEG data is the required useful part of the EEG which comprises of the wave bands – alpha, beta, theta, delta and gamma.

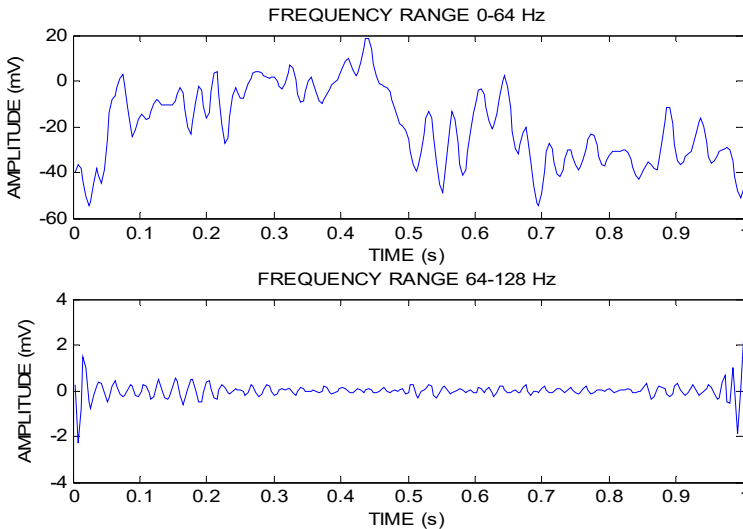
To obtain the individual wave bands, a2 component is further decomposed into a3, lower resolution component (0–32) Hz and d3, higher resolution component (32–64 Hz).

The d3 component corresponds to gamma band of the EEG. The a3 component is decomposed to a4 and d4, further a4 into a5 and d5 and a5 into a6 and d6. The d4, higher resolution component (16–32 Hz) of a3 corresponds to beta band of the EEG, the d5, higher resolution component (8–16 Hz) of a4 corresponds to alpha band of the EEG, the d6, higher resolution component (4–8 Hz) of a5 corresponds to theta band of the EEG while the a6, lower resolution component (0–4 Hz) of a5 corresponds to delta band of the EEG (Huang et al., 2006a). The first and second level of decomposition is shown in Figures 5 and 6 respectively.

