Locked-in patients' activities enhancement via brain-computer interface system using neural network

Anas.A.Magour[1], K.Sayed[1], Wael A. Mohamed[2], and M M El Bahy[2]

Abstract**—** Nowadays, there are millions of people around the world suffer from the disability caused by big stroke. In recent years we have seen a rising interest in brain computer interface (BCI) systems that help those patients to practice their normal lives. Therefore, this work presents a GUI application based on an offline BCI system to test their mental capacities. This application was designed based on three tests are alphabet, arithmetic operations and Raven's progressive matrices. The success of this system depends on the choice of the processing techniques. Therefore, Discrete Wavelet Transform (DWT) and Principal Components Analysis (PCA) were used to extract a set of statistical features from the recorded brain signals. These features were classified into four classes are head movement to up, down, right or left using three classifiers are Artificial Neural Network (ANN), Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA). The performance of classifiers was measured using the most frequently statistical parameters: the sensitivity, specificity, precision, classification accuracy, and area under receiver operating characteristics (ROC) curve (AUC). It was concluded that when DWT was used as a feature extraction, ANN and SVM achieved the highest classification accuracy with a value of 95.24% but when using PCA, ANN achieved the highest classification accuracy with a value of 92.86%. On the other hand, LDA classifier was the worst among the three classifiers.

Keywords**—** Artificial Neural Network (ANN), Big Stroke, Brain Computer Interface (BCI), Discrete Wavelet Transform (DWT).

I. Introduction

ACCORDING to the World Health Organization (WHO), there are 15 million patients suffer from a stroke annually all over the world. A stroke or brain attack is defined as a sudden interruption in the blood supply to the brain and it can be occurred to any person at any time [1]. Stroke leads to the death of the brain cells which leads to the memory loss or the disability [2]. The effect of the stroke depends on the place of it in the brain and how much the brain is damaged. Therefore, the patient who has small stroke, he may suffer from a temporary weakness of his arm or leg. While the person who has big stroke, he may be permanently paralyzed or may lose the ability to speak [1]. The disability that result from the stroke affects on an individual in performing one or more of the functions that are essential to daily life, such as self-care, social interaction and the inability to obtain self-sufficiency and to make it in constant need to help others so that he can overcome his disability. However, there is a lot that can be done to improve the quality of life for these patients.

This paper presents a GUI application based on an offline BCI system to test the mental capacities of the patients who suffer from big stroke. There are many tests that seek to test the mental capacities of these patients such as reading, arithmetic operations, alphabet, memory, general knowledge and Raven's progressive matrices (RPM). Raven's progressive matrices (RPM) is a nonverbal test where the questions consist of visual patterns. It is the most common and popular test which is used to measure the ability to think clearly about complex ideas and the ability to store and recall information for the persons who ranging from 5 year to the elderly [3]. Therefore, the proposed application was designed based on three tests: alphabet, arithmetic operations and Raven's progressive matrices (RPM).

A brain computer interface (BCI) is a sophisticated technological system that conveys the commands from user's brain to control the external devices such as a computer, wheelchair, artificial limbs, or other applications without using his muscles [4][5]. BCI is sometimes called a mind-machine interface (MMI), direct neural interface (DNI), synthetic telepathy interface (STI) or brain–machine interface (BMI).This technology contributes to provide a comfortable life for the disabled patients through improving their cognitive abilities and motor skills [6][7]. The general framework of the BCI system is a closed loop system that consists of five stages: signal acquisition, signal pre-processing, feature extraction, signal classification and feedback control of external application as shown in Fig. 1 [8][11].

[1] Biomedical Engineering Department, Benha University, Benha 13511, Egypt
[2] Electrical Engineering Department, Benha University, Benha 13511, Egypt

Anas.A.Magour is a Demonstrator at the Biomedical Engineering Department at Benha Faculty of Engineering in Benha University (corresponding author to provide phone: 002-0109-069-1781; e-mail: anas.abdelraouf@bhit.bu.edu.eg).

K.Sayed is an Assistant Professor at the Biomedical Engineering Department at Benha Faculty of Engineering in Benha University (corresponding author to provide phone: 002-0100-576-9386; e-mail: khaled.sayed@bhit.bu.edu.eg).

Wael A. Mohamed is an Assistant Professor at the Electrical Engineering Department at Benha Faculty of Engineering in Benha University (e-mail: wael.ahmed@bhit.bu.edu.eg).

M M El Bahy is a Tenured Professor at the Electrical Engineering Department at Benha Faculty of Engineering in Benha University (e-mail: mmbahy@bhit.bu.edu.eg).

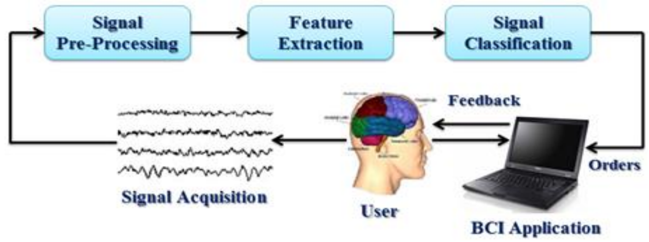

Fig. 1 the general framework of the BCI system

Brain signal acquisition is the measurement of the neurophysiologic state of the brain. EEG signals are extracted by using various types of data collection techniques such as Electro Encephalography (EEG), Functional Magnetic Resonance Imaging, Near Infra-Red Spectroscopy (NIRS) and Magneto Encephalography (MEG). The BCI systems have been extensively studied in research laboratories over the last two decades. Nowadays, the researchers aim to make this technology accessible to everyone. Due to the high prices of the medical EEG recording devices in global market, BCI applications have become difficult to implement. This led to the appearance of many low cost alternative devices such as the Emotiv Epoc headset [9][10]. The acquired brain signals are classified on a frequency basis into five different rhythms [11]:

1) Delta waves (δ): 0 - 4 Hz
2) Theta waves (θ): 4 - 8 Hz
3) Alpha waves (α): 8 - 13 Hz
4) Beta waves (β): 13 - 30 Hz
5) Gamma waves (γ): 30 - 100 Hz

Signal pre-processing or signal enhancement is an important stage because the acquired EEG data could be contaminated by artifacts and noise. This stage aims to improve the signal quality without losing a lot of information to make the signal in a best form for the next two processing stages: feature extraction and signal classification. In this step, artifacts are removed from the EEG data by various techniques such as Common Average Referencing (CAR), Surface Laplacian (SL), Independent Component Analysis (ICA), Common Spatial Patterns (CSP), Principal Component Analysis (PCA), Single Value Decomposition (SVD), Common Spatio-Spatial Patterns (CSSP), Frequency Normalization (Freq-Norm), Local Averaging Technique (LAT), Robust Kalman Filtering, Common Spatial Subspace Decomposition (CSSD), Notch Filtering, etc [6][12].

