Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 143 (2018) 242–249

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Conference on Advances in Computing and Communication (ICACC-2018)

Classification of EEG Data using *k*-Nearest Neighbor approach for Concealed Information Test

Annushree Bablani[a], Damodar Reddy Edla[a,*], Shubham Dodia[a]

[a] *National Institute of Technology Goa , Ponda (403401), India*

Abstract

In this paper, an EEG based Concealed Information Test is developed. EEG is an acquisition technique of brain signal from brain scalp using electrodes. The main task here is to classify the EEG data into innocent and guilty. Data acquisition of 10 subjects has been carried out. The signal preprocessing is performed by passing the raw EEG signals through a band-pass filter. From these preprocessed EEG signals, it is necessary to extract significant features. In the time domain, the extraction of the statistical parameters such as mobility, activity and complexity is done from the EEG signals. The binary classification of the guilty and innocent classes in performed using *k*-nearest neighbor classifier. In order to validate the deceit identification system, 5-fold cross validation has been applied on the each of the subjects. To validate the performance of the classifier, performance measures such as accuracy, sensitivity, and specificity are taken into consideration. Out of three Hjorth parameters, mobility yielded better classification accuracy of up to 96.7%.

Keywords: Electroencephalography; *k*-nearest neighbor classifier; Hjorths' Parameter; P300; Event Related Potential.

1. Introduction

Brain Computer Interface (BCI) is an advanced technology which helps us to use the computational power of brain. Until few recent years, developing BCI was compared to a science fiction. The discovery of electroencephalography (EEG) changed this mindset, and researchers started working hard on to decode these EEG signals captured from the brain [1]. BCI is mainly divided into four phases namely Signal Acquisition, Signal Processing, Feature Extraction followed by Classification. During signal acquisition the recording of EEG signals is performed from the surface (or scalp) of the brain. These signals are recorded using electrodes, amplified using amplifier to increase power of signals and then digitized. In signal processing, noise reduction and artifact removal is performed. Signal processing stage thus helps in converting the extracted signals in required form for processing the next step. Feature extraction

* Damodar Reddy Elda. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: dr.reddy@nitgoa.ac.in

10.1016/j.procs.2018.10.392

phase extracts feature that would hopefully decode subjects message or command. BCI uses various features based on time domain or frequency domain or even a combination of them to increase the accuracy and performance. The task of extracting useful and relevant information out of these features is very challenging task. At times to reduce the complexity, the feature vector is reduced into lower dimension, but it always has a chance of losing the relevant information. With the use of feature vectors, classification of the signals is carried out. Therefore, to choose the features that helps to discriminate well is an essential task. Finally, these classified signals are translated into meaningful messages or commands for any application on a computer or a connected device such as a wheelchair or prosthetic devices. BCI has wide range of applications in the field of gaming, robotics, forensics, medicine etc. The main concern of our work is to use BCI in lie detection. Lie detection is an evaluation of a stated statement with the aim to disclose hidden intentions to lie. Traditional lie detection mechanism includes the most famous polygraph test. Here, a series of questions are asked by the interrogator and the person has to answer those questions in terms of yes or no. While the person answers these questions, few physiological indices are monitored such as blood pressure, heart rate, skin conductivity, and respiration. If a person is lying, it is observed that these indices change suddenly. But, the results of this technique are not acceptable as a strong evidence in judiciary system. A smart person can control his physiological indices and a innocent person may found guilty due to fear. EEG signals help us to solve this problem, as they are involuntary, and cannot be controlled by a user. Figure 1 illustrates the basic BCI architecture followed in the paper with all four phases and a feedback given back to user.

