Comparison of Preprocessing Algorithms using an Affordable EEG Headset

Sadiq J. Abou-Loukh, PhD
Assistant Professor
Engineering College-
Baghdad University

Arwa Ra'ad Obaid
M.Sc. Researcher
Engineering College-
Baghdad University

ABSTRACT

Brain Computer Interface is a technology make a communication with the outside world via brain thoughts. The performance of the BCI system depends on the choice of approaches to process the signals of the human brain at each step. The recording signals of a human brain having bad or small signal to noise ratio (SNR) made brain patterns hard to be distinguished. So, the signal quality need to be enhanced, i.e. enhancing the SNR. The electroencephalogram (EEG) signals are composed of true signal and noise signals so that in order to have high SNR, the EEG signals should be transformed so that the undesired components (noise signal) will be isolated and the true signal will remain.

Methods proposed in this paper are for preprocessing, feature extraction and classification of EEG signals (brain signals) recorded from Emotiv EPOC. The raw EEG data is preprocessed to remove noise and then is handled in order to eliminate the artifacts using Principal Component Analysis (PCA), Common Spatial Pattern (CSP), and Common Average Reference(CAR). Power Spectral Density (PSD) is computed from filtered data as a feature. Finally, Support Vector Machine method used to interpret the EEG patterns. The PCA algorithm showed good performance with a value 94.28% compared to other algorithms.

Keywords
Electroencephalogram (EEG), Brain Computer Interface (BCI), Emotiv EPOC.

1. INTRODUCTION
The BCI system is an important technology to express human thoughts to the outside environment. To design a BCI system, the user should produce various brain activity patterns which are captured in the form of Electroencephalogram (EEG) and converted to commands by identifying the patterns by the system. In most BCI, the identification of pattern is based on a classification [1]. EEG offers inexpensiveness, non-invasiveness, ease of use, the possibility of implementation in real-time, and high temporal resolution. BCI researches have essentially contained the recording of EEG signals from electrodes placed on the scalp. Many BCI researches have contained the improving signal pre-processing approaches to generate reliable and precise control of BCI applications [2].

Many paradigms for constructing EEG-based BCI systems were tried out in the last 20 years. The EEG rhythms is a well-known phenomenon occurring in EEG when interacting with stimuli using sophisticated machine learning algorithms to classify the EEG [1]. Alonso and Gomez-Gil offered reviews of the modern BCIs by considering the steps that designing the standard BCI starting from signal acquisition towards classification and designing the control interface steps [3]. Waldert et al. focused on extracting motor cortical

information that causes the movement of the corresponding body part to design an efficient BCI system [4].

Lakshmi et al. concentrated on addressing the different methodologies necessary to be applied in each step of BCI designing with their advantages and drawbacks [5]. Leeb et al. designed a BCI system that controlling a wheelchair to go from one avatar to another and stopping at each one and speaking to it [6]. Teich implemented a BCI system using an affordable EEG headset (Emotiv EPOC EEG headset) [7]. Vidaurre et al. solved the "BCI Illiteracy" issue, that there is (15% - 20%) of people who don't have the ability to control a BCI. Consequently, they also try to make the BCI more adaptive such that they do not solely rely on off-line calibration when using supervised learning [8].

Carlson and Millán also designed a BCI that controlling a wheelchair by "left" and "right" commands and moreover prepared the wheelchair to avoid the obstacles to make sure that is safe to be used [9].

BCI system building paradigms use specific changes in the EEG that occur through the presentation of controlled external stimuli and they were implemented with great success. As these EEG changes were through a stimulus, they are called Evoked potentials. For example, many BCI systems that utilized Steady-State Visually Evoked Potentials (SSVEP) which can be elicited by ordering the subject to fixate on a box/checkerboard on an LCD screen flickering steadily were built. A corresponding power increase is identified in the subject's EEG at the same frequency and in the flickering's harmonic frequencies. SSVEP can control a computerized device by flickering many different stimuli at various rates while at the same time allowing a user to shift his/her gaze between different stimuli [1].

Imaginings of the movements of the left or the right hand produce a variation in the event-related de-synchronization (ERD) for both α (8-13) Hz and β (14-30) Hz components of sensorimotor rhythms at the contra-lateral sensorimotor region of the human brain. It is particularly useful for severely paralyzed patients. It also supports a good tool for normal subjects to improve the communications ability with peripheral devices [2].