Feature extraction is the first processing stage that aims to describe the used brain signals in the BCI systems by some relevant statistical properties called the features. These features are collected in a vector named as feature vector. This stage is implemented by various techniques such as Adaptive Auto Regressive parameters (AAR), bilinear AAR, multivariate AAR, Fast Fourier Transformations (FFT), PCA, ICA, Genetic Algorithms (GA), Wavelet Transformations (WT), and Wavelet Packet Decomposition (WPD) [6][12].

Signal classification is an important processing stage that translates the features of the used signals into commands to control the external devices such as computer or wheelchair. This stage is implemented using various classifiers such as Artificial Neural Networks (ANN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Non-linear Bayesian Classifiers (NBC) and Nearest Neighbour Classifiers (NNC) [6][12][13].

Feedback is the procedure that is happening repeatedly in a closed loop to help the user to identify his mental state to help him to perform his tasks. The mental state determines what the user has to do in order to produce brain signals that are used in the BCI system [12].

There are four different types of control signals in the current BCI applications: Visual Evoked Potentials (VEPs), Slow Cortical Potentials (SCPs), P300 Evoked Potentials and Sensorimotor Rhythms (Beta Rhythms). These signals are defined as a certain brain signals which have unique properties and are created unconsciously by stimulations or consciously by performing a cognitive task [14][15].

The BCI systems can be classified into (I) exogenous or endogenous, (II) synchronous or asynchronous and (III) invasive or non-invasive. The BCI systems are classified into exogenous or endogenous on the basis of the nature of the brain signals that are used as input. Exogenous BCI systems use the neuron activity in the brain that is caused by an external stimulus such as P300 evoked potentials or VEPs. On the other hand, endogenous BCI systems depend on self-regulation of brain rhythms and potentials without external stimuli. The BCI systems are also classified into synchronous or asynchronous on the basis of the methods of input data processing. The synchronous BCI systems analyze the brain signals during predefined time windows but, the asynchronous BCI systems analyze the brain signals regardless of the time when the user acts. The BCI systems are also classified into invasive BCI systems that are based on signals recorded from electrodes implanted over the brain cortex and non-invasive BCI systems that are based on signals recorded from electrodes placed on the scalp. Non-invasive BCI system is preferred by the medical researchers because it is very safe, more practical and flexible in the extraction of the brain signals [6][15].

The BCI applications define the method that a BCI is used. Therefore, the BCI applications are divided into five main fields: locomotion, neuroprosthesis, environmental control, entertainment (games) and communication [6][12][15]. Fig. 2 shows the relationships between the types of BCI applications relating to the information transfer rate (ITR) and the user's capabilities for control.

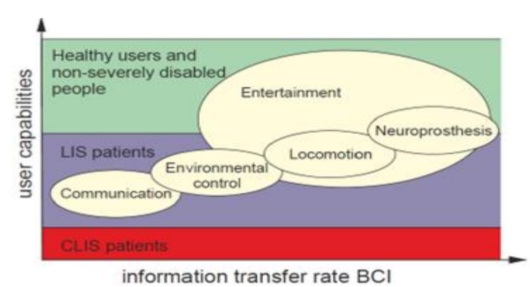

Fig. 2 the relationships between BCI applications relating to ITR and the users capacities [6]

In view of Fig. 2, we will find that most BCI applications are designed for entertaining purposes. The capabilities provided to healthy users and non-severely disabled people are significantly higher than those provided to locked-in syndrome patients. However, there are no applications are offered to completely locked-in syndrome patients. Among the five main areas of BCI applications, communication BCIs has the lowest ITR and the capabilities provided to the users. On the other hand, neuroprosthesis BCIs have the highest ITR and the capabilities provided to the users [6].

I. METHODOLOGY

A. Subjects

In this work, the used dataset was acquired from female students of the biomedical engineering department at Benha faculty of engineering in the morning. Their ages were 22 years and their body weight was almost 65 kg. Those students did not complain of any diseases. There are many external factors that effect on the purity and shape of the brain signal represented in using hair styling products and taking caffeine. Therefore, on the day before the test, students were asked to stop drinking coffee, tea or cola and not using hair sprays, gels, or oils. All recordings were performed according to medically ethical standards and took 2 hours almost. Before recording the EEG signals, students were given all information about the proposed application. They were asked to move their heads many attempts to the up, down, right and left as shown in Fig. 3.

Fig. 3 head movements to the up, down, right and left

B. EEG Data Acquisition

In the current work, the used EEG data was acquired using the Emotiv Epoc headset. The Emotiv Epoc is a device worn on the user's head to record the brain signals that represent the user’s thoughts, feelings, facial expressions and mental commands. Over the last few years, the Epoc has been extensively developed for non-critical BCI applications such as games and communication systems. The Emotiv includes 14 electrodes (plus CMS/DRL references, P3/P4 locations) which are applied to the subject of the experiment according to the 10–20 international electrode placement system. During the recording process, all the standard available electrodes of the Emotiv headset were used where a saline solution was used to reduce the impedance of these electrodes.

The acquired EEG signals are non-stationary signals so they could be contaminated by noise and two types of artifacts are physiological artifacts and non-physiological artifacts. These artifacts can often lie in the same frequency range as the brain signals being recorded. Contamination of EEG data leads to lose large amounts of data that makes analysis of EEG signals very difficult and make the classification results worthless [6]. Therefore, after visual inspection for these artifacts, they were eliminated by discarding the affected EEG signals to make the used signals in a best form for the next two processing stages: feature extraction and signal classification. Fig. 4 illustrates the difference between the typical recorded EEG signal and the contaminated EEG signal for head movement to the left as an example. In addition to, the recorded EEG signals were also filtered using a digital 5th order Sinc notch filter to reject 50 Hz.

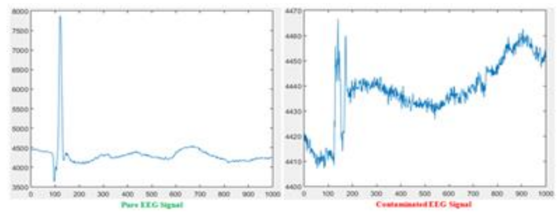

Fig. 4 pure EEG signal and contaminated EEG signal for head movement to the left

C. Feature Extraction Using DWT

Extraction of the statistical information from raw signals is a crucial step in the classification of signals because of its direct effect on the performance of the classification techniques. The decomposition of the signals can be done using various decomposition techniques based on the type of the signal: stationary or non-stationary signal. The signal is considered as a stationary signal if it does not vary much with respect to time while the signal is considered as a non-stationary signal if it varies with respect to time [16]. Therefore, DWT is a good method to analyze the non-stationary EEG signals because it captures transient features and localizes them in both time and frequency [17]. DWT provides high-time resolution if the frequency is high and high-frequency resolution if the frequency is low because it uses long time windows at low frequencies and short time windows at high frequencies to make the time-frequency analysis better [20][22][23].