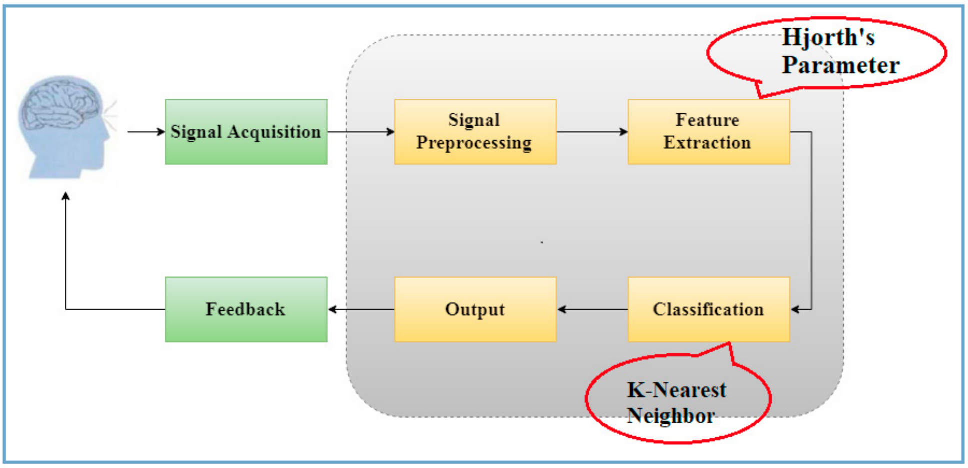

Fig. 1. Architecture of Brain Computer Interface.

The P300 (P3) wave is a deeply studied event related potential. This can be observed in response to unfamiliar, significant stimuli often called “oddball” paradigm. P3 is a wave which has positive deflection with a typical latency of 300-1000ms from stimulus onset. It can be observed that it is maximum at parietal lobe (at Pz) and minimum at frontal lobe (at Fz), and it takes intermediate values at central lobe (at Cz). The P300-based “Concealed Information Test (CIT)” or “Guilty Knowledge Test (GKT)” utilizes P300 amplitude which relies on the hypothesis of evoking different responses when presented the known items in the series of a large number of similar unknown items. The provided proof is that if a person is shown with a familiar object, P300 is generated. We now show an innocent person crime related object and P300 is not generated. Same is done in case of a guilty person, he/she says no, but still P300 is generated. This whole concept is used in the experiment to detect lie.

Many researchers have performed various tests and applied different statistical and machine learning approaches to classify EEG data into guilty and innocent. Farwell et. al in [2] have performed a “Guilty Knowledge Test” with the use of Event Related Brain Potentials (ERPs). GKT is used to test whether a person is guilty or innocent. Here, three types of stimuli are used, namely, probe stimuli, target stimuli, and irrelevant stimuli. A P300 response is elicitated when a person comes across a familiar object like target stimuli. A person can be convicted for a crime, if one has ”guilty information” and tells lie, but a P300 response is elicitated. This indicates that the person knows some special information about the object displayed. Using Bootstrapped analysis, the performance of the system was analysed and there was no false positives or no false negatives achieved in the experiment but 12.5% of the cases provided intermediate results. Haider et. al in [3] have proposed a lie detection method using LDA to separate the positive and negative samples. They have used sixteen channels and extracted signals using different extraction methods. This work was done using MATLAB and Xilinx tool. They also implemented the complete system on FPGA to check the efficiency. They achieved an accuracy of 85% using their proposed method. This method was claimed to be easy and much more convenient than the previously other proposed methods. Simbolon et. al in [4] have proposed a method based on SVM. To differentiate if a person is guilty or innocent, Event Related Potential (ERP) was used. The whole work was implemented in MATLAB. Eleven males were chosen for the experiment with the age varying in range of 20 and 27. The data was divided in training and testing and then various models were build. An accuracy of upto 70.83% was obtained using this approach. In this work, Hjorth parameters are used for feature extraction and k-nearest neighbor as a classifier. The paper consists of section 2 methodology, section 3 Results, and section 4 conclusion followed by references.

2. Methodology

2.1. Hjorth's Parameters

Hjorth in 1970 developed three statistical parameters [5] using time domain which are activity, mobility and complexity to get time domain information of signal. The parameters have been used for EEG feature extraction earlier for emotion recognition [6].

- *Activity:* It is square of standard deviation of signal *x(n)*, as given in equation 1, where T belongs to number of time samples.

$$A(x(n)) = \frac{\Sigma_{t=1}^{T}[x(n)-\mu]^2}{T} \tag{1}$$

- *Mobility:* Mobility gives root of the ratio between derivative of signals' variance and its variance [5] as shown in equation 2

$$M(x(n)) = \sqrt{\frac{\sigma(x'(n))}{\sigma(x(n))}} \tag{2}$$

  where $x'(n)$ represents differential of EEG signal *x(n)* and σ represents variance of the data.
- *Complexity:* Complexity provides ratio of derivative of mobility to the mobility of signal as shown in equation 3. Value of complexity lies in range of [0,1], here, 1 shows that signal is similar to sine wave [5].

$$C(x(n)) = \frac{M(x'(n))}{M(x(n))} \tag{3}$$

  Where $x'(n)$ represents differential of EEG signal *x(n)*.