In the present work, a comparison is made in terms of accuracy between three preprocessing algorithms which are, Principle Component Analysis (PCA), Common Average Reference (CAR), and Common Spatial Patterns (CSP) to remove noise and artifacts from brain signals recorded by 14-electrodes Emotiv EPOC. The EEG signals pass through several steps starting from acquisition the data till the classification of data. These steps will be illustrated in Figure 1.
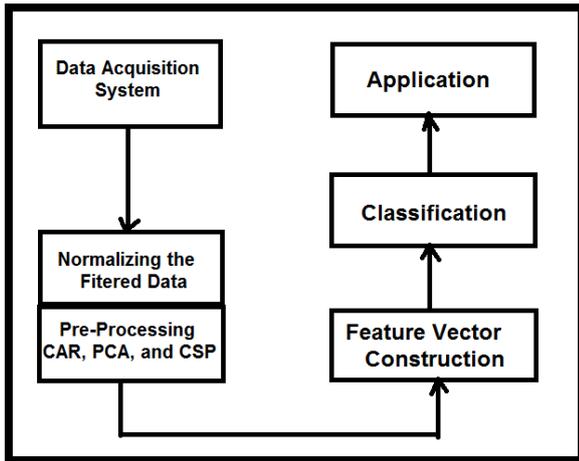
Fig 1: EEG signal processing steps to detect the brain patterns

2. EEG DATA SET

An EEG is a recording of the very weak electrical potential generated by the brain. This section consist of two subsections, the first one illustrates how EEG signals were recorded and by which device this process was accomplished, the second one illustrates how the channel was selected to get the relevant information from brain signals.

2.1 Recording of EEG Signals

The Electroencephalogram data set was recorded from 52 year old man. He did not have any abnormalities in his health and the recording's sampling rate is 128 Hz. The database consists of EEG signals of nodding his head with four positions (right, left, forward, and backward). Every single row in the EEG database matrix represents one electrode only. EEG signal has different frequency bands from which one the features can be extracted depending on the task. Table 1 describes the frequency bands of EEG signals [7].

The EEG signals were filtered by the EEG headset and automatically by the notch filter to remove the 50 AC line noise. Then, the signals were filtered by Sinc band pass filter (order 5) to get the frequency band from 0.2 Hz to 45 Hz. The obtained frequency band is the band that is more informative according to [10] that shows the substantial is increasing in the EEG spectral content below the frequency of 5Hz (delta band brain activity), a medium increase in the range of 5-20Hz (alpha and low beta bands), and a smaller increase of spectra in the higher frequency range can be observed. These EEG data were recorded from Emotiv EPOC which it has 14 electrodes only and 2 reference electrodes (CMS and DRL). The position of each electrode will be identified in Figure 2.

2.2 Channel Selection

Most of the EEG electrodes (channels) seemed to represent redundant information. Furthermore, there are eight channel locations are usually used for Movement-Related Cortical Potentials (MRCP) analysis that covers the areas between central and frontal positions (FC3, FCZ, FC4, C3, C1, CZ, C2, and C4) [11]. According to the EEG headset, the signal that includes the relevant information about head movement can be found on FC4.

Table 1. The frequency range and amplitudes of EEG bands

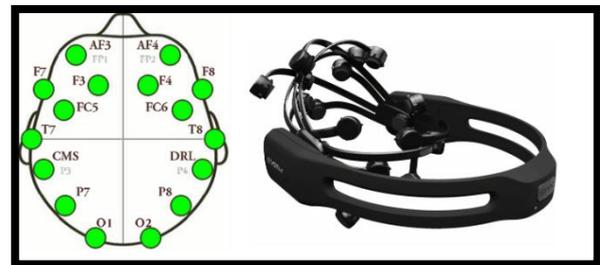| Type of EEG	Frequency (Hz)	Amplitude (μV)
Delta (δ)	1-3	10-300
Theta (θ)	4-7	< 50
Alpha (α)	8-13	~ 50
Beta (β)	14-30	< 30
Gamma (γ)	> 30	very low.

Fig 2: Emotiv EPOC EEG headset and the electrode positions [7].