The main idea of the DWT is analyzing the EEG signal at different frequency bands with different resolutions by decomposing the EEG signal x[n] into approximation and detail coefficients. The detail coefficient D1 is produced from passing the EEG signal x[n] through the high pass filter h[n] while the approximation coefficient A1 is produced from passing the EEG signal x[n] through the low pass filter g[n] then the filtered signals are down-sampled by 2 as shown in Fig. 5. To get the next level of DWT coefficients, the approximate coefficient A1 is again passed through HPF and LPF. Then the output of these filters is down-sampled by 2. The bandwidth of each level of decomposition is half of the bandwidth of the previous level [17][24].

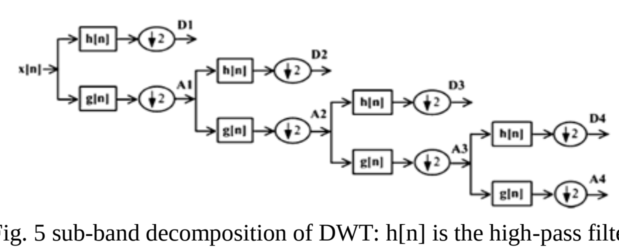

Fig. 5 sub-band decomposition of DWT: h[n] is the high-pass filter and g[n] the low-pass filter [39]

When using the DWT in analysis of the EEG signals, the best wavelet and the suitable number of decomposition levels must be identified. The number of decomposition levels is selected according to the prevailing frequency components of the EEG signal. In this work, four decomposition levels were selected to save time and to keep extra un-useful data that may be obtained. Therefore, the recorded EEG signals were decomposed into four detail coefficients D1–D4 and one final approximation coefficient A4 [18][21][30].

Usually in the wavelet analysis, tests are performed with various types of wavelet families such as daubechies (db), coiflets (coif), symlets (sym), and biorthogonal (bior). The wavelet family that achieves the maximum classification accuracy will be selected. In the present work, daubechies order 4 (db4) was used in the wavelet analysis because it achieved the maximum classification accuracy and it has near optimal time-frequency localization properties, smoothing feature and its waveform is similar to the EEG signal [17][19].

The frequency band $[\frac{f_d}{2} : f_d]$ of each detail coefficient of the DWT is directly related to the sampling frequency of the original EEG signal, which is given by $f_d = \frac{f_s}{2^l}$ where f_s is the sampling frequency and l is the level of decomposition. Table 1 shows the frequency bands of the different decomposition levels for (db4) with a sampling frequency 128 Hz.

Table 1. Decomposition of EEG signals into different frequency bands with a sampling frequency of 128 Hz

Decomposition Level	Frequency Range (Hz)	Frequency Bands
D1	32-64	Gama (γ)
D2	16-32	Beta (β)
D3	8-16	Alpha (α)
D4	4-8	Theta (θ)
A4	0-4	Delta (δ)

The selection of statistical features is an important step in designing the used classification techniques because the classifiers will perform poorly if the used statistical features are not chosen well. The computed discrete wavelet coefficients provide a compact representation that shows the energy distribution of the signal in time and frequency. Therefore, these computed coefficients of the EEG signals of each record were used as the feature vectors. Some statistical values over a group of the wavelet coefficients were used to more decrease the dimensions of the extracted feature vectors. The following statistical features were used to represent the time-frequency distribution of the signals under study [18][20][21]:

1) Maximum of the wavelet coefficients in each sub-band.
2) Minimum of the wavelet coefficients in each sub-band.
3) Mean of the wavelet coefficients in each sub-band.
4) Standard deviation of the wavelet coefficients in each sub-band.

Maximum and minimum values describe the range of observation in the reconstructed signal. Mean value is the center of a group of values. Standard deviation value is used to measure the variability of a data set [40]. These features were extracted from the frequency bands of D1–D4 and A4. Table 2 shows the features extracted from four EEG signals for head movements to the up, down, right and left using DWT.

Table 2. The features extracted from four recorded signals for head movements to the up, down, right and left using DWT

Signal	Features	D1	D2	D3	D4	A4
Up	Max	194.1	1015.8	2435.7	9640.2	114276.2
	Min	-190.1	-2673.7	-1727.3	-5457.5	47315.7
	Mean	-0.014	-10.9	21	70.7	59962.4
	Sta dev	21.6	187.5	306.8	1359.3	12059.5
Down	Max	27.1	120.6	168.3	852.3	68996.5
	Min	-26.3	-97.7	-244.9	-847.8	59277.8
	Mean	0.008	-0.679	-0.022	-7	62810.7
	Sta dev	9.3	32.5	64.5	222.5	2285.1
Right	Max	141.3	1519	2849.4	24893.7	97034.9
	Min	-97.6	-736.1	-1501.4	-11209.8	58630.2
	Mean	0.03	-0.646	15.5	249.3	69543.6
	Sta dev	16	117.6	358.1	3425.1	3961.9
Left	Max	180.8	693.1	754	14772.3	124343.4
	Min	-199.1	-763.4	-1271.5	-7959.8	49477.9
	Mean	-0.007	-1.1	-13.8	118.9	70775.6
	Sta dev	21.1	111.7	244.4	2155.4	15393.2

D. Feature Extraction Using PCA

PCA is maybe the oldest and best known multivariate statistical technique and it is used by almost all researchers to extract a group of statistical features. It was invented in 1901 by Karl Pearson and later developed independently by Harold Hotelling in (1930) [25]. It is dimensions-reduction tool that uses an orthogonal linear transformation to convert a group of correlated variables into a smaller number of linearly uncorrelated variables that is called principal components. These principal components are arranged according to their variance where the first principal component has the maximum possible variance. This variance allows PCA to separate the brain signal into different components [26].

PCA is a well-established technique for reducing the dimensions of the extracted features because the number of principal components is less than the number of original variables. This decrease in the dimensions can reduce the complexity of the classification stage in the BCI systems [6]. PCA was implemented to extract the statistical features from the recorded EEG signals according to the following steps [6]:

Step 1: Read the data matrix and symbolized it by x.

$$x = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix} \quad (1)$$

Step 2: Calculate the mean μ of the data matrix x.

$$\mu = \frac{1}{n} \sum_{i,j=1}^{n} x_{i,j} \quad (2)$$

Step 3: Subtract the mean μ from the data matrix x.

$$X = \sum_{i,j=1}^{n} (x_{i,j} - \mu) \quad (3)$$

Step 4: Calculate the covariance matrix of the data matrix x.

$$Cov = \frac{XX^T}{n-1} \qquad (4)$$

Step 5: Calculate the eigenvalues ($\lambda_i, i = 1,2, \ldots, n$) of the covariance matrix from equation 5 and the eigenvectors ($\alpha_i, i = 1,2, \ldots, n$) of the covariance matrix from equation 7.

$$\det(Cov - \lambda_i I) = 0 \qquad (5)$$
$$Cov\, \alpha_i = \lambda_i\, \alpha_i \qquad (6)$$
$$(\lambda_i I - Cov)\, \alpha_i = 0 \qquad (7)$$

Where, Cov is the square covariance matrix, α_i is the eigenvector corresponding to the eigenvalue λ_i, I is the unit matrix and n is the number of rows.

