Detection of epileptics during seizure free periods

Conference Paper · May 2013

DOI: 10.1109/WoSSPA.2013.6602363

CITATIONS
2

READS
384

3 authors, including:

M. Amine Hadj-Youcef
Université Paris-Sud 11
6 PUBLICATIONS **5** CITATIONS
SEE PROFILE

Mourad Adnane
National Polytechnic School of Algiers
16 PUBLICATIONS **292** CITATIONS
SEE PROFILE

DETECTION OF EPILEPTICS DURING SEIZURE FREE PERIODS

M.A. Hadj-Youcef[1], *M. Adnane*[1], *A. Bousbia-Salah*[2]

[1] Ecole Nationale Polytechnique, Electrical Engineering Department
10 Avenue Hassen Badi, El-Harrach, Algiers, Algeria
[2] University of Sciences And Technology Houari Boumediene, Electrical Engineering Department
BP-32 El-Alia, Bab-Ezzouar, Algiers, Algeria

ABSTRACT

In this paper the problematic of epileptic detection is treated. An algorithm of EEG signal classification into two classes: Healthy and Epileptics is developed. The difference with conventional methods is the use of free seizure epileptic records. A good classification accuracy means that it is possible to detect an epileptic in normal state or at an early stage of epilepsy. The raw EEG signal is decomposed using discrete wavelet transform (DWT). Then, principal component analysis (PCA) allows dimensionality reduction and better representation of the data. Several features are extracted and used in support vector machine (SVM) classifier. Results show satisfactory classification accuracy comparable or better than those reported in literature.

1. INTRODUCTION

Epilepsy is an illness that affects the nervous system. The number of persons worldwide suffering from that illness is more than 1%. If an early detection and diagnosis are made, 80% of the seizure activity can be controlled or treated effectively [1]. These data statistics show us how important it is to make an early diagnosis of this illness. The conventional way to diagnose epilepsy involves the study of the electroencephalography signal (EEG). EEG can be defined as the electrical activity of the brain and its duration is very long. In addition, these EEG recordings are visually inspected by trained neurophysiologists for detecting epileptic seizures or other abnormalities which make the task very difficult. Researches in this field often make the classification of EEG signals using two sets: healthy and epileptic, where epileptic set contain seizures [2–4]. This approach can not be helpful for an early stage diagnosis of epilepsy. Then, the aim of this work is to develop a method of detection of epileptics during seizure free periods which can somehow be assimilated to an early stage diagnosis of the illness. Then, this method can be used as computer aided diagnosis (CAD) tool by neurophysiology practitioners.

The method uses discrete wavelet transform (DWT) to decompose the signal. Then, valuable features are extracted. Then, principal component analysis (PCA) is applied to obtained features in order to take off redundancy and reduce dimensionality and make better representation of the signal. Obtained data is used in a training/validation scheme using support vector machine (SVM) to classify EEG signals into two classes Healthy and Epileptcs. Results are very satisfactory and the method proved its reliability. The block diagram depicted in Figure 1 illustrates the proposed method steps.

2. PROPOSED METHOD

2.1. EEG data set

We used a well known and publicly available EEG database [5]. A complete description of the database is available in [6]. The data set contains five sets (denoted A-E) of 100 single-channel EEG segments. Sets A and B correspond to healthy volunteers whereas sets C, D and E correspond to epileptics. Sets A and B consist of recordings made in awake state with eyes open and eyes closed, respectively. Sets C and D were measured during seizure free intervals. Only set E contains seizure activity. All EEG signals were recorded with the same amplifier system and an average common reference is used. The data were digitized at 173.61 Hz using 12 bit A/D converter resolution. The band-pass filter range is 0.53-40 Hz (12 dB/oct). In this work, we used set A and set C. Example given in Figure 2 shows two EEG signals one from set A and one from set C.