2.2. k-Nearest Neighbor Classifier

k-Nearest Neighbor classifier [7] is a non-parametric approach, which classifies a given data point according to the majority of its neighbors. The KNN algorithm completes its execution in two steps, first finding the number of nearest neighbors and second classifying the data point into particular class using first step. To find the neighbor, it makes use of distance metrics like euclidean distance as given in equation 4

$$Distance(x, y) = \sqrt{\Sigma_i (x_i - y_i)^2} \tag{4}$$

It chooses nearest k samples from the training set, then takes majority vote of their class where k should be an odd number to avoid ambiguity. Figure 2 illustrates architecture of KNN classifier. There are 2 classes, namely class 1 and

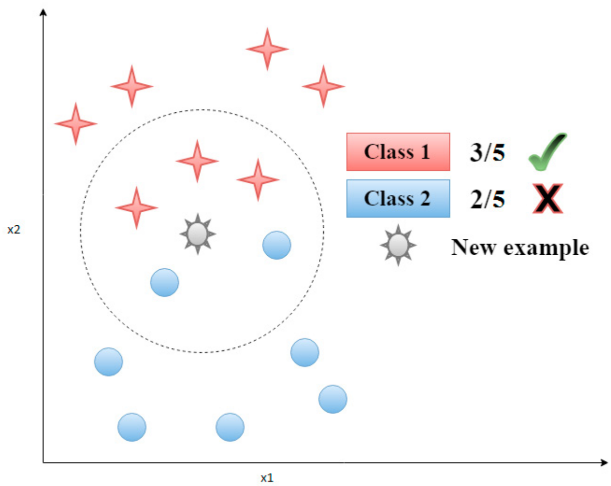

Fig. 2. K-nearest neighbor

class 2. The red asterisks indicate class 1 and blue circles indicate class 2. K chosen is 5, and among the 5 nearest neighbors, 3 samples belong to class 1 and 2 samples belong to class 2. The KNN classifier works on the principle of giving new sample to the class with majority of votes in the defined K. So, the new test example is assigned to class 1.

2.3. Data Acquisition

Data acquisition has been performed using an EEG device from 10 participants among which 9 are male and one is female. The age to participants vary from 20 years to 25 years. They are not found with any type of psychological disorder and are having normal and corrected vision. They have given a written informed consent to the experimenter before starting the data recording. The EEG data recording is done by placing Ag/AgCl electrodes at Fz, FC1, FC2, C3, Cz, C4, CP5, CP1, CP2, CP6, P3, Pz, P4, O1, Oz and O2 sites following 10-20 international system for electrode placement as shown in figure 3. Electro-Occulograph (EOG) recordings are done by capturing Vertical EOG (VEOG) and Horizontal EOG (HEOG) from right eye. For VEOG an electrode has been placed above and below the eye and for HEOG electrode was placed on outer canthus. As reference, an electrode has been placed on mastoid and another on forehead as ground. For signal acquisition, we have used EasyCap [8] (a 32-Channel EEG Standard Cap Set (Munich, Germany)), a V-amp amplifier, set of 16 electrodes and brain vision recorder [9]. Electrode placement protocol is similar as in [10].

For Concealed Information Test (CIT) participants are instructed to perform a mock crime scenario for two sessions. The participants are divided into two groups i.e. “guilty” and “innocent” groups. Instead of asking question [11], certain set of images are shown on the screen which will act as the stimulus for them and will generate different ERP responses in subjects’ brain. Following the same procedure as performed in previous studies [11][12], in this work also three types of stimuli are presented to the subject i.e. target, irrelevant and probe. Stimuli images are categorized as follows:

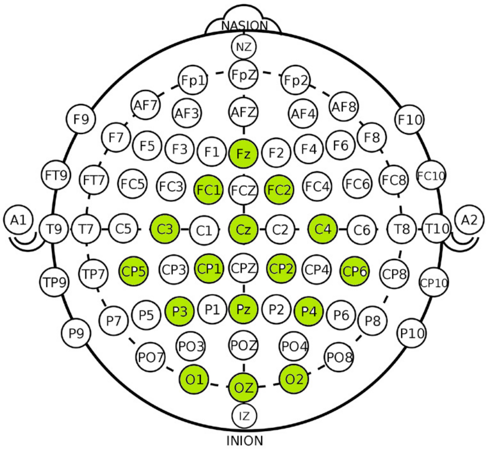

Fig. 3. Electrodes location during experiment.