**Figure 5** First level of wavelet decomposition of the signal (see online version for colours)

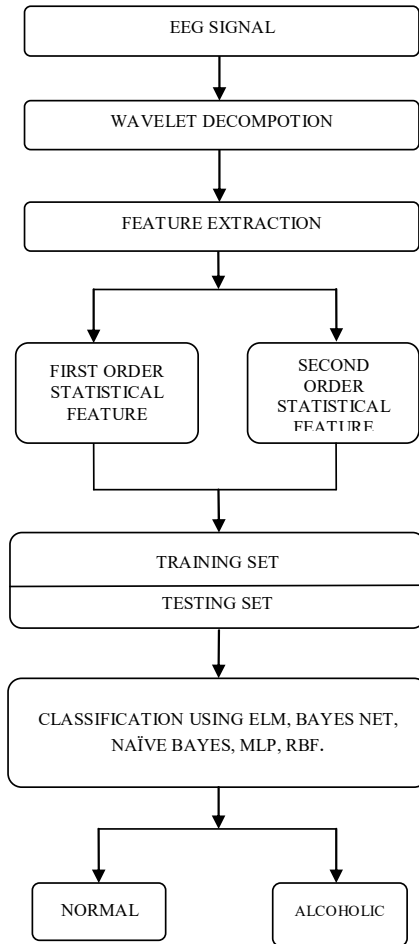


**Figure 6** Second level of wavelet decomposition of the signal (see online version for colours)

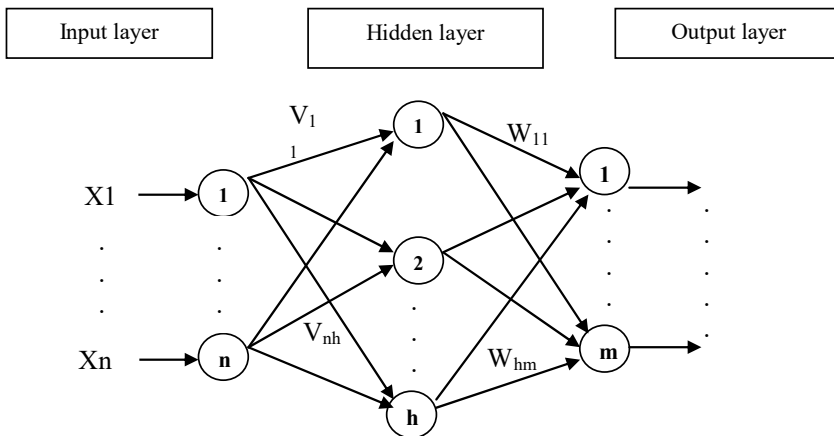


The block diagram of the proposed methodology to separate alcoholics from non-alcoholics is shown in Figure 7.

**Figure 7** Block diagram of the proposed methodology



**Figure 8** Structure of ELM network





### 3 EEG classifications

An artificial neural network is an emulation of biological neural system. It is a parallel system, capable of resolving complex problems. The greatest advantage over other hard computing classifiers is that it is tolerant to missing data, mistakes in the data and allows approximation of the data. Adding to that, generally it has higher efficiency compared to others.

The performance of our automatic alcoholic and non alcoholic classification system in this study is evaluated by the probabilistic classifiers: Bayes classifier, Naïve Bayes classifier and neural networks: ELM, RBFN and MLP.

#### 3.1 Bayes and Naïve Bayes classifiers

Bayesian decision theory is a fundamental statistical approach to the problem of pattern classification. It makes the assumption that the decision problem is posed in probabilistic terms, and that all of the relevant probability values are known. It encodes dependencies among all variables and readily handles situations where some data entries are missing. Naïve bayes classifier is also a probabilistic classifier based on Bayes theorem. This classifier assumes that the effect of value of each attribute on a given class is independent of value of other attributes. This is known as conditional independency. An advantage of the Naive Bayes classifier is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

#### 3.2 Extreme learning machine

The ability of the neural network to produce greater classification efficiency is the most celebrated aspect of using neural network. The time taken for computation is as important aspect for a classifier as efficiency is. Number of algorithms has been used to develop neural networks and a major and a common disadvantage these algorithms pose is their learning speed. The slow learning rates may be attributed to their gradient descent or iterative approach in training the network.

Recently, Huang et al. (2006b) proposed a new learning algorithm named extreme learning machine which reduces the computational time substantially yet not compromising on efficiency. This algorithm as proposed by Huang et al. uses single layer feed forward network architecture (SLFN) (Malar et al., 2012; Suresh et al., 2010).

#### 3.3 Radial basis function neural network

Radial basis function network (RBFN) is a type of artificial neural network. It has a feed forward architecture with an input layer, a hidden layer, and an output layer. It is applied to the problems of supervised learning and associated with radial basis functions.

### 3.4 Multilayer perceptron

The design of the MLP is completely characterised by an input layer, one or more hidden layers, and an output layer. All layers consist of at least one neuron. Analytically, the MLP performs a nonlinear approximation using sigmoidal kernel functions as hidden units and linear weights. During the training of feature vectors, the weights are continuously tailored by diminishing the error between desired outputs and the computed outputs. MLPs are highly nonlinear interconnected systems and serve for both nonlinear function approximation and nonlinear classification tasks.

## 4 Performance analysis

In general the performance of classifiers is required to be evaluated in order to compare their sensitivity, specificity along with overall accuracy. For this analysis, usually the following confusion matrix in Table 3 is designed based on the trade-off between actual and classifier generated outputs. Also the performance of the classifiers is tested as per the specifications in Table 4.

**Table 3** Confusion matrix

<i>Classifier output</i>	<i>Non-alcoholic</i>	<i>Alcoholic</i>
Positive	TP	FP
Negative	FN	TN

**Table 4** ELM based alcoholic and non alcoholic classification system specification

<i>Classifier</i>	<i>Number of hidden layer</i>	<i>Lambda (<math>\lambda</math>)</i>	<i>Activation function</i>
ELM	100	1	Gaussian

Sensitivity: it is a measure of accuracy of identification of alcoholics. It is mathematically defined as

$$\text{Sensitivity} = \frac{TP}{TP + TN} \quad (1)$$

Specificity: it is a measure of accuracy of identification of non-alcoholics. It is mathematically defined as