3. ALGORITHMS
3.1 Principal Component Analysis

Principle Component Analysis (PCA) [12] is a linear transformation technique which converts the correlated variables into uncorrelated variables named principal components. The PCA detects the interior structure of the data by explaining the alteration for the data. PCA may be used in the preprocessing step of BCI design. The PCA may also be used as a dimensionality decreasing method by fixing the alteration of the EEG signals, by accounting for the correlation among variables. The PCA finds the solution to this equation

$$RW = WA \qquad (1)$$

Where R is the covariance matrix of EEG data matrix, W is the eigenvector matrix and finally, A is the diagonal eigenvalue matrix. The PCA accomplishes this purpose by rotating the coordinate structure so that each axes of the new coordinate system will have the directions that the alteration of the data will be maximized when projected for them. The new axes were the eigenvectors of the covariance matrix of the EEG data matrix. The tutorial of this algorithm is as follows:

1: Get some data
In this work, the recorded EEG data set described in Section 2.1 is used. The EEG data set matrix A has two dimensions n×m, where n is the rows represent the channels of the Emotiv EEG headset and m is the columns represent samples of the timed EEG signal.

2: Subtract the mean
For PCA to work properly, the mean should be subtracted from each of the data dimensions.

The mean subtracted is the average across each dimension.

This produced B which is a data set matrix which having a mean equal to be zero.

3: Calculate the covariance matrix of the data set resulted from step 2

This is done in precisely by the following equation:

$$C = B^T B / (n - 1) \tag{2}$$

Where B is the EEG data matrix with zero means.

4: Calculate the eigenvectors and eigenvalues of the covariance matrix

Since the covariance matrix is square, the eigenvectors and eigenvalues of the matrix resultant from step 3 can be calculated. The resultant eigenvectors are considered as principle components and will be organized in descended form (from eigenvector having highest eigenvalues toward eigenvector having the lowest eigenvalue).

5: Constructing the new data set

The originated principal component matrix is multiplied by the data set matrix generated from step 2 according to the following formula:

$$Data_{new} = PC' \times B \tag{3}$$

3.2 Common Average Reference

Common Average Referencing (CAR) [13] is a simple technique in computations, and therefore CAR is applicable for both real–time and on–chip applications. It is usually used in EEG signals, where it is essential to detect signal sources at actual noisy recordings, the CAR of array where each row represents single source and the columns represent samples of EEG signals that can be computed by evaluating the mean of all "good" recording channels (sample by sample), generating only one main reference of all sites. So that the noise will be removed by subtracting the mean from the desired source. In this case, the SNR will be enhanced.

There are various recording references electrode locations such as linked-ears, vertex (Cz), ipsilateral-ear, linked-mastoids, contralateral-ear, bipolar references, C7 reference, and the tip of the nose. The best reference may be chosen which will produce topographic distortion if the relatively electrically neutral area is not employed. The techniques that don't depend on the references will not suffer from problems associated with an actual physical reference. In this work, the DRL electrode will be used as a reference in the CAR approach.

Below is the function to calculate CAR for FC6 electrode (FC6 was selected as an example, and the whole number of electrodes was 14 electrodes):

$$CAR = FC6 - \frac{AF3 + AF4 + \cdots + Oz + O2 + DRL}{15} \tag{4}$$

3.3 Common Spatial Patterns

Common spatial pattern (CSP) [14] is a spatial filter which has been the most popular technique in brain computer interface (BCI). It enables discriminating ERD/ERS effects with ease. The CSP finds directions that maximize variance for one class and at the same time minimize variance for the other class given two different classes embedded in a high-dimensional space. This method is based on the simultaneous diagonalization of two covariance matrices and is applied successfully to the classification of movement-related EEG which is a mathematical method used in signal preprocessing for splitting a multivariate signal into additive subcomponents that have maximum differences in between two windows. CSP was first defined by Koles and it can distinguish abnormal EEG activity. CSP implements the transformation of EEG signal into a variance matrix that maximally discriminates between various classes by using spatial filtering which detects the patterns in EEG. It is sensitive to the electrode positions and artifacts. During the training process, the identical electrode positions are to be preserved to detect the same signals. The increase in efficiency may occur due to the change in electrode positions.

Think through an arbitrary variable $x \in R^N$ that representing the EEG data, which is recorded by N channels which is desired to infer the intention of the user of the BCI system and let c be the class label of this variable x, $c \in C = \{c1,\ldots,cM\}$.