- Target images: Images of well-known personalities known to all subjects.
- Irrelevant images: Set of random unknown images shown to subjects.
- Probe Images: Images related to crime. Probe will be image of person whom guilty subject knows well and is asked to imagine that he/she has committed crime with (or image of victim).

Before starting the experiment the participants are given a brief description about the experiment and its procedure. After understanding the experimental scenario, the experiment begins with flashing of images (or stimuli) on screen in front of participants. They have to recognize those images and have to answer either “yes” or “no” after each image. The images are presented on a 15.4-inch display screen for 31 seconds. Each image has been displayed for 1.1 seconds followed by a blank image for 2 seconds. So, a total of ten images are presented to subject, out of ten images, seven are irrelevant images, two are target images and one is probe image. These images are randomly presented to the subject. The experiment goes into two sessions, where each session have been conducted for 30 trials for a single subject, so we have 600 trials ($2 \times 30 \times 1 \times 10$). During guilty session, subjects are instructed to respond “yes” for target image, whereas “no” for probe image and irrelevant image. During innocent session, subjects are instructed to respond “yes” for target image, whereas “no” for probe image and irrelevant image.

The analysis of recorded EEG data has been performed using Brain Vision Analyzer 2.1 [9]. EEG signals have been recorded in real environment, hence consist lot of artifacts. To remove high frequency signal, band pass filter have been applied. Band pass filter removes a given band of frequency from the signal without degrading signal quality. In this study, band pass filter is applied in range of 0.3 Hz to 30 Hz. This is the range of frequency usually analyzed when mental task is performed by the subject [13].

3. Results

For analysis of Concealed Information Test , the feature extraction and classification has been applied on EEG data recorded using an acquisition device. The CIT has been performed to analyze human behavior whilst lying. The experiments have been discussed in section 2. The aim here is to perform a binary classification or classify data into two classes namely “guilty” or “innocent”. Identifying guilty is an crucial task hence classifying EEG data properly is one of the major mile to reach. To analyze more precisely the data collected during two sessions, Independent component analysis (ICA) has been applied on raw EEG data. ICA has been applied on guilty session data and innocent session data separately. Infomax ICA has been used on EEG data using MATLAB toolbox. The component maps generated after applying ICA are shown in Figure 4 and 5. Each scalp map in these figures shows the components’ activity for each channel during recording. For further analyses the EEG data is converted into numerical attributes using Brain Vision Analyzer 2.1 [9]. The validation of data is performed using 5-fold cross validation (5-FCV) applied on subject wise EEG data. Here, in this study subject wise single trial analysis has been performed. Experiment is conducted

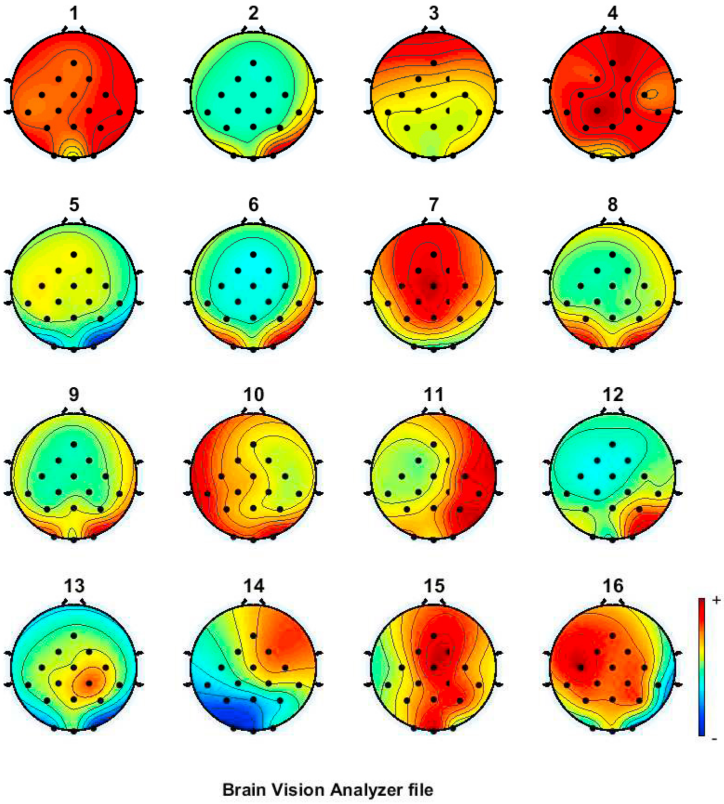

Fig. 4. Component maps for guilty session.

