RESEARCH ARTICLE OPEN ACCESS

Performance Comparison between DWT and PSD as Feature Extraction with SVM Classifier Using DEAP

Maher A. Aljohani[1], Mohammed J. Abdulaal[2]
[1] *Electrical and Computer Department, King Abdulaziz University, Saudi Arabia*
[2] *Electrical and Computer Department, King Abdulaziz University, Saudi Arabia*

ABSTRACT
Support Vector Machine (SVM) is a selective classifier and achieved a good result in terms of performance and accuracy. Thus, this paper present two methods and compare both of them in terms of accuracy and F1-Score. The first system is SVM with Discrete Wavelet Transform (DWT) as feature extraction and the other system with Power Spectral Density (PSD). Principal Component Analysis (PCA) used as a spatial filter and a reduction tool.
***Keywords* –** SVM, PSD, DWT, PCA

Date of Submission: 18-05-2020 Date of Acceptance: 03-06-2020

I. INTRODUCTION

This paper will focus on the field of Machine Learning (ML) for Electroencephalography (EEG) data and will study the performance of DWT and PSD as feature extraction along with SVM and compare them. Thus, there will be an EEG data phase, then spatial filtering phase for extraction the features and classification phase for SVM. In the spatial filtering concerns on reduces a large set of features by defining the dominant features will reflect the large set so the Principal Component Analysis (PCA) will be used. The classification phase will be using Support Vector Machine.

II. METHODOLOGY

The methodology for this research is to design a system with five consecutive blocks and each block has certain operation to fulfil the goal as shown in Figure 1 Each block will be described briefly in this section.

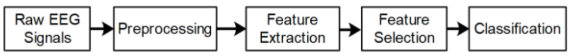

Figure. 1. Block diagram for the proposed framework.

EEG data will be acquired from an open-source dataset. In addition, there are multiple online datasets for emotion recognition such as Temple University hospital repository dataset, DEAP dataset, PhysioNet dataset. DEAP dataset will be used in this study and it is basically a physiological EEG signals database for emotions either positive and negative from different subjects. The data collected from 32 participants while they are watching multiple music videos and each video is rated in terms of valence, arousal, dominance, like, dislike, and familiarity [1].

Most real-world datasets have a large number of features; we might have to deal with thousands of dimensions or features. Dimensionality reduction aims to reduce the number of features - but not simply by selecting a sample of features from the feature-set, which is something else — Feature Subset Selection or simply Feature Selection.

Feature extraction is a process to reduce the number of initial sets in order to be more manageable for processing to decrease the time of computing processes. In other words, it is a method to select or combine variables or both into feature, while the describing and accurately is as the original data set. There are many types of algorithms for time domain and frequency domain and both such as, Fast Fourier Transform (FFT), Wavelet Transform (WT), Eigenvectors, Autoregressive, Common Spatial Patterns (CSPs), and Power Spectral Density (PSD) and so on.

A Power Spectral Density (PSD) is the measure of signal's power content versus frequency. A PSD is typically used to characterize broadband random signals. The amplitude of the PSD is normalized by the spectral resolution employed to digitize the signal [2]. PSD is used to indicate the input channel signals and control it and PSD represents the power of a certain frequency in the brain in EEG-based system.

Wavelet Transform is an important algorithm to the field of recognition and diagnostic. WT reduces data points to a few parameters that represents the signal by compresses the time-varying signal. EEG signal is non-stationary so, the WT is most suitable feature extraction from raw data since it is a time-frequency domain which is a spectral estimation technique thus, any general function can be expressed as infinite series [3-5]. There are two types of WT which are Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT).

CWT can be expressed as:

$$CWT(a,b) = \int_{-\infty}^{\infty} x(t)\psi_{a,b}^{*}(t)dt \qquad (1)$$

x(t) is the unprocessed EEG, a represents dilation and b represents translation factor, where ψ_(a,b) (t) denotes the complex conjugate and it is expressed as follow:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \qquad (2)$$

ψ(t) is stands for wavelet. The scaling parameter a and translation parameter b change continuously as it is major weakness of CWT [6].

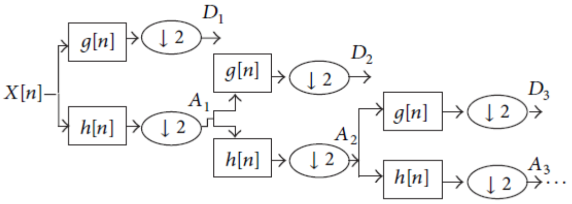

Figure. 1. Decomposition of DWT

DWT has been defined on multiscale future representation and with not that, each scale is under consideration in order to represent a unique thickness of EEG signal [4]. As shown in Figure 2, the decomposition of multiresolution for the raw data is x(n) and each step contains two digital filters g(n), h(n) and two down-sample by two [7]. The component of EEG data with dominant frequency is the factor of choosing the level numbers of wavelet decomposition.