Let as denoting the class probability by $P(c_i)$, where $i = 1, \ldots, M$, and suppose that the EEG data will be conditioned on any class follows a Gaussian distribution with zero means, i.e., $p(x|c_i) = N(0, R_{x|c_i})$, $i = 1, \ldots, M$. This is no restriction in the setting considered here for these details:

1. The linear transformation of x is considered only, and hence any mean can be first subtracted and added again in the end.
2. BCIs based on sensorimotor rhythms typically give the information about the intention of the user from variations in the EEG data in specific frequency bands through conditions. As long as no information restricted in higher moments of x is being used for inference, no information is lost by supposing p(x|c) to follow a Gaussian distribution. A linear transformation is needed to be determined where $W \in R^N \times L$ with $L << N$, such that for finite training data by means of reducing the dimension $\hat{x} = W^T x$ for inferring the intention of the BCI-user leads to an improved classification accuracy, in contrast, to use x.

3.3.1 Binary Common Spatial Pattern

This section deals with two class paradigm i.e., $C = \{c1, c2\}$. The CSP algorithm then solves the following optimization problem:

$$w^* = argmax_{x \in R^N} \left\{ \frac{w^T R_{x|c1} w}{w^T R_{x|c2} w} \right\} \tag{5}$$

Where Rx|c1, Rx|c2 are the covariance matrices of x with the given c1, c2 respectively. Since (5) is in the form of the well-known Rayleigh quotient, solutions to (5) are given by eigenvectors of the generalized eigenvalue problem:

$$R_{x|c1} w = \Lambda R_{x|c2} w \tag{6}$$

The eigenvectors of (6) correspond to the desired spatial filters. Also, for the given eigenvector w^* the corresponding eigenvalue determines the value of the cost function:

$$\lambda^* = \frac{w^{*T} R_{x|c1} w^*}{w^{*T} R_{x|c2} w^*} \tag{7}$$

These eigenvalues are a measure of the quality of the acquired spatial filters, i.e., the eigenvalue associated with a spatial filter states the ratio of the variance between conditions of the component of the EEG data extracted by the spatial filter. The preprocessing is then generally completed by combining the L eigenvectors from the equation (6) with the selected eigenvalues to form $W \in R_N \times L$ and computing $\hat{x} = W^T x$.

The binary CSP algorithm is detailed as follows:

```
Binary CSP Algorithm

Input: a set of l preprocessed EEG trials Xi
with known class
Yi 2 f1; 2g.
Output: W projection matrix
1: for i = 1 to l do
2: PXi = Cov (Xi) {Compute Sample
Covariance matrix}
3: end for
4: P1 = A(PXi ; Yi = 1) {Arithmetic mean of
covariance matrices for class 1}
5: P2 = A(PXi ; Yi = 2) {Arithmetic mean of
covariance matrices for class 2}
```

3.3.2 Multi-class Common Spatial Pattern

The binary CSP can be extended to multi-class CSP by execution two-class CSP on various combinations of classes, i.e. by determining the CSP for one class against all other classes, or by joint approximate diagonalization (JAD). Given EEG data from M different classes, the goal of CSP by JAD is to find a transformation $W \in R_N \times N$ that diagonalizes the covariance matrices $Rx|ci$, i.e.

$$W^T R_x|c_i\ W\ =\ Dc_i\ , i\ =\ 1, \ldots, M \qquad (8)$$

With $Dci \in RN \times N$ diagonal matrices.

There are several methods to this problem. The idea of using JAD for multi-class CSP lies in the point that CSP for two classes can be assumed as diagonalizing two covariance matrices. More exactly, if the eigenvector of the generalized eigenvalue equation (6) is combined in a matrix W, then $W^T R_x|c_i\ W\ =\ Dc_i$ i = 1, . . . , 2. It then seems probable to extend CSP to multi-class CSP algorithm by finding the transformation matrix W which is approximately diagonalized multiple covariance matrices. A total of L columns of the obtained matrix W is then taken as the chosen spatial filters. Here, the CSP for multi-class is achieved by determining the CSP for one class against all other classes.