2.2. Discrete wavelet transform

The Discrete wavelet transform (DWT) is used to make a representation of a signal (EEG) in time and frequency. DWT is used for analyzing non-stationary signals and provides time-frequency representation of a signal at different scales which fits our need because EEG signal is a non-stationary signal which contains several waves at different frequency bands. Many researchers interested in DWT for analyzing EEG signals [2–4]. DWT decomposes a signal using high pass and low pass filters applied at consecutive levels. The high-pass filter, g corresponds to the discrete mother wavelet and the low-pass filter, h is its mirror version.

The down-sampled signals are called approximation (CA1) and detail (CD1) coefficients. Then, approximation and detail coefficients of next level are obtained by decomposing, in the same way, the approximation coefficient of the previous level. Six levels of decomposition

978-1-4673-5540-7/13/$31.00 ©2013 IEEE

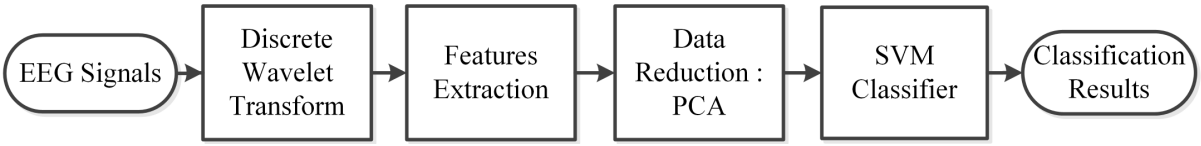

Fig. 1. Block diagram of the proposed method

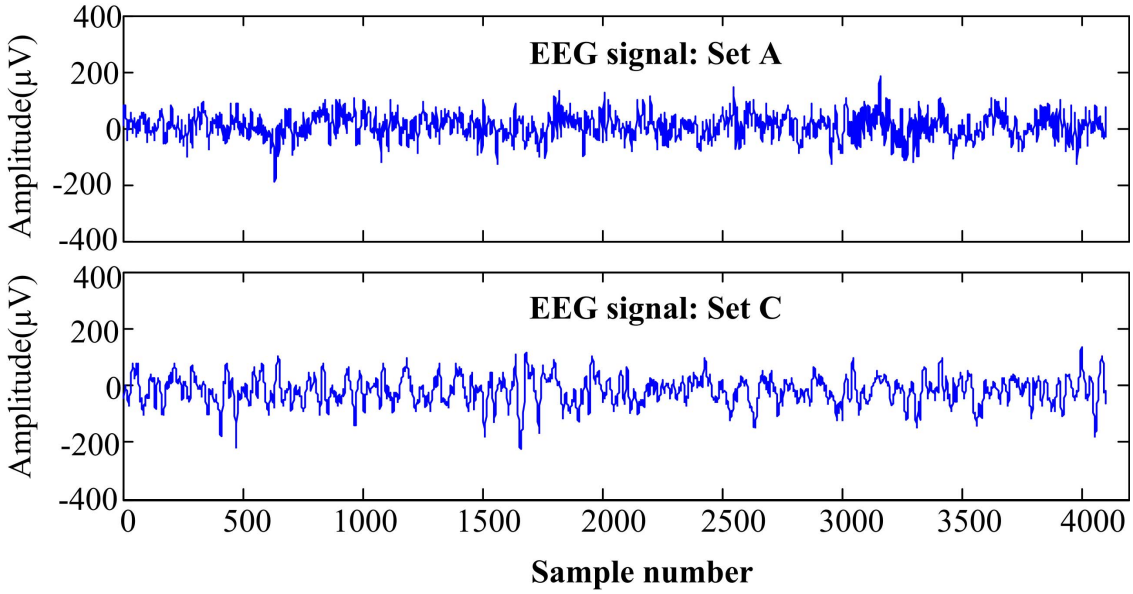

Fig. 2. Example of EEG signal from set A (Healthy) and set C (Epileptic).