Principal component analysis is a Spatial filter and a reduction tool that can be used to reduce a large set of variables to a small set that still contains most of the large set [8]. So, the idea behind the reduction process is there are multiple of variables does not add any information which will be more logically to reduce it to the minimum as much as possible. This method is commonly used for linear dimension reduction. Let suppose we have data with more than 4-dimentional and need to reduce to 2-dimensional so by using this method will be able to use 2 directions that will capture the data variation.

The input matrix X of dimension $N \times p$:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Np} \end{pmatrix} \qquad (3)$$

The sample covariance matrix of X is given as:

$$S = X^T X / N \qquad (4)$$

If you do the Eigen decomposition of $X^T X$:

$$\begin{aligned} X^T X &= (UDV^T)^T (UDV^T) \\ &= VD^T U^T UDV^T \\ &= VD^2 V^T \end{aligned} \qquad (5)$$

By projecting X onto the principal component's directions, you get the principal components:

$$z_1 = \begin{pmatrix} X_{1,1}v_{1,1} + X_{1,2}v_{1,2} + \ldots + X_{1,p}v_{1,p} \\ X_{2,1}v_{1,1} + X_{2,2}v_{1,2} + \ldots + X_{2,p}v_{1,p} \\ \vdots \\ X_{N,1}v_{1,1} + X_{N,2}v_{1,2} + \ldots + X_{N,p}v_{1,p} \end{pmatrix} \qquad (6)$$

$$z_j = Xv_j = u_j d_j \qquad (7)$$

Function terms:

$U = (u1, u2, \ldots, u_N)$ is an N × N orthogonal matrix.

$V = (v1, v2, \ldots, v_P)$ is an p × p orthogonal matrix.

D is a N x p rectangular matrix with nonzero elements along the first p x p submatrix diagonal.

$D^2$is the diagonal part of matrix D with every element on the diagonal squared.

u_j is the projection of the row vectors of X.

z_j are the principal components of X.

SVM classifier is a selective classifier that will separate the features over a hyperplane in order to facilitate or make the decision easier to discriminate among features. So, it is a model that analyzes the input data and learn from classification and regression analysis. The equation of finding the hyperplane is:

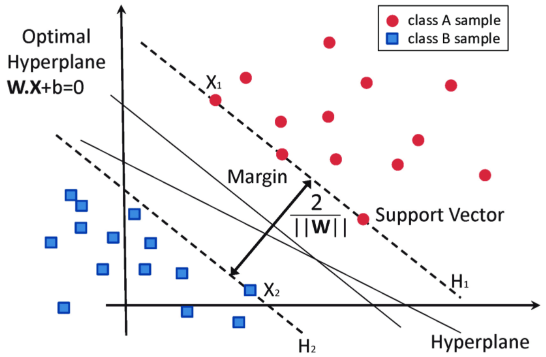

Figure 3: Representation of SVM separation of two features.

There are tuning parameters to control SVM such as Kernel, Regularization, Gamma and Margin. This algorithm uses multiple kernels which are sets of mathematical functions such as linear, nonlinear, polynomial, and radial basis function (RBF). The kernel will take the input data and then it will be transformed into the required form.

III. RESULTS

The aim is to classify emotions using only EEG signals as features with a deep learning classifier using Transfer Learning and compare classification performance with Support Vector Machines classifier. Notably, DEAP dataset has 40 channels to records the data and containing different type of signals such as Temperature, Eye movement, Respiration and EEG. Thus, EEG signals will be used only which is the core of this study and that means will be used only 32 channels out of 40. For preprocessing, the labels binary with values greater than 5 were assigned ‘High’ and values less than or equal 5 were assigned ‘Low’.

Signals comprise multiple frequencies combined at different instances. Discrete wavelet transformation (DWT) was used to decompose the signals into different bands using a succession of high-pass/ low-pass filters. Alternatively, Power spectral density (PSD) was used to obtain the frequency spectrum along with power distribution at each frequency. Statistical features were derived from both DWT and PSD. Then, PCA was used for dimensionality reduction and the resulting data for emotion classification in the case of SVM classifier only since we need as much as we get for deep learning. All these in order to extract the maximum performance from a classifier and minimize the computational resources required and then decomposed the signals into component frequencies and relevant statistical features are derived which characterize the entire signal.

PSD describes the distribution of power over frequency. Very similar to FFT and power spectral density is a very popular transformation method for continuous signals like EEG. For each channel, Power and frequency values for the top 10 peaks were derived and combined with values derived for all other EEG channels to complete the feature set for all the videos. As shown in Figure 4, the EEG signal for channel 2 and then apply it to PSD as shown in Figure 5.

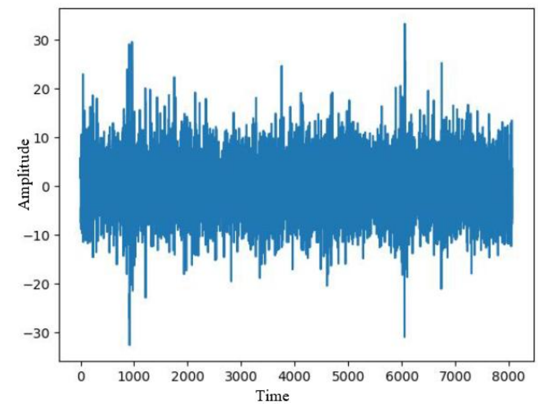

Figure 4: EEG signal sample from subject 1 - video 1 - channel 2.