3.4 Support Vector Machine

Classification and detection of brain patterns are essential steps. The human's brain creates patterns using several mental tasks corresponding to the mental schemes, so BCI will detect and classify /interpret them into suitable commands. This step is accomplished using Support Vector Machine (SVM) [15] which is a supervised learning model correlating with learning classification algorithm. This method is used by a large number of researchers because of its properties like, empirical and structural risk minimization principles that are considered as good features for SVM. A hyperplane makes a perfect gap among different sets of data to be located in its true class. Support Vectors will be the nearest points towards the hyperplane and SVM model guarantee maximum space from the hyperplane for nearby support vectors of the two classes. The Data that will not separate linearly will be converted into higher dimensional mapping for classification. The non-linear SVM converting the original sample data into higher dimensional mapping is called Feature Mapping and its mapping function is denoted as Φ(xi). The kernel function is used to determine the value of mapping function Φ.

$$X_i^T\ X_j = K(X_i, X_j) = \Phi(xi)^T \Phi(xj) \qquad (9)$$

$K(X_i , X_j)$ is called the Kernel function which is based on the inner product of two variants X_i and X_j . In original space, the dot product is used in the calculation and it is converted into higher space that can replace dot products as the kernel function. In this paper, the Polynomial Kernel will be used for good separated classes. The kernel function is as follows:

$$K(X_i, X_j) = [(Xi^T\ Xj\ ) + 1]^d \qquad (10)$$

3.4.1 Linearly Separable

Linearly Separable classification as illustrated in Figure 3 separates the high dimensional data into two sets = {+1, -1} without any misclassification or overlapping. SVM produced a number of decision boundaries or margins where the best margin is recognized by using perceptron algorithm.

The main objective of the Support Vector Machine (SVM) is maximizing the margin width in order to decrease the misclassification error.

$$w_1x_1 + w_2x_2 - b = 0 \qquad (11)$$

$$w_1x_1 + w_2x_2 - b = 1 \qquad (12)$$

$$w_1x_1 + w_2x_2 - b = -1 \qquad (13)$$

Where w_1 and w_2 are the positions of the hyperplane H1 and H2 respectively. x_1 and x_2 are data points and takes the value of +1, 0, - 1 which shows how far hyperplanes away from the original line.
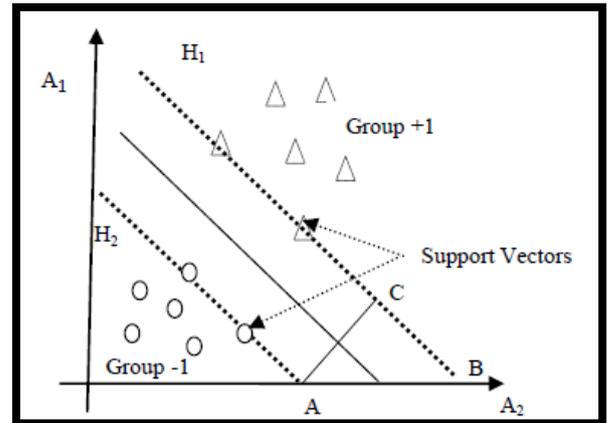

Fig 3: Linear hyperplane that separates one class from another [15]

3.4.2 Linearly Inseparable

The SVM supports for the data holds noisy, errors, and faulty data which having some possibilities of error rate. It is unthinkable to construct a linear hyperplane without error for binary classification data as shown in Figure 4. Linearly inseparable classification can produce solutions for high-dimensional data sets with overlapped or misclassified or erroneous data. Slack variable ξ is used to represent the error term with slight modification in constraint (equations (14 and 15)) and allow misclassified points.

$$wx_i - b \geq +1\text{-}\ \xi \text{ for } y_i = +1 \text{ where } \xi \geq 0 \qquad (14)$$

$$wx_i - b \geq -1+ \xi \text{ for } y_i = -1 \text{ where } \xi \geq 0 \qquad (15)$$

$$y_i(wx_i - b) \geq \xi - 1\ , \xi \geq 0 \qquad (16)$$
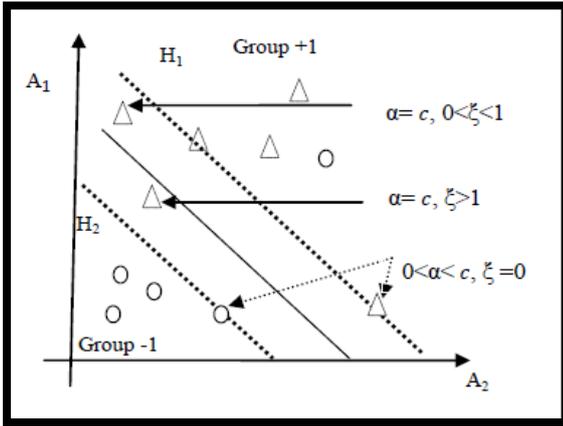
Fig 4: Linear hyperplane that separates one class from another with a slack variable [15].