was chosen to be used in this work, thus the signal is decomposed into the details CD1-CD6 and one approximation, CA6. The structure of decomposition at level 6 is shown in Figure 3. Figures 4 and 5 show six different levels of approximation and details of an EEG signal taken from a healthy and epileptic subjects, respectively. These approximation and detail coefficients are reconstructed from the Daubechies 2 (db2) wavelet filter.
The DWT is obtained by using two principal equations: the scaling function $\varphi_{jk}(x)$ based on low-pass filter (g), and the wavelet function $\psi_{jk}(x)$ based on high-pass filter (h) , where:

$$\varphi(x) = 2^{j/2} h(2^j x - k) \quad (1)$$

$$\psi(x) = 2^{j/2} g(2^j x - k) \quad (2)$$

Where $x = 0, 1, 2, ..., M-1, j = 0, 1, 2, ..., J-1, J = log_2(M)$, M represents the length of the signal that is chosen as 2^J [7]. k is the sampling rate and resolution j specify the positions and widths of the x-axis function, respectively. The parameter j commands the dilation or compression. The parameter k commands the translation.
The Approximation coefficients $CA_i(k)$ and the detail coefficients $CD_i(k)$ in ith level are described as:

$$CA_i = 1/\sqrt{(2)} \sum_x f(x)\psi_{jk}(x) \quad (3)$$

$$CD_i = 1/\sqrt{(2)} \sum_x f(x)\varphi_{jk}(x) \quad (4)$$

for k=1, 2,...,$2^j - 1$

2.3. Feature extraction

The extracted wavelet coefficients procure a representation that shows EEG signal energy distribution in time and frequency. Then, six features were extracted using wavelet coefficients. Four statistical features were used that are:

- Maximum (MAX) of the wavelet coefficients
- Minimum (MIN) of the wavelet coefficients
- Range (RNG) of the wavelet coefficients
- Standard deviation (STD) of the wavelet coefficients

and two non-statistical features that are:

- Energy (ENG):

$$ENG_i = \sum_{j=1}^{N} |CD_{i,j}|^2, i = 1, 2, ..., l \quad (5)$$

- Entropy (ENT):

$$ENT_i = -\sum_{j=1}^{N} CD_{i,j}^2 \log(CD_{i,j}^2), i = 1, 2, ..., l \quad (6)$$

The extracted features for two records of sets A and C are shown in Table 1.

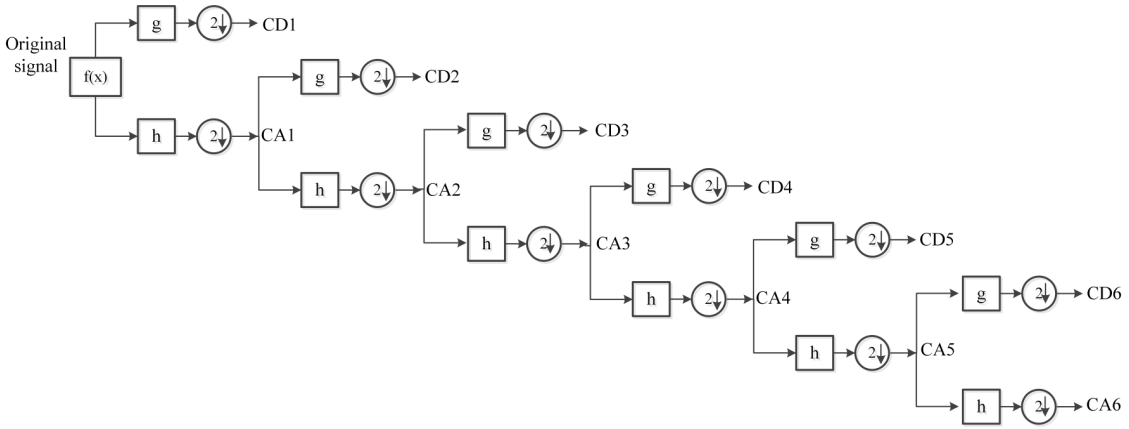

Fig. 3. Structure of decomposition with DWT at level 6.