Step 6: Choosing the eigenvector with the highest eigenvalue which is the principle component of the original data named as a feature vector W.

Step 7: Deriving the low-dimensional new data Y by taking the transpose of the feature vector $W^T$and multiply it by the data after subtracting the mean from it X.

$$Y = W^T X \qquad (8)$$

Finally, some statistics values will be calculated from the low-dimensional new data. The statistical values that were used to more decrease the dimensions of the new data are maximum, minimum, mean and standard deviation. Table 3 shows the statistical features extracted from four recorded signals for head movements to the up, down, right and left using PCA.

Table 3. The statistical features extracted from four recorded signals for head movements to the up, down, right and left using PCA

Signal	Features Extracted	Value
Up	**Maximum**	12972.6
	Minimum	5056.7
	Mean	6419.4
	standard deviation	1348
Down	**Maximum**	7625.5
	Minimum	6362.4
	Mean	6782.9
	standard deviation	252.7
Right	**Maximum**	13663.7
	Minimum	5603.9
	Mean	7516.9
	standard deviation	598.7
Left	**Maximum**	13655.7
	Minimum	5138.4
	Mean	7680.9
	standard deviation	1756.3

E. *Signal Classification Using ANN*

Artificial neural networks (ANNs) are simple processing systems or computational models that are inspired from natural nerve cells. ANNs consist of a huge number of strongly interconnected nodes (neurons) that are used to process the data. In ANNs, the connections between the nodes are named as the weights where the knowledge about the problem has been distributed through these weights. ANNs are widely used by the medical researchers for classifying the biomedical signals because of their special properties such as robustness, self-learning, adaptability and performing massively parallel computations [27][28].

The multi-layer perceptron neural network (MLPNN) as shown in Fig. 6 is one of the most widely used neural network models. It is a nonparametric method to detect and estimate many tasks and also called a feed-forward artificial neural network. The MLPNN is preferred in classifying the brain signals by medical researchers because it has many features such as the ability to learn and generalize, small training group, quick operation and easy performance [17].

The MLPNN consists of several layers of nodes where each layer is connected to the next layer in a directed diagram. A simple MLPNN consists of two layers are an input layer that contains the input variables of the problem and an output layer that contains the solution of the problem. This type of MLPNNs is suitable for linear problems. While, for non-linear problems an additional intermediate (hidden) processing layer is used to convey the information from an input layer to an output layer in one direction to apply some mathematical transformation [29].

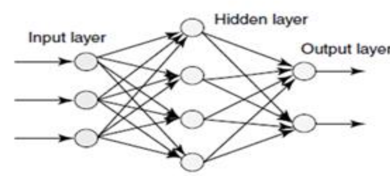

Fig. 6 multi-layer perceptron neural network [27]

The neural network has to be trained to adjust the connection weights and biases in order to produce the desired mapping. Learning in ANNs is accomplished through special training algorithms that are developed based on the learning rules that are supposed to emulate the learning mechanisms of the biological systems [17][28]. There is a number of training algorithms that are used to train the MLPNN where the back-propagation training algorithm is one of the most common training algorithms. The back-propagation training algorithm means that the artificial neurons are organized in layers and send their signals forward and then the errors are propagated backwards. The main idea of the back-propagation training algorithm is to obtain a desired output when certain inputs are given. The training algorithm is an important part of the ANN and becomes suitable when it has a short training time that leads to good classification results [17][27].

F. *Signal Classification Using SVM*

In 1963, Vladimir N. Vapnik and Alexey Ya. Chervonenk invented the original algorithm of support vector machine (SVM). The SVM is one of the most common machine learning methods that classifies the brain signals according to the neural activity of the brain due to its accuracy and ability to deal with a large number of predictors [31][33]. The basic idea of the SVMs is to choose the optimal hyper-plane or a group of hyper-planes to classify the feature vectors to many classes. The optimal hyper-plane is chosen based on the largest margin which is defined as the maximum distance between the nearest training samples as shown in Fig. 7. For the two dimensional data, a single hyper-plane is enough to classify the data. While, two hyper-planes are used to classify the three dimensional data [6][32]. The SVM is robust with regard to the problem of the dimensions that means a large training set is not required for good results even with the high dimensional data [34]. These advantages come at the expense of execution speed [35].

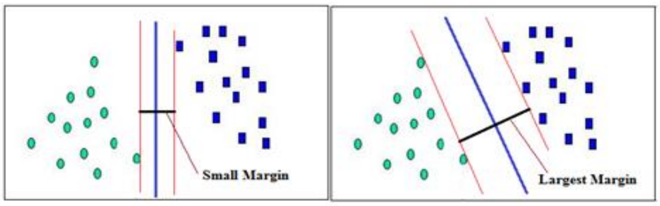

Fig. 7 SVM found the optimal hyper-plane for classification two classes: the circles and the squares

SVM techniques are classified into two types: (i) linearly separable classification and (ii) non-linearly separable classification. Linearly separable classification is used to classify the high dimensional data to two groups without any misclassification or overlapping. While, the SVM that uses non-linearly separable classification is preferred by medical researchers because it leads to a more flexible decision boundary in the data space which may increase the classification accuracy [31]. The SVM that uses non-linearly separable classification is created by some popular kernel functions such as:

1) Linear kernel: $K(u,v) = u^T v$ (9)
2) Polynomial kernel: $K(u,v) = (\gamma u^T v + c)^d, \gamma > 0$ (10)
3) RBF kernel: $K(u,v) = exp(-\gamma * |u - v|^2)$ (11)
4) Sigmoid: $K(u,v) = tanh(\gamma * u^T * v + c)$ (12)

Where $K(u,v)$ is called the Kernel function which is based on the inner product of two variables u and v, γ is the gamma, c is the constant coefficient and d is the polynomial degree. The radial basis function (RBF) is usually used in the BCI applications [32][6].

G. Signal Classification Using LDA

Linear Discriminant Analysis (LDA) is one of the most widely used classification techniques for the BCI systems. The LDA is a very simple classifier that achieves acceptable accuracy without high computation requirements. However, it can lead to completely wrong classifications in the presence of extreme values or strong artifacts [36]. Although the computational requirements of the LDA are limited, LDA is a good choice for designing the online BCI applications with a rapid response [37]. The LDA has been used successfully in many BCI systems such as P300 speller, multi-classes, motor imagery based on BCI or synchronous BCIs [38].

The LDA transforms the data linearly from a high dimensional space to a low dimensional space where the decision is made [38]. The main idea of the LDA is to choose the best discriminant function to classify the data into two classes or more. For two classes, a linear discriminant function that represents by a hyper-plane in the feature space will be used as shown in Fig. 8 (a) while several hyper-planes are used to classify more than two classes as shown in Fig. 8 (b). The hyper-plane can be represented mathematically according to the following equation [6]:

$$f(x) = w^T x + w_0 \quad (13)$$

Where, w is the weight vector, x is the input feature vector and w_0 is the threshold. The input feature vector is assigned to one class or the other based on the sign of $f(x)$.