3.4.3 Non-Linearly Separable

Data are not linearly separable can be converted into higher dimensional mapping for classification as shown in Figure 5. The non-linear mapping of original sample data is transformed into higher dimensional mapping are called Feature Mapping and its mapping function is denoted as $\Phi(xi)$. The kernel functions are used to find the value of mapping function Φ.

$$x_i^T x_j = K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j) \qquad (17)$$

$K(x_i, x_j)$ is called the Kernel function which is based on the inner product of two variants x_i and x_j. In original space dot product of $x_i \cdot x_j$ is used for calculation and it is converted into higher space can be replaced dot products as the kernel function.
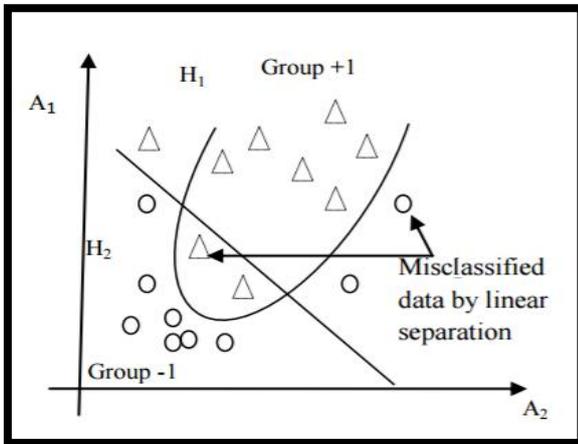
Fig 5: Non-Linear hyperplane that separates one class from another [15].

Some of the popular Kernel functions are as follows:

1. Radian Kernel Function (RBF)

$$K(x_i, x_j) = e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (18)$$

2. Linear Kernel Function
$$K(x_i, x_j) = x_i^T x_j \qquad (19)$$

3. Polynomial Kernel Function
$$K(x_i, x_j) = [(x_i^T x_j) + 1]^d \qquad (20)$$

4. Gaussian Function
$$K(x_i, x_j) = \exp(-[||x_i - x_j||^2]/2\sigma^2) \qquad (21)$$

4. PREPROCESSING

The EEG signal that comes from the human brain, which represents any activity that happens is mixed with various interior and exterior noises which named by artifacts. The largest proportion of these artifacts can be removed or reduced by applying simple filters. For that reason, the desired frequency bands which can be selected by using BPF. The preprocessing step aims at simplifying subsequent processing operations without losing relevant information. An important goal of preprocessing is to improve signal quality by improving the so called signal-to-noise ratio (SNR). A bad or small SNR means that the brain patterns are buried in the rest of the signal (e.g. background EEG), which makes relevant patterns hard to detect. A good or large SNR, on the other hand, simplifies the BCI's detection and classification task. Transformations combined with filtering techniques are often employed during preprocessing in a BCI. Scientists use these techniques to transform the signals so unwanted signal components can be eliminated or at least reduced. These techniques can improve the SNR as given in [2]. In the present paper, EEG signal will be processed by three steps as illustrated below:

1. The noise included in EEG signals caused by AC line will be eliminated by applying a Notch filter to remove (50 Hz), then a Sinc digital BPF was applied to extract (0.2-45) Hz frequency band.

2. EEG data will be normalized to have zero mean and unit variance.

3. Finally, the artifacts will be removed by using these algorithms (PCA, CAR, and CSP).

5. FEATURE EXTRACTION AND CLASSIFICATION

The brain patterns used in BCIs are characterized by certain features or properties. For instance, amplitudes and frequencies are essential features of sensorimotor rhythms and SSVEPs. The firing rate of individual neurons is an important feature of invasive BCIs using intracortical recordings. The feature extraction algorithms of a BCI calculate (extract) these features. Feature extraction can be seen as another step in preparing the signals to facilitate the subsequent and last signal processing stage, detection, and classification. Detection and classification of brain patterns are the core signal processing task in BCIs. The user elicits certain brain patterns by performing mental tasks according to mental strategies, and the BCI detects and classifies these patterns and translates them into the appropriate commands for BCI applications.