Table 1. The extracted features for two records of set A and Set C

Data	Features	CA6	CD6	CD5	CD4	CD3	CD2	CD1
SET A	MAX	375.658	273.796	221.230	219.504	166.693	80.577	28.556
	MIN	-471.456	-278.909	-226.240	-221.937	-168.786	-78.625	-27.805
	RNG	47.114	552.705	447.471	441.441	335.479	159.203	56.361
	STD	169.468	110.184	81.086	72.786	51.385	22.627	6.873
	ENG	$4.742\ 10^8$	$8.499\ 10^7$	$8.842\ 10^7$	$1.421\ 10^8$	$1.426\ 10^8$	$5.667\ 10^7$	$1.080\ 10^7$
	ENT	$-1.873\ 10^7$	$-4.356\ 10^6$	$-4.273\ 10^6$	$-6.585\ 10^6$	$-6.253\ 10^6$	$-2.049\ 10^6$	$-2.675\ 10^5$
SET C	MAX	483.129	516.020	374.661	249.823	139.347	59.842	25.180
	MIN	-611.205	-493.418	-379.055	-260.597	-133.484	-57.927	-24.592
	RNG	1094.335	1009.438	753.716	510.420	272.832	117.769	49.772
	STD	225.613	198.525	130.837	78.042	36.423	13.453	4.602
	ENG	$6.684\ 10^8$	$2.987\ 10^8$	$2.510\ 10^8$	$1.801\ 10^8$	$8.782\ 10^7$	$2.568\ 10^7$	$6.179\ 10^6$
	ENT	$-2.568\ 10^7$	$-1.803\ 10^7$	$-1.398\ 10^7$	$-8.990\ 10^6$	$-3.891\ 10^6$	$-8.963\ 10^5$	$-1.542\ 10^5$

2.4. Data Reduction : Principal Components Analysis (PCA)

PCA is a method of multivariate statistics which involves the transformation of correlated variables into new uncorrelated variables named "principal components" or axis. Using PCA, one can make a better representation of data. In addition, the data dimension is reduced to a limited number of components without significant loss of information [8]. Using XLSTAT software [9] in the case of standard deviation (STD) feature, we noticed that we can use only 91.88% of original data to represent the whole data with minimal loss of information.

2.5. Classification : Support Vector Machine (SVM)

SVM classifies data into different classes (2 in present work) by calculating the optimal hyperplane separating those classes. The best hyperplane for an SVM is the one offering the largest margin between two classes. The support vectors are the data points that are used to calculate the separating hyperplane and these points are on the boundary of the slab. The Figure 6 illustrates these definitions, with $+$ indicating data points of class 1, and $-$ indicating data points of class -1.

2.6. Performance measures

To evaluate the performance of our method, we use a set of statistical measures :

- TP (True Positive): Number of epileptic records identified as epileptic records;
- TN (True Negative): Number of healthy records identified as healthy records;
- FP (False Positive): Number of healthy records identified as epileptic records;
- FN (False Negative): Number of epileptic records identified as healthy records;

Moreover, considering these above mentioned statistical measures, performance of the method can be evaluated using the following measures:

1. Accuracy (ACC): It measures the correct classification rate of healthy and epileptic classes.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}.100\% \quad (7)$$

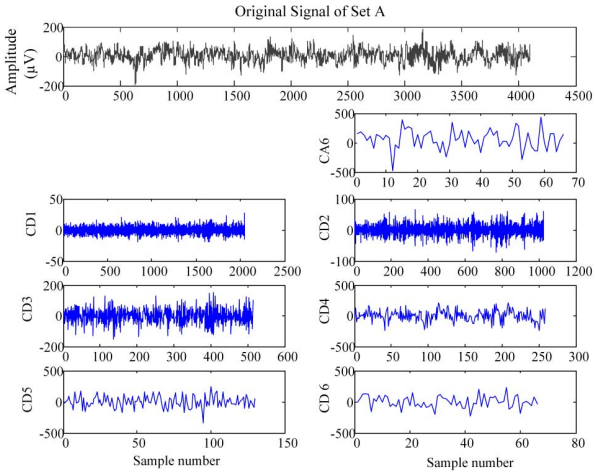

Fig. 4. Approximate and detail coefficients of a sample EEG signal segment taken from a healthy patient with eyes open (set A).