using Intel(R) Core(TM) i7-4790 CPU @ 3.60 GHz, with 8 GB RAM. Furthermore, the implementation is performed using MATLAB 2016a. For calculating the classification performance, various performance measures like accuracy, sensitivity, and specificity have been utilized [14]. From the results as tabulated in Table 1, among all three Hjorth

Table 1. Performance of KNN using different Hjorth Parameters on EEG based CIT data

	Accuracy			Sensitivity			Specificity		
Subject Number	Activity	Complexity	Mobility	Activity	Complexity	Mobility	Activity	Complexity	Mobility
1	60.0	78.2	80.0	65.0	63.3	73.3	64.0	96.0	88.0
2	56.4	66.0	85.5	60.0	76.0	92.0	80.0	56.0	80.0
3	50.0	73.3	78.3	90.0	93.3	70.0	50.0	53.3	86.7
4	54.0	86.0	94.0	44.0	92.0	96.0	56.0	80.0	92.0
5	50.0	70.0	96.0	68.0	68.0	96.0	52.0	72.0	96.0
6	60.0	36.7	96.7	70.0	50.0	93.3	50.0	23.3	100.0
7	43.3	61.8	70.0	56.7	56.7	66.7	50.0	68.0	73.3
8	55.0	65.0	68.3	80.0	80.0	86.7	56.0	50.0	50.0
9	45.0	56.7	76.7	76.7	86.6	96.7	53.3	26.7	56.7
10	61.7	56.7	73.3	90.0	73.3	80.0	53.3	40.0	66.7
Average	53.5	65.0	**81.9**	70.0	73.9	**85.1**	56.5	56.5	**78.9**

parameters, mobility gives better results on recorded EEG data. It has achieved an accuracy of 96.7% for subject 6. An average accuracy of 81.9%, sensitivity of 85.1% and specificity of 78.9% has been achieved. A comparison with existing approaches have been performed applying this work dataset and the results are depicted in table 2. From the table 2 it can be inferred that the accuracy with *k*-NN as classifier has been improved for CIT data. A comparison

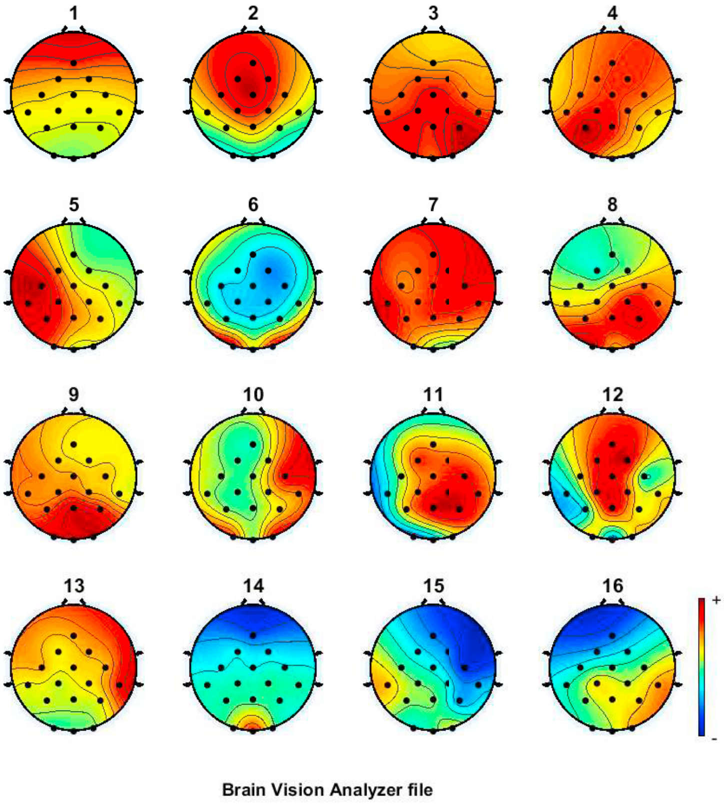

Fig. 5. Component maps for innocent session..

with different dataset applying same feature extraction approach has also been shown. In [6] on emotion recognition dataset, authors have achieved an average accuracy of 35.9%.