Power Spectral Density (PSD) [16] was considered as a feature that formulates EEG signals for simplifying the discovery of EEG patterns. Finally, the features are fed to the SVM classifier to recognize the EEG patterns. The PSD is the most important application area in the digital signal. The data segment can be represented as:

$$x_i(n) = \begin{cases} x(n+iD), & n = 0,1,2,\dots,M-1 \\ & i = 0,1,2,\dots,L-1 \end{cases}$$
(22)

Where iD is the starting point for the ith sequence. If D= M, the segment does not overlap and the L of the data sequence is identical to the data segment of Bartlett method. Then the EEG data will be segmented to computing the periodogram.

$$\check{P}_{xx}^i(f) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} x_i(n)\, w(n)\, e^{-j\pi f n}\right|^2 \qquad (23)$$

$$i = 0, 1, \ldots, L - 1$$

Where U is a normalization factor for the power

$$U = \frac{1}{M} \sum_{n=0}^{M-1} w^2(n) \tag{24}$$

The Welch power spectrum estimate is the average of determined periodogram, is

$$P_{XX}^W = \frac{1}{L} \sum_{i=0}^{L-1} \breve{P}_{XX}^{\ i}(f) \tag{25}$$

Mean value of Welch estimate

$$E[P_{XX}^W(f)] = \frac{1}{L} \sum_{i=0}^{L-1} E\, \breve{P}_{XX}(f) \tag{26}$$

6. EXPERIMENTAL RESULTS

EEG data from four various mental states have been considered as illustrating in Section 2.1. The signal has been preprocessed and the feature is extracted by computing PSD to the EEG patterns and fed them to the classifier (SVM algorithm with polynomial kernel and quadratic programming method). The performance of a BCI can be measured in various ways. A simple measure is classification performance (also termed classification accuracy or classification rate). It is the ratio of the number of correctly classified trials (successful attempts to perform the required mental tasks) and the total number of trials. In order to estimate the correct performance of the classifier, the cross-validation technique (CV) will be applied. The Random subsampling cross-validation (RSCV) method performs 7 data splits of the entire data set and the accuracy will be determined in each one. Table 2 represents the efficiency of the three preprocessing methods using seven folds for random subsampling cross-validation experiments. In Table 2, the PCA has the best accuracy and specificity followed by CAR which has lower accuracy.

Table 2. The random subsampling cross-validation results

The Accuracy of Cross-Validation Folds (%)							
Algorithms	F1	F2	F3	F4	F5	F6	F7
CSP	75	85	77.5	82.5	82.5	65	80
CAR	95	82.5	90	80	77.5	87.5	85
PCA	90	92.5	92.5	97.5	100	87.5	100

7. CONCLUSION

This paper focuses on the comparison between the efficiency of classification of EEG patterns of nodding the head in four positions using three preprocessing algorithms (PCA, CAR, and CSP) and PSD as a feature. The noise and artifacts will be removed and then the PSD was computed as a feature of EEG patterns. The EEG signals from FC4 electrode was selected in the classification step to get the result shown in Figure 6. The

PCA algorithm outperforms all recommended preprocessing algorithms with accuracy equal to 94.2%. The CAR algorithm ranked second with efficiency equal to 85.4% because that CAR algorithm does not need to improve signal quality because it's not in any way optimized to do that. The CAR can even lower your signal quality, e.g. a single artifactual channel that affects into all other channels. CSP and PCA methods give a set of filters with the same size as your channels, but CSP algorithm affected by the number of electrodes and its positions, for this reason, this algorithm ranked third with accuracy equal to 78.2%.