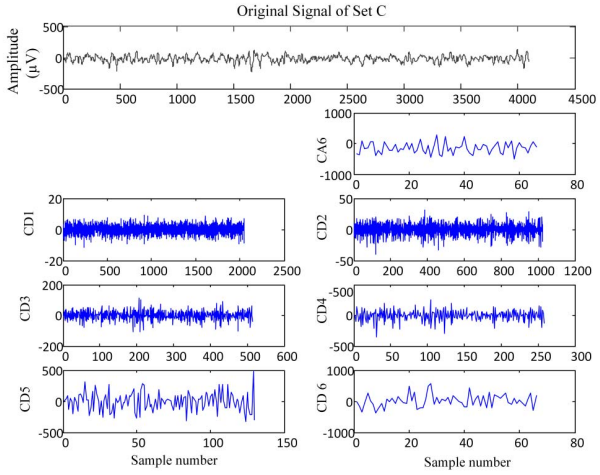

Fig. 5. Approximate and detail coefficients of a sample EEG signal segment taken from a epileptic patient (set C).

2. Specificity (SP): probability that healthy event will be detected when it is actually present.

$$SP = \frac{TN}{TN + FP}.100\% \tag{8}$$

3. Sensitivity (SE): probability that epileptic event will be detected when it is actually present.

$$SE = \frac{TP}{TP + FN}.100\% \tag{9}$$

3. RESULTS AND DISCUSSION

Performance results of the proposed method, using different features (taken separately) is shown in Figure 8. Best result is found when using standard deviation (STD) feature with 10-fold cross-validation. On the other hand, tests were fulfilled combining the two best features which are standard deviation (STD) and entropy (ENT). Results

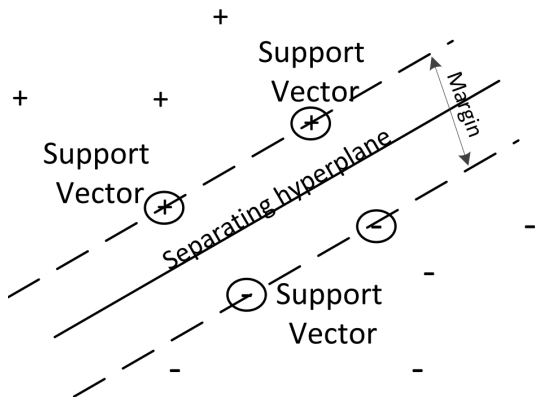

Fig. 6. Support Vector Machine classifier

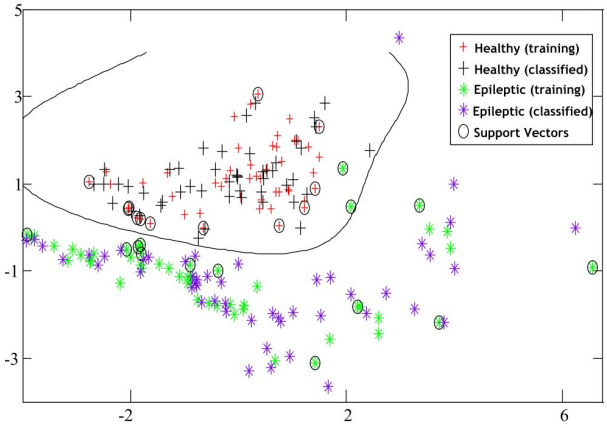

Fig. 7. Example of classification using SVM classifier and standard deviation feature (std) with 10-fold cross-validation