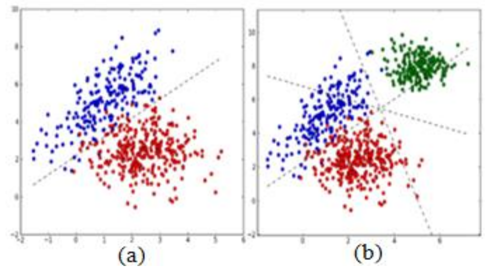

Fig. 8 (a) a hyper-plane separates 2 classes and (b) several hyper-planes separate more than 2 classes [6]

H. Proposed BCI Application

In this work the proposed application was designed based on three tests are alphabet, arithmetic operations and Raven's progressive matrices. This application has been implemented by using MATLAB GUI. It can be divided into six stages. First and second stages consist of four sub-stages; each sub-stage takes ten seconds. The other four stages are separate from each other, and each stage takes 20 seconds. The user can select the suitable stage from the six stages by using six push buttons numbered from 1 to 6 and he can start a new test using NEW TEST push button. During the first stage, the patient is asked to look at a small letter then he is asked to choose the correct capital letter by moving his head to the up, down, right or left. While in the second stage, the patient is asked to solve an arithmetic operation by choosing the correct answer. During the last four stages, the disabled person is asked to complete the Raven's progressive matrix by selecting the suitable picture. Fig. 9 shows stage 1, stage 2, stage 4, and stage 5 of the proposed application.

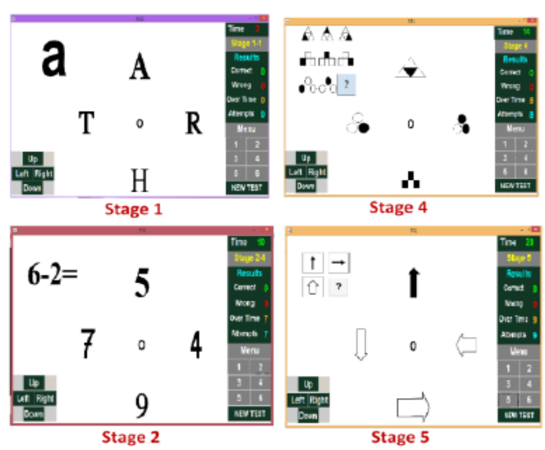

Fig. 9 stage 1, stage 2, stage 4, and stage 5 of the proposed application

II. Experimental Results

The purpose of this paper is to present a GUI application based on an offline BCI system to test the mental capacities of the patients who suffer from big stroke. The proposed BCI system was implemented according to the following block diagram shown in Fig. 10.

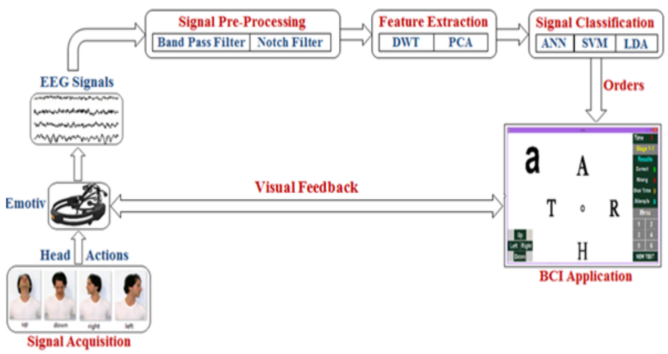

Fig. 10 the block diagram of the BCI system

In the BCI system, the used EEG data was recorded from female students using Emotiv Epoc headset. In the recording process, the students were asked to move their heads to up, down, right and left. The recording data included 135 samples: 28 for head movement to up, 31 for head movement to down, 47 for head movement to right and 29 for head movement to left. In this study, the success of this system depends on the selection of the processing methods. Therefore, this system was implemented using two feature extraction techniques: DWT and PCA to extract a group of statistical features from the EEG signals which were classified into four classes are head movements to up, down, right and left by three classifiers: ANN, SVM and LDA.

The EEG signals were decomposed into four detail coefficients D1–D4 and one final approximation coefficient A4 by using DWT with daubechies wavelet of order 4 (db4). For decreasing the dimensions of the extracted features vectors, a group of statistical features was extracted from the obtained coefficients of each frequency sub-band. These features are the maximum, minimum, mean and standard deviation values. In addition to using DWT, PCA had been used as another feature extraction method to reduce the dimensions of the recorded EEG data to a low dimensional new data. A set of statistical features had been extracted from the low-dimensional new data. These features are the same that were extracted from DWT coefficients.

A two-layer feed-forward network was implemented to classify the recorded EEG signals. The extracted features vectors from the recorded brain signals were divided randomly into 70% of these vectors (93 samples) for training the MLPNN and 30% of the these vectors (42 samples) for testing the MLPNN. The activation function (f) in the hidden layer was sigmoid function while it was softmax (normalized exponential function) in the output layer.

Identifying the best training algorithm is very important in designing the MLPNN. Therefore, scaled conjugate gradient back-propagation algorithm was used for training the desired MLPNN and updating weights and bias values according to gradient descent. Identifying the appropriate number of hidden neurons is also very important in designing the MLPNN. This number of hidden neurons had been determined empirically and the result was that 20-15-4 MLPNN was the optimum model for classification the extracted statistical features from wavelet coefficients. While 4-10-4 MLPNN was the optimum model for classification the extracted statistical features by PCA. The other parameters of the designed neural network are displayed in Table 4.

Table 4. Desired neural network parameters

Neural Network parameters	Feature Techniques	
	DWT	**PCA**
No. Input Neurons	20	4
No. Hidden Neurons	15	10
No. Output Neurons	4	4
Training Function	Trainscg	Trainscg
Performance Function	Crossentropy	Crossentropy
Max No. iterations (epochs)	1000	1000
Initial Weights and Biases	Random	Random

In our system, the library that was developed by Chih-Chung Chang and Chih-Jen Lin (LIBSVM tools) was used for creating the SVM. The choice of kernel function and its parameters are very important in creating the SVM. Therefore, some common kernel functions that are previously presented and different parameters with each kernel function were used separately in building the SVMs. The experimental results suggested that the SVM with a polynomial of degree 3 as a kernel function with the constant coefficient (c) = 0 and the slope gamma (γ) = 30 was the best model for classification the extracted statistical features by DWT. While the SVM with a polynomial of degree 3 as a kernel function with the constant coefficient (c) = 2 and the slope gamma (γ) = 1 was the best model for classification the extracted statistical features by PCA. The Polynomial kernel function is a non-stationary kernel and it can be represented mathematically as [41]:

$$K(u,v) = (\gamma u^T v + c)^d, \gamma > 0 \qquad (14)$$

Discriminant analysis is a classification method that assumes that different classes generate data based on different Gaussian distributions. To train (create) a classifier, the fitting function fitcdiscr in the MATLAB R2015a toolbox is used to estimate the parameters of a Gaussian distribution for each class. The experimental results suggested that the linear discriminant analysis (LDA) was the best model for classification the extracted statistical features by DWT and PCA.