Table 2. Comparison with existing approaches

Classification Approach	Feature Extraction Technique	Accuracy	Specificity	Sensitivity	G-measure
KNN [15]	Non papRametric LDA	76.8%	70.0%	73.1%	76.5 %
QDA (Emotion dataset)[6]	Various (power, wavelet, Hjorth parameters etc.)	35.9 %	-	-	-
LDA [16]	EMD	80.1 %	75.7 %	75.7 %	77.8 %
KNN (Proposed)	Hjorth parameters	81.9 %	85.1%	78 .9%	-

4. Conclusion

In this paper, we have proposed an approach to identify deception using brain EEG signals. Deceit identification is one of the challenging tasks as no innocent should be convicted to a crime which he/she has not performed. Hence, the results obtained from this system must be accurate and precise. The proposed approach consists of Hjorth parameters such as activity, mobility and complexity. After performing subject wise analysis, the best results are attained from the mobility parameter yielding upto 96.7% for subject 6. Other performance measures to evaluate the classifier performance are sensitivity, specificity and G-measure. This approach has provided promising results in the binary classification of guilty and innocent classes.

References

[1] Berger, H., 1930. "Ueber das elektrenkephalogramm des menschen." *Journal für Psychologie und Neurologie*

[2] Farwell, L.A., Donchin, E., 1991a. "The truth will out: Interrogative polygraphy (lie detection) with event-related brain potentials". *Psychophysiology* **28**, 531–547.

[3] Haider, S.K., Daud, M.I., Jiang, A., Khan, Z., 2017. "Evaluation of p300 based lie detection algorithm". *Electrical and Electronic Engineering* **7**, 69–76.

[4] Simbolon, A.I., Turnip, A., Hutahaean, J., Siagian, Y., Irawati, N., 2015. "An experiment of lie detection based eeg-p300 classified by svm algorithm", *Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT)*, IEEE. 68–71.

[5] Hjorth, B., 1970. "Eeg analysis based on time domain properties". *Electroencephalography and clinical neurophysiology* **29**, 306–310.

[6] Jenke, R., Peer, A., Buss, M., 2014. "Feature extraction and selection for emotion recognition from eeg". *IEEE Transactions on Affective Computing* **5**, 327–339.

[7] Altman (1992) "An introduction to kernel and nearest-neighbor nonparametric regression". *The American Statistician* **46**, 175–185.

[8] Accessed Date 18-05-2017. easycap. http://www.easycap.de/e/products/products.htm15.

[9] Accessed Date 18-05-2017. Brain products. URL: http://www.brainproducts.com/.

[10] Gao, Lu, Yang, Yu, Na, and Rao (2012) "A novel concealed information test method based on independent component analysis and support vector machine". *Clinical EEG and neuroscience* **43**, 54–63.

[11] Abootalebi, Moradi, and Khalilzadeh (2009) "A new approach for eeg feature extraction in p300-based lie detection". *Computer methods and programs in biomedicine* **94**, 48–57.

[12] Farwell, and Donchin (1991) "The truth will out: Interrogative polygraphy (lie detection) with event-related brain potentials". *Psychophysiology* **28**, 531–547.

[13] Rosenfeld, Soskins, Bosh, and Ryan (2004) "Simple, effective countermeasures to p300-based tests of detection of concealed information". *Psychophysiology* **41**, 205–219.

[14] Zhu, Zen, and Wang (2010) "Sensitivity, specificity, accuracy, associated confidence interval and roc analysis with practical sas implementations". *NESUG proceedings: health care and life sciences, Baltimore, Maryland* **19**. .

[15] Deng Wang, Duoqian Miao, and Gunnar Blohm. 2013 "A new method for eeg-based concealed information test". *IEEE transactions on information forensics and security*, **8(3)**,520-527.

[16] Abdollah Arasteh, Mohammad Hassan Moradi, and Amin Janghorbani,2016. "A novel method based on empirical mode decomposition for p300-based detection of deception". *IEEE Transactions on Information Forensics and Security*, **11**, 2584-2593