are shown in Table 2. Best accuracy result is ($ACC = 99\%$). This is a good result comparable or better than most published results so far. This is shown when comparing our results to other researchers results. Researchers in [10] used the same EEG database used in this paper but used three sets, such as A (Healthy), C (Epileptics with no seizures) and E (Epileptics with seizures). Since discrimination between healthy and epileptics during seizures is much easier than discrimination between healthy and epileptics with no seizures, we consider that the work made in [10] is somehow comparable to the case of two classes: healthy and epileptics with no seizures which is the case of our present work. Best result obtained in [10] is ($ACC = 98.1\%$) which is lower than our obtained result. Similarly, researchers in [11] used three sets, such as A (Healthy), C (Epileptics with no seizures) and E (Epileptics with seizures). The best result obtained in [11] is ($ACC = 99\%$) which is equal to our obtained result. This brief comparison shows that our method performed very well since it obtains accuracy comparable or better than published results so far. Main findings of this work are summarized in several points:

- The results demonstrates the efficiency of DWT applied to EEG signal. This goes along with our first assumptions.

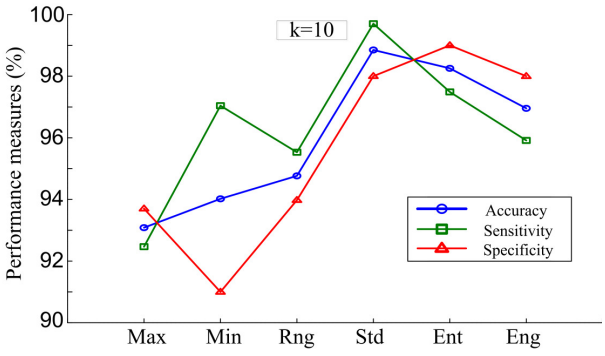

Fig. 8. Performance of classification using different features with 10-fold cross-validation

Table 2. Classification results using standard deviation combined to entropy for different k-fold cross-validation.

Performance	k=2	k=5	k=10
ACC (%)	98,84	98,99	99
SE (%)	98,97	99	99
SP (%)	98,89	99	99

- Best results are obtained when combining features standard deviation (STD) and entropy (ENT). This result can not be commented without deep investigation about feature selection and feature comparison which is our next research work.

- This work can be generalized and enhanced by using different EEG databases. This is necessary if one wants to avoid over fitting problem.

4. CONCLUSION

In this paper, we described a method for EEG signal classification. The two classes scheme was chosen such as class 1 is healthy EEG records and class 2 is the epileptic EEG records with no seizures. The method uses the decomposition of the EEG signals using discrete wavelet transform (DWT). Then features which best represents EEG variability are extracted. Moreover, PCA is used so vectors of features obtained in this operation are transformed to a new data vectors. The new data vectors are less correlated and their dimension is reduced in comparison to initial data. Finally, training and classification are applied to data by the mean of support vector machine (SVM). Results are very satisfactory which means that there is possiblity to detect epliepsy at an early stage of the illness using the developed method.

5. REFERENCES

[1] T. K. Gandhi, P. Chakraborty, G. G. Roy, B. K. Panigrahi. Discrete harmony search based expert model for epileptic seizure detection in electroencephalography, *EXPERT Syst. Appl.*39 (4) (2012) 4055-4062.

[2] E. D. Übeyli, İ. Güler, Features extracted by eigenvector methods for detecting variability of EEG signals, *Pattern Recognit. Lett.* 28 (5) (2007) 592-603.

[3] A. Subasi, M. Ismail Gursoy, EEG signal classification using PCA, ICA, LDA and support vector machines, *EXPERT Syst. Appl.* 37 (12) (2010) 8659-8666.

[4] A. Subasi, EEG signal classification using wavelet feature extraction and a mixture of expert model, *EXPERT. Syst. Appl.* 32 (4) (2007) 1084-1093.

[5] EEG time series data, Department of Epileptology University of Bonn, *http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html* (accessed: 21 December 2012).

[6] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke,P. David, C. E. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, *