The classification results are displayed by a confusion matrix which is a simple analysis tool used in measuring the performance of the classification techniques. In a two-by-two confusion matrix shown in Fig. 11, each row of the matrix represents the samples in a predicted class, while each column represents the samples in an actual class [42]. The Internal data in the confusion matrix have the following meaning:

1) True positive (TP): - If the class is positive and it is correctly classified as positive.
2) False negative (FN): - If the class is positive and it is incorrectly classified as negative.
3) True negative (TN): - If the class is negative and it is correctly classified as negative.
4) False positive (FP): - If the class is negative and it is incorrectly classified as positive.

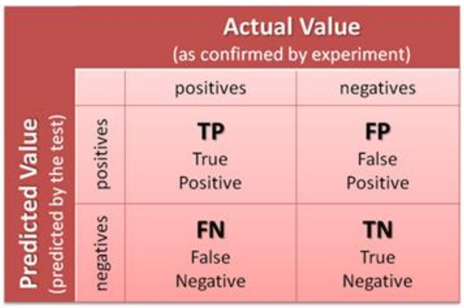

Fig. 11 a two-by-two confusion matrix

The performance of the classifiers was measured by using the most frequently statistical parameters, are the sensitivity, specificity, precision, the total classification accuracy, and the area under receiver operating characteristics (ROC) curve (AUC). These parameters are defined according to the following equations [39]:

1) Sensitivity = TP/ (TP+FN) × 100 (15)
2) Specificity = TN/ (TN+FP) × 100 (16)
3) Precision = TP/ (TP+FP) × 100 (17)
4) Accuracy = Number of Correctly Classified Samples / Total Number of Samples × 100 (18)

The ROC curve is a two-dimensional imaging of classifier performance. It is created by plotting the true positive rate (TPR) (sensitivity) against the false positive rate (FPR) (1 − specificity). The area under ROC curve (AUC) is used to reduce ROC performance to a numeric value that represents the performance of classifier. The value of AUC always ranges between 0 and 1 because AUC is a portion of the area of the unit square. The AUC of a classifier has a significant statistical property which is equal to the probability that the classifier will rank randomly selected positive samples higher than randomly selected negative samples. The AUC is used to distinguish between a pair of classes. If the number of classes equals two, the AUC is a single numeric value. But if the number of classes more than two, the AUC is defined according to Hand and Till equation which is based on calculating an AUC for every pair of classes without using information from the other classes [43].

$$AUC_{total} = \frac{2}{c(c-1)} \sum_{i<j} AUC(c_i, c_j) \qquad (19)$$

Where c is the number of classes and $AUC(c_i, c_j)$ is the area under the two-class ROC curve involving classes $c_i$and c_j. The summation is calculated over all pairs of distinct classes, irrespective of order. In this work, R programming language was used to measure the multi-class AUC as defined by Hand and Till. The relationship between the AUC and classification accuracy was summarized in Table 5 [44].

Table 5. The relationship between the AUC and classification accuracy [44]

Classification Accuracy	AUC Range
Excellent	0.9 - 1.0
Very Good	0.8 - 0.9
Good	0.7 - 0.8
Sufficient	0.6 - 0.7
Bad	0.5 - 0.6
Test Not Useful	< 0.5

After, training and testing the extracted statistical features from DWT and PCA by using ANN, SVM and LDA. The classification results of ANN, SVM and LDA were summarized by four-by-four confusion matrices that are displayed in Table 6, Table 7 and Table 8 respectively. In the confusion matrices that were shown in the following tables, each row of the matrix represents the samples in an actual class, while each column represents the samples in a predicted class.

Table 6. The confusion matrix of ANN using DWT and PCA

ANN Classifier					
Confusion Matrix					
DWT	**Class Type**	**predicted**			
	Actual	**Up**	**Down**	**Right**	**Left**
	Up	8	0	0	0
	Down	1	10	0	0
	Right	0	0	13	1
	Left	0	0	0	9
Confusion Matrix					
PCA	**Class Type**	**predicted**			
	Actual	**Up**	**Down**	**Right**	**Left**
	Up	8	0	0	0
	Down	0	11	0	0
	Right	0	0	13	1
	Left	0	0	2	7

Table 7. The confusion matrix of SVM using DWT and PCA

SVM Classifier					
Confusion Matrix					
DWT	**Class Type**	**predicted**			
	Actual	**Up**	**Down**	**Right**	**Left**
	Up	6	2	0	0
	Down	0	11	0	0
	Right	0	0	14	0
	Left	0	0	0	9
Confusion Matrix					
PCA	**Class Type**	**predicted**			
	Actual	**Up**	**Down**	**Right**	**Left**
	Up	8	0	0	0
	Down	7	4	0	0
	Right	0	0	13	1
	Left	0	0	0	9

Table 8. The confusion matrix of LDA using DWT and PCA

LDA Classifier					
Confusion Matrix					
DWT	Class Type	predicted			
	Actual	Up	Down	Right	Left
	Up	8	0	0	0
	Down	9	2	0	0
	Right	0	0	12	2
	Left	0	0	1	8
Confusion Matrix					
PCA	Class Type	predicted			
	Actual	Up	Down	Right	Left
	Up	6	2	0	0
	Down	8	3	0	0
	Right	0	0	14	0
	Left	0	0	4	5

To measure the performance of the three classifiers, the classification accuracy, AUC, sensitivity, specificity and precision were measured from the confusion matrices. We concluded that when DWT was used as a feature extraction, ANN and SVM achieved the highest classification accuracy with a value of 95.24% while ANN achieved the highest AUC with a value of 0.9865. But when using PCA, ANN achieved the highest classification accuracy and AUC with values of 92.86% and 0.9755 respectively. On the other hand, the performance of LDA classifier was bad. Table 9 shows a summary of the classification accuracy and AUC of the classifiers. The other statistical parameters are displayed in Fig. 12 when DWT was used as a feature extraction while they are displayed in Fig. 13 when PCA was used.