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (2)$$

Precision: it is a measure of repeatability of true classification (Non-alcoholic). It is mathematically defined as

$$\text{Precision } (P_r) = \frac{TP}{TP + FP} \quad (3)$$

Accuracy: it is a measure of exactness of true classification (Non-alcoholic and alcoholic). It is mathematically defined as

$$\text{Accuracy } (A_c) = \frac{TP + TN}{P + N} \quad (4)$$

where

- True positive (*TP*)    Person predicted alcoholic when really consumed  
 True negative (*TN*)    Person predicted alcoholic when not consumed  
 False positive (*FP*)    Person predicted non-alcoholic when really consumed  
 False negative (*FN*)    Person predicted non-alcoholic when not consumed

## 5 Results and discussions

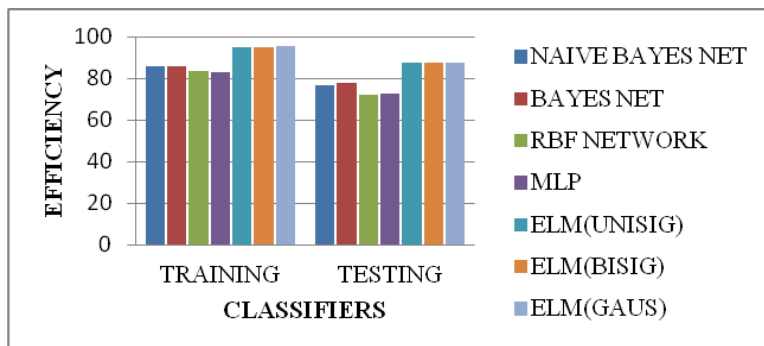
We have evaluated the performance of the our proposed alcoholic and non-alcoholic classification system using 1,600 datasets for training the network in which 800 samples are of alcoholic while remaining 800 belong to non-alcoholic. Then 960 datasets are used for testing the network which comprised of 480 alcoholic and 480 non-alcoholics. Here we have used the classifiers: Bayes net, Naive Bayes, RBF network, multilayer perceptron and extreme learning machine. Performances of the classifiers used are given in Table 5.

**Table 5** Performance analysis of classifiers

<i>S. no.</i>	<i>Classifier</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Precision</i>	<i>Accuracy</i>	<i>Training efficiency</i>	<i>Testing efficiency</i>
1	Naive Bayes net	75.81%	77.42%	77.70%	76.46%	85.93%	76.46%
2	Bayes net	77%	79.34%	80.20%	78.12%	86%	78.12%
3	RBF network	72.49%	71.48%	70.83%	71.9%	83.31%	71.9%
4	Multilayer perceptron	72.18%	73.03%	73.54%	72.60%	82.93%	72.60%
5	Extreme learning machine	87.37%	87.84%	87.91%	87.6%	95.3%	87.6%

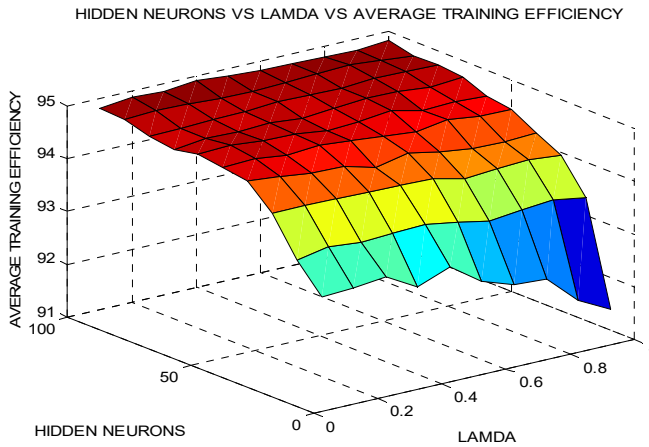
The graph shown in Figure 9 compares the training and testing efficiency of all the classifiers used.