In the future, the objective is to improve this system in online applications, like controlling the robot movement. Moreover, the extra data sets have to be classified to get more knowledge in extraction and perfect decision rules.
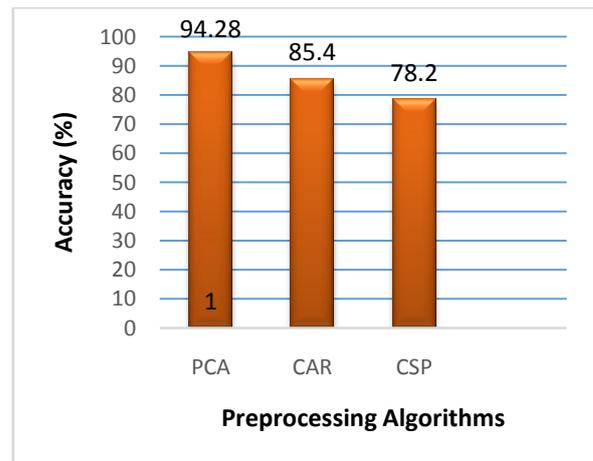

Fig 6: Final classification results.

8. REFERENCES

[1] Padmavathi, R. and Ranganathan, V. 2014. A Review on EEG Based Brain Computer Interface Systems. International Journal of Emerging Technology and Advanced Engineering. Vol. 4, No. 4, 683-696.

[2] Graimann, B., Allison, B.Z., and Pfurtscheller, G. 2010. Brain Computer Interface, 4th (Ed.), Springer, Germany.

[3] Alonso, F., and Gomez-Gil, J. 2012. Brain Computer Interfaces, A Review. Sensors. Vol. 12, No. 2, 1211–1279.

[4] Waldert, S., et al. 2009. A Review on Directional Information in Neural Signals for Brain-Machine Interfaces. Journal of Physiology Paris. Vol. 103, 244–254.

[5] Lakshmi, M., Prasad, T., and Prakash, V. 2014. Survey on EEG Signal Processing Methods. International Journal of Advanced Research in Computer Science and Software Engineering. Vol. 4, No. 1, 84-91.

[6] Leeb, R. et.al. 2007. Self-Paced (Asynchronous) BCI Control of a Wheelchair in Virtual Environments: A Case Study with a Tetraplegic Computational Intelligence and Neuroscience. Computational Intelligence and Neuroscience. Vol. 2007, 1-8.

[7] Teich, P. 2015. Designing a Brain Computer Interface Using an Affordable EEG Headset. Freie University, Berlin, Germany, MSc. Thesis.

[8] Vidaurre, C., Sannelli, C., Müller, K.-R., and Blankertz, B. 2011. Machine-Learning-Based Coadaptive

Calibration for Brain-Computer Interfaces. Neural computation. Vol. 23, No. 3, 791–816.

[9] Carlson, T., and Millán, J. 2013. Brain-Controlled Wheelchairs: A Robotic Architecture. IEEE Robotics and Automation Magazine. Vol. 20, 65–73.

[10] Mihajloví´c, V., Patki, S., and Grundlehner, B. 2014. The Impact of Head Movements on EEG and Contact Impedance: An Adaptive Filtering Solution for Motion Artifact Reduction. Engineering in Medicine and Biology Society (EMBC), 36th Annual International Conference of the IEEE, pp. 5064-5067.

[11] Alomari, M., Samaha, A., and AlKamha, K. 2013. Automated Classification of L/R Hand Movement EEG Signals using Advanced Feature Extraction and Machine Learning. International Journal of Advanced Computer Science and Applications. Vol. 4, No. 6, 207-212.

[12] Mahajan, K., Vargantwar M., and Rajput, S. 2011. Classification of EEG using PCA, ICA and Neural Network. International Journal of Computer Applications, Vol. 1, 80-83.

[13] Ludwig, K., et al. 2009. Using a Common Average Reference to Improve Cortical Neuron Recordings from Microelectrode Arrays. Journal of Neurophysiology. Vol. 101, 1679-1689.

[14] Grosse-Wentrup, M., and Buss, M. 2008. Multi-class Common Spatial Patterns and Information Theoretic Feature Extraction. IEEE. Vol. 55, pp.1991-2000.

[15] Bhuvaneswari, P. and Kumar, J.S. 2013. Support Vector Machine Technique for EEG Signals. International Journal of Computer Applications. Vol. 63, 121-167.

[16] Gupta, H., and Mehra, R. 2013. Power Spectrum Estimation using Welch Method for Various Window Techniques. International Journal of Scientific Research Engineering & Technology. Vol. 2, 389-392.