Table 9. The classification accuracy and AUC of the classifiers

Classifier Name	DWT		PCA	
	Accuracy	AUC	Accuracy	AUC
ANN	95.24%	0.9865	92.86%	0.9755
SVM	95.24%	0.9792	80.95%	0.941
LDA	71.43%	0.9107	66.67%	0.8815

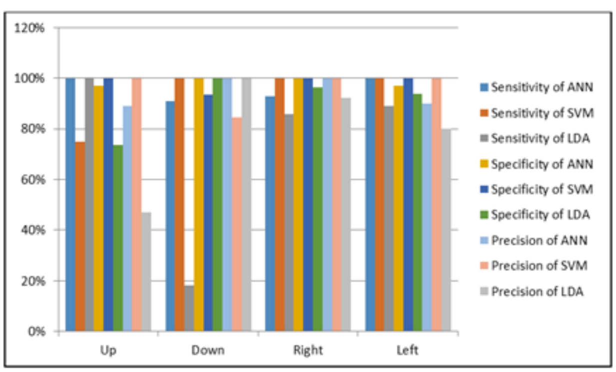

Fig. 12 the comparative performance of the three classifiers with respect to sensitivity, specificity and precision using DWT

As seen in Fig. 12. When ANN was used to classify the features that were extracted by DWT: the sensitivity was the highest in head movement signals to the up and left, the specificity was the highest in head movement signals to the down and right and the precision was the highest in head movement signals to the down and right. But, when using SVM to classify the features from DWT: the sensitivity was the highest in head movement signals to the down, right and left, the specificity was the highest in head movement signals to the up, right and left and the precision was the highest in head movement signals to the up, right and left. While, when LDA was used to classify the features from DWT: the sensitivity was the highest in head movement signals to the up, the specificity was the highest in head movement signals to the down and the precision was the highest in head movement signals to the down.

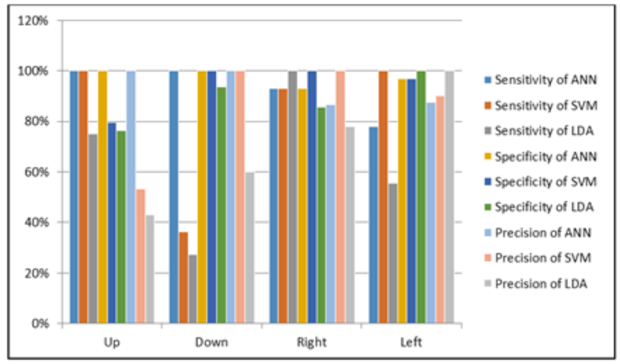

Fig. 13 the comparative performance of the three classifiers with respect to sensitivity, specificity and precision using PCA

While, by looking at Fig. 13, the sensitivity, specificity and precision were the highest in head movement signals to the up and down when ANN was used to classify the features that were extracted by PCA. But, when using SVM to classify the features from PCA: the sensitivity was the highest in head movement signals to the up and left, the specificity was the highest in head movement signals to the down and right and the precision was the highest in head movement signals to the down and right. While, when LDA was used to classify the features from PCA: the sensitivity was the highest in head movement signals to the right, the specificity was the highest in head movement signals to the left and the precision was the highest in head movement signals to the left.

IV. CONCLUSION

This paper presented a GUI application based on an offline BCI system to test the mental capacities of the patients who suffer from big stroke. This application was designed based on three tests: alphabet, arithmetic operations and Raven’s progressive matrices. In our BCI system, the used EEG data was recorded using Emotiv Epoc headset. The success of this system depends on the choice of the processing techniques. Therefore, the proposed BCI system consists of two feature extraction methods: DWT with daubechies wavelet of order 4 and PCA to extract a group of statistical features from the recorded brain signals. These features were classified into four classes are head movement to up, down, right or left using ANN, SVM and LDA. To measure the performance of these classifiers, the classification accuracy, AUC, sensitivity, specificity and precision were measured from the classifiers’ confusion matrices. We concluded that when DWT was used as a feature extraction, ANN and SVM achieved the highest classification accuracy with a value of 95.24% while ANN achieved the highest AUC with a value of 0.9865. But when using PCA, ANN achieved the highest classification accuracy and AUC with values of 92.86% and 0.9755 respectively. On the other hand, the performance of LDA classifier was bad. From the results presented previously it turned out that the ANN classifier is the best method to classify the recorded EEG signals in our BCI system because the values of its parameters were the highest.

ACKNOWLEDGMENT

First and foremost, I thank Allah Almighty for giving me strength and perseverance to accomplish this research successfully. This work was supported by the Department of Biomedical Engineering at Benha Faculty of Engineering. I would like to express my sincere thanks and appreciation to my friendly supervisors, Professor Dr. Mahmoud Mohamed EL-Bahy, Dr. Khaled EL-Sayed Ahmed and Dr. Wael Abd EL-Rahman Mohamed for their advices and suggestions.

REFERENCES

[1] B. M. Gund, P. N. Jagtap, V. B. Ingale, and R. Y. Patil, "Stroke: A Brain Attack," IOSR Journal Of Pharmacy, vol. 3, pp. 1–23, 2013.

[2] J. L. Saver, M. Goyal, A. Bonafe, H. C. Diener, E. I. Levy, and V. M. Pereira, "Solitaire™ with the Intention for Thrombectomy as Primary Endovascular Treatment for Acute Ischemic Stroke (SWIFT PRIME) trial: protocol for a randomized, controlled, multicenter study comparing the Solitaire revascularization device with IV tPA alone in acute ischemic stroke," International Journal of Stroke, vol. 10, pp. 439–448, 2015.

[3] B. W. B, H. J. A, B. C. M, R. E. Jan, R. J, and G. R. C, "Development of abbreviated nine-item forms of the Raven's standard progressive matrices test," Assessment, vol. 19, pp. 354–369, 2012.

[4] E. M. Holz, L. Botrel, T. Kaufmann, and A. Kübler, "Long-Term Independent Brain-Computer Interface Home Use Improves Quality of Life of a Patient in the Locked-In State: A Case Study. Archives of Physical Medicine and Rehabilitation," International Journal of Stroke, vol. 96, pp. 16–26, 2015.

[5] R. Xu, N. Jiang, and C. Lin, "Enhanced Low-Latency Detection of Motor Intention From EEG for Closed-Loop Brain-Computer Interface Applications," IEEE Transactions on Biomedical Engineering, vol. 61, pp. 288–296, 2014.

[6] L. F. Nicolas-Alonso, and J. Gomez-Gil, "Brain Computer Interfaces, a Review," Sensors, vol. 12, pp. 1211–1279, 2012.

[7] R. Mampilly, N. Jos, N. Rose, and A. A. M, "Survey on Brain Computer Interface for Paralyzed People," International Journal of Innovative Research in Science, Engineering and Technology, vol. 5, pp. 16565–16568, 2016.

[8] X. Chen, Y. Wangb, M. Nakani, X. Gaoa, T. P. Jungb, and S. Gao, "High-speed spelling with a noninvasive brain–computer interface," PNAS, vol. 112, pp. 6058–6067, 2015.

[9] M. Duvinage, T. Castermans, M. Petieau, T. Hoellinger, G. Cheron, and T. Dutoit, "Performance of the Emotiv Epoc headset for P300-based applications," BioMedical Engineering OnLine, vol. 12, pp. 1–15, 2013.