**Figure 9** Performance comparisons of the classifiers (see online version for colours)

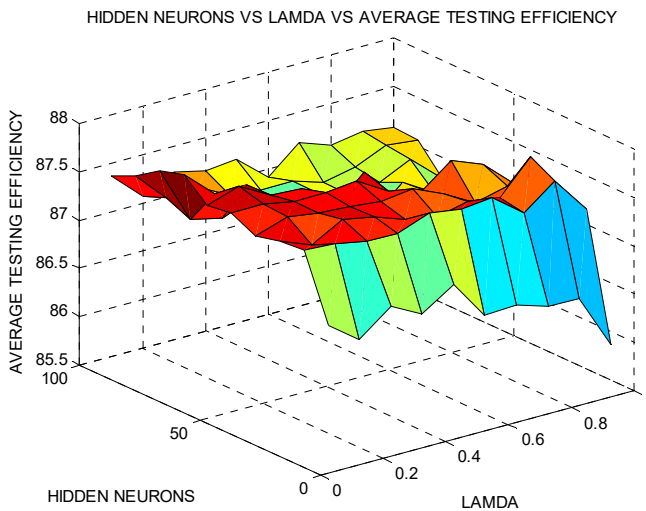


From the above results, we conclude that the ELM classifier outperforms others. Further, the performance of ELM is assessed using three different activation functions – unipolar sigmoidal, bipolar sigmoidal and Gaussian. Training and testing efficiency of the network for various activation functions is evaluated for different numbers of hidden neurons and different values of lambda, a parameter in the activation function that determine the threshold. Number of hidden neurons is varied from 0 to 100 while the value of lambda was incremented from 0.1 to 1. Each combination is evaluated for 20 trials and average efficiency was calculated. The performances of ELM for all the possible combinations are depicted in the 3D plots shown in Figures 10–15.

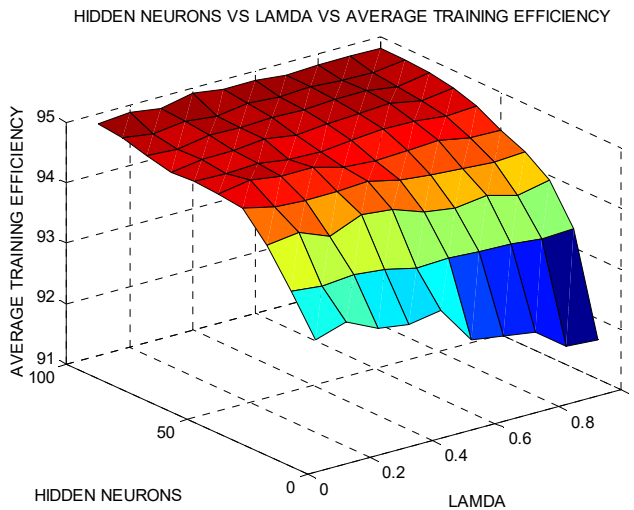
**Figure 10** ELM training efficiency under unipolar sigmoidal activation function (see online version for colours)



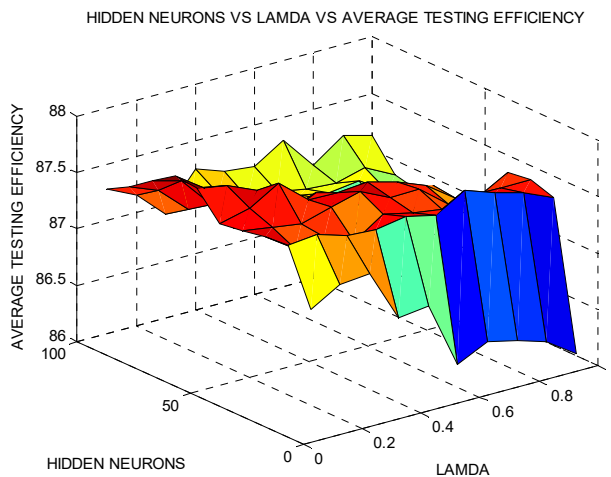
**Figure 11** ELM testing efficiency under unipolar sigmoidal activation function (see online version for colours)