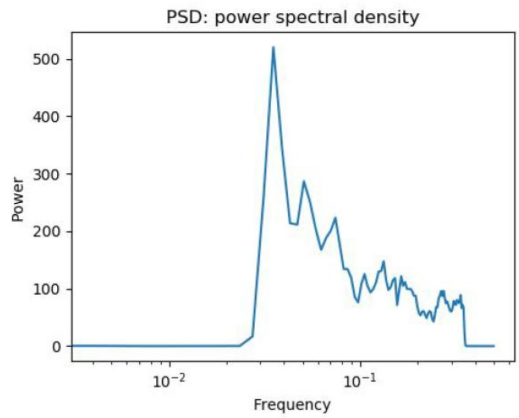

Figure 5: The result of PSD for previous EEG signal.

Wavelet Transformation aims to represent a signal as a linear combination of wavelets, each having a different scale. DWT is a cascade of high/ low pass filters, was used to derive statistical features from the resulting sub bands as shown in Figure 5.

Statistical features are derived from the sub bands: Shannon’s Entropy, Mean, Median and Standard Deviation.

SVM hyperparameters (C, kernel, gamma, tol) were optimized using Grid search with 3 custom validation splits having data for 1, 2, 3 subjects respectively. Then, hyperparameter tuning for best SVM classifier was retrained on the entire training set and the ‘valence’ label class was predicted. SVM classifier folded 32 times to ensure the data validity by reserve one subject each iteration for testing and then all the results will be combined and divided by 32 to get the average accuracy and F1-Score.

With SVM, PCA is used as dimensionality reduction. Using PSD and wavelet transformation the dimensionality of the feature set is greatly reduced but still the number of derived features for each video are high. PCA was used to further reduce the features set length as shown below Figure 6 and 7 for block diagrams of the classifier. For PCA after DWT 30 components were used and PCA after PSD transformation 20 components explaining around 99% variance were used.

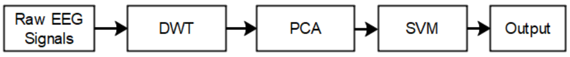

Figure 6: Block diagram for SVM classifier with DWT as feature extraction.

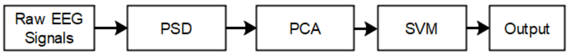

Figure 7: Block diagram for SVM classifier with PSD as feature extraction.

Table 1: SVM classifier results.

SVM with DWT		SVM with PSD	
F1-Score (%)	Accuracy (%)	F1-Score (%)	Accuracy (%)
70.7	55.3	71.9	55.9

According to Table 1, SVM classifier with DWT have better results in terms of accuracy and f1-score which the average accuracy is 55.9% and average f1-score is 71.9% while SVM classifier with PSD is 55.3% and 70.7%.

IV. CONCLUSION

In this paper, the SVM classifier was design with two different feature extraction methods, which are DWT and PSD. The both systems present nearly the same accuracy performance. F1-score are used instead of accuracy in the case of having False Negatives and False Positives which mean having imbalanced confusion matrices. F1 Score is type of accuracy measure as weighted harmonic mean of precision and recall. Overall, there is no big difference between both system.

ACKNOWLEDGEMENTS

The authors of this paper would like to acknowledge and thank King Abdulaziz University for its support in making this project success.

REFERENCES

[1]. Christian M., Touradj E., Jong-Seok L., Anton N. (2011). *DEAP: A Database for Emotion Analysis Using Physiological Signals*, IEEE Transactions on Affective Computing.

[2]. Muhammad L., Henry M. (2016). *Signal Processing for Power Spectral Density.*

[3]. Ales P., Jaromir k., Oldrich V. (2008). *Wavelet transform use for feature extraction and EEG signal segments classification*, in Proceedings of the 3rd International Symposium on Communications, Control, and Signal Processing(ISCCSP'08),pp.719–722.

[4]. I. Daubechies (1990). *Wavelet transform, time-frequency localization and signal analysis*, IEEE Transactions on Information Theory, vol.36, no.5,pp.961–1005.

[5]. S. Soltani, (2002). *On the use of the wavelet decomposition for time series prediction*, Neurocomputing, vol. 48, no. 1, pp. 267–277.

[6]. Alhagry S (2017) *Emotion recognition based on EEG using LSTM recurrent neural network Emotion*.

[7]. Qiao R, Qing C, Zhang T, Xing X and Xu X (2017) *A novel deep-learning based framework for multi-subject emotion recognition*, ICCSS 2017—2017 Int. Conf. Information, Cybernetics and Computational Social Systems.

[8]. Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A. and Yger, F. (2018). *A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update*. Journal of Neural Engineering.