[10] M. J. Khan, M. J. Hong, and K. S. Hong, "Decoding of four movement directions using hybrid NIRS-EEG brain-computer interface," Frontiers in Human Neuroscience, vol. 8, pp. 1–10, 2014.

[11] A. K. Maity, R. Pratihar, A. Petieau, A. Mitra, S. Dey, V. Agrawal, and S. Sanyal, "Multifractal Detrended Fluctuation Analysis of alpha and theta EEG rhythms with musical stimuli," Chaos, Solitons and Fractals, vol. 81, pp. 52–67, 2015.

[12] M. Ahn, M. Lee, J. Choi, and S. C. Jun, "A Review of Brain-Computer Interface Games and an Opinion Survey from Researchers, Developers and Users," Sensors, vol. 14, pp. 14601-14633, 2014.

[13] N. Naseer, M. J. Hong, and K. S. Hong, "Online binary decision decoding using functional near-infrared spectroscopy for the development of brain–computer interface," Experimental Brain Research, vol. 232, pp. 555–564, 2014.

[14] R. M. Alkhater, "Real-time Detection of P300 Brain Events: Brain-computer Interfaces for EEG-based Communication Aids," thesis submitted to Auckland University of Technology, Auckland, New Zealand, 2012.

[15] A. E. Hassanien, and A. T. Azar, "Brain-Computer Interfaces," Intelligent Systems Reference Library, vol. 74, pp. 273–287, 2015.

[16] P. K. Bhatia, and A. Sharma, "Epilepsy Seizure Detection Using Wavelet Support Vector Machine Classifier," International Journal of Bio-Science and Bio-Technology, vol. 8, pp. 11–22, 2016.

[17] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," Expert Systems with Applications, vol. 32, pp. 1084–1093, 2007.

[18] P. Jahankhani, V. Kodogiannis, and K. Revett, "EEG Signal Classification Using Wavelet Feature Extraction and Neural Networks," IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing, 2006, pp. 52-57.

[19] M. MURUGAPPAN, M. RIZON, R. NAGARAJAN, and S. YAACOB, "EEG Feature Extraction for Classifying Emotions using FCM and FKM," The WSEAS Conferences, 2008, pp. 299-304.

[20] E. D. Übeyli, "Combined neural network model employing wavelet coefficients for EEG signals classification," Digital Signal Processing, vol. 19, pp. 297–308, 2009.

[21] D. Cvetkovic, E. D. Übeyli, and I. Cosic, "Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: A pilot study," Digital Signal Processing, vol. 18, pp. 861–874, 2008.

[22] Y. U. Khan, and F. Sepulveda, "Brain-computer interface for single trial EEG classification for wrist movement imagery using spatial filtering in the gamma band," IET Signal Process, vol. 4, pp. 510-517, 2010.

[23] M. HEKiM, "The classification of EEG signals using discretization-based entropy and the adaptive neuro-fuzzy inference system," Turkish Journal of Electrical Engineering & Computer Sciences, vol. 24, pp. 285-297, 2013.

[24] R. Sharma, R. B. Pachori, and U. R. Acharya, "An Integrated Index for the Identification of Focal Electroencephalogram Signals Using Discrete Wavelet Transform and Entropy Measures," Entropy, vol. 17, pp. 5218-5240, 2015.

[25] I. T. Jolliffe, "Principal Component Analysis (Second ed.)," Springer, 2002.

[26] H. Abdi, and L. J. Williams, "Principal Component Analysis," Wiley Interdisciplinary Reviews: Computational Statistics, vol. 2, pp. 1-47, 2010.

[27] P. H. Sydenham, and R. Thorn, "Handbook of Measuring System Design," Wiley, vol. 3, pp. 901-908, 2005.

[28] Y. Kumar, M. L. Dewal, and R. S. Anand, "Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network," Signal, Image and Video Processing (SIViP), vol. 8, pp. 1323-1334, 2012.

[29] A. Subasi, and E. Ercelebi, "Classification of EEG signals using neural network and logistic regression," Computer Methods and Programs in Biomedicine, vol. 78, pp. 87-99, 2005.

[30] A. Subasi, "Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients," Expert Systems with Applications, vol. 12, pp. 1–11, 2004.

[31] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," Proceedings of the fifth annual workshop on Computational learning theory, 1992, pp. 144.

[32] P. Bhuvaneswari, and J. S. Kumar, "Support Vector Machine Technique for EEG Signals," International Journal of Computer Applications, vol. 63, pp. 1-5, 2013.

[33] A. Subasi, and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and Support Vector Machines," Expert Systems with Applications, vol. 37, pp. 8659–8666, 2010.

[34] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition," J. Neural Eng, vol. 2, pp. L14–22, 2005.

[35] M. Thulasidas, G. Cuntai, and W. Jiankang, "Robust classification of EEG signal for brain-computer interface," IEEE Trans. Neural Sys. Rehabil. Eng, vol. 14, pp. 24–29, 2006.

[36] K. R. Muller, C. W. Anderson, and G. E. Birch, "Linear and nonlinear methods for brain computer interfaces," IEEE Trans. Neural Sys. Rehabil. Eng, vol. 11, pp. 165–169, 2003.

[37] V. Bostanov, "BCI competition 2003-data sets Ib and IIb: Feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram," IEEE Trans. Biomed. Eng, vol. 51, pp. 1057–1061, 2004.
[38] F. Lotte, M. Congedo, A. Lecuyer, A. Mitra, F. Lamarche, and B. Arnaldi, "A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces," HAL, vol. 4, pp. 24–48, 2007.
[39] H. U. Amin, A. S. Malik, R. F. Ahmad, N. Badruddin, N. Kamel, and M. Hussain, "Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques," Australasian Physical & Engineering Sciences in Medicine, vol. 38, pp. 139–149, 2015.
[40] A. S. Muthanantha, A. S. Anitha, M. Manjureka, and T. Mohanapriya, "MULTICLASS SUPPORT VECTOR MACHINE WITH NEW KERNEL FOR EEG CLASSIFICATION," International Research Journal of Engineering and Technology. Eng, vol. 3, pp. 1191-1197, 2016.
[41] M. M. Murugavel, D. Akshaya, S. Anitha, M. Manjureka, and T. Mohanapriya, "MULTICLASS SUPPORT VECTOR MACHINE WITH NEW KERNEL FOR EEG CLASSIFICATION," International Research Journal of Engineering and Technology, vol. 3, pp. 1192–1197, 2016.
[42] W. M, and D. P, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation," Journal of Machine Learning Technologies, vol. 2, pp. 37-63, 2011.
[43] T. Fawcett, "An introduction to ROC analysis," SCIENCE DIRECT, vol. 27, pp. 861–874, 2006.
[44] A. M. Šimundić, "Measures of Diagnostic Accuracy: Basic Definitions," The electronic Journal of the International Federation of Clinical Chemistry and Laboratory Medicine, vol. 19, pp. 1–9, 2009.