**Figure 12** ELM training efficiency under bipolar sigmoidal activation function (see online version for colours)

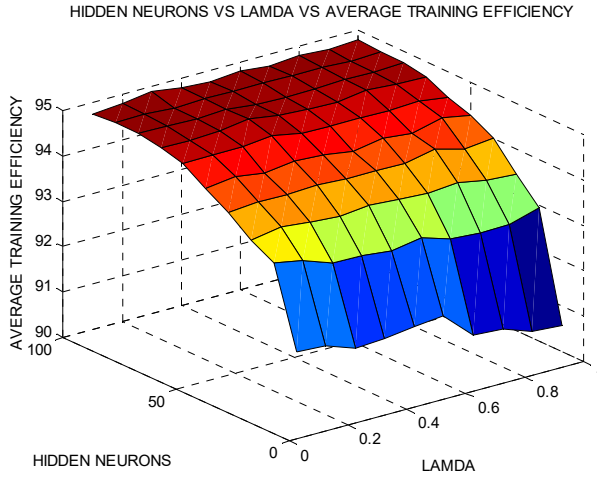


**Figure 13** ELM testing efficiency under bipolar sigmoidal activation function (see online version for colours)

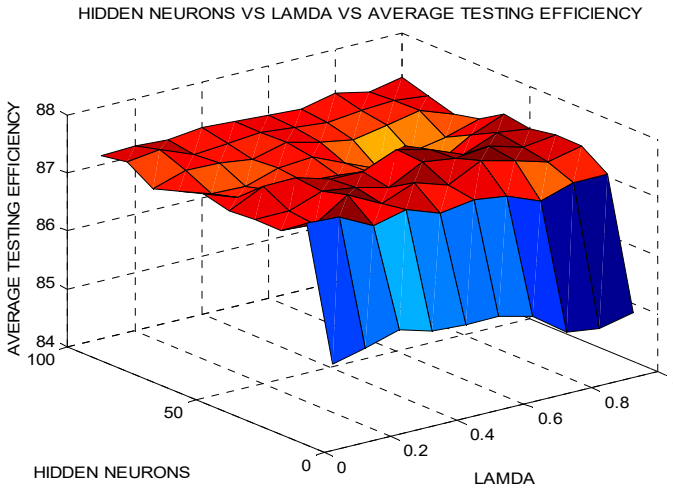


ELM with Gaussian activation function has consistency over the classification efficiency for different combinations of hidden neurons and the value of lambda compared to other two activation functions. Table 5 compares the maximum training and testing efficiency of the three activation functions. In Table 6, a comparison of existing methods in identification of alcoholics is provided.

**Figure 14** ELM training efficiency under Gaussian activation function (see online version for colours)



**Figure 15** ELM testing efficiency under Gaussian activation function (see online version for colours)



**Table 6** Performance of ELM under unipolar sigmoidal, bipolar sigmoidal and Gaussian activation function

<i>S. no.</i>	<i>Activation function</i>	<i>Maximum training efficiency</i>	<i>Maximum testing efficiency</i>
1	Unipolar sigmoidal	95.2%	87.3%
2	Bipolar sigmoidal	95.2%	87.5%
3	Gaussian	95.3%	87.6%

**Table 7** Comparison of existing methods in identification of alcoholics

<i>Authors</i>	<i>Classifier</i>	<i>No. of samples</i>	<i>No. of features</i>	<i>Accuracy</i>
Palaniappan (2007)	MLP	800	123	91.25%
Damousis et al. (2011)	SVM	80	19	85%
Proposed method	ELM	2,560	7	87.6%

Palaniappan (2007) used 800 samples of the same database to classify alcoholic and non-alcoholic. Nearly 123 features of EEG are employed in a multilayer perceptron neural network. Though classification accuracy of 91.25% is achieved, use of large number of features increases the computational time of the network. Damousis et al. (2011) used different data set to categorise alcoholics and obtained an accuracy of 85%. Almost 80 samples and 19 features of EEG data are used for classification. Here the number of samples used is very much less when compared to the proposed method. And also number of features used is large.

## 6 Conclusions

Electroencephalograph can be used as a tool to differentiate alcoholics from non-alcoholics. The ground truth that EEG is influenced by alcohol intake is validated through power spectral density analysis. Wavelet transforms were used to extract the useful section of EEG, the segment which includes tangible information from which statistical features were extracted. Classification was performed using classifiers such as Bayes, Naive Bayes, RBFN, MLP and ELM and is found out that ELM provides with higher rates of efficiency compared to other classifiers. Therefore the proposed system discriminates alcoholics from non-alcoholics with higher sensitivity, specificity, precision and accuracy. The proposed method can be used for screening chronic alcoholic subjects. According to the studies of Dr. Henry Begleiter, some neurophysiologic anomalies in chronic alcoholics were already present in their young offspring before any exposure to alcohol and drugs. Presence of this neural excitability is a predisposing factor leading to the development of alcoholism, substance abuse, conduct disorder and antisocial personality disorder (Begleiter and Platz, 1972). Therefore, automatic detection of chronic alcoholic will be welfare to the society and for certain job recruitments that require more concentration.

## References

- Begleiter, H. and Platz, A. (1972) 'The effects of alcohol on the central nervous system in humans', in Kissin, G. and Begleiter, H. (Eds.): *The Biology of Alcoholism*, pp.293–343, Plenum Press, Boston, MA.
- Damousis, G., Argyropoulos, S. and Muzet, A. (2011) 'Classification of physiology indicators for the automatic detection of potentially hazardous physiological states', *Applied Computational Intelligence and Soft Computing*, Vol. 2011, No. 4, pp.1–8.
- Friedman, N., Geiger, D. and Goldszmidt, M. (1997) 'Bayesian network classifiers', *Machine Learning*, Vol. 29, Nos. 2–3, pp.131–163.
- Hommer, D.W., Momenan, R., Kaiser, E. and Rawlings, R.R. (2001) 'Evidence for a gender related effect of alcoholism on brain volume', *Am. J. Psych.*, Vol. 158, No. 2, pp.198–204.

- Huang, G.B., Zhu, Q.Y. and Siew, C-K. (2006a) 'Extreme learning machine: theory and applications', *Neurocomputing*, Vol. 70, Nos. 1–3, pp.489–501.
- Huang, G.B., Zhu, Q.Y. and Siew, C-K. (2006b) 'Universal approximation using incremental constructive feed forward networks with random hidden nodes', *IEEE Trans. Neural Netw.*, Vol. 17, No. 4, pp.879–892.
- Kousarrizi, M.R.N., Ghanbari, A.A., Gharaviri, A., Teshnehlab, M. and Aliyari, M. (2009) 'Classification of alcoholics and non-alcoholics via EEG using SVM and neural networks', *Proc. 3rd Int. Conf. Bioinformatics and Biomedical Engineering (ICBBE)*, Tehran, Iran, June, pp.1–4.
- Malar, E., Gauthaam, M. and Chakravarthy, D. (2011) 'A novel approach for the detection of drunken driving using the power spectral density analysis of EEG', *IJCA*, Vol. 21, No. 7, pp.10–14.
- Malar, E., Kandaswamy, A., Chakravarthy, D. and Dharan, A.G. (2012) 'A novel approach for detection and classification of mammographic microcalcifications using wavelet analysis and extreme learning machine', *Computers in Biology and Medicine*, Vol. 42, No. 9, pp.898–905.
- Ocak, H. (2008) 'Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm', *Signal Process*, Vol. 88, No. 7, pp.1858–1867.
- Palaniappan, R. (2007) 'Screening for chronic alcoholic subjects using multiple gamma band EEG: a pilot study', *JCS&T*, Vol. 7, No. 2, pp.182–185.
- Suresh, S., Saraswathi, S. and Sundararajan, N. (2010) 'Performance enhancement of extreme learning machine for multi-category sparse data classification problems', *Eng. Appl. Artif. Intel.*, Vol. 23, No. 7, pp.1149–1157.
- Wu, D., Chen, Z., Feng, R., Li, G. and Luan, T. (2010) 'Study on human brain after consuming alcohol based on EEG signal', *Proc. 3rd IEEE Int. Conf. Computer Science and Information Technology (ICCSIT)*, July, Vol. 5, pp.406–409, Dalian, China.
- Yazdani, A. and Kamaledin, S.S. (2007) 'Classification of EEG signals correlated with alcohol abusers', *Int. Sym. Signal Processing and its Applications (ISSPA)*, Tehran, Iran, February, pp.1–4.
- Zhang, X.L., Begleiter, H., Porjesz, B. and Litke, A. (1997) 'Electrophysiological evidence of memory impairment in alcoholic patients' *biological psychiatry*, Vol. 42, No. 12, pp.1157–1171.

## Website

<http://kdd.ics.uci.edu/databases> (accessed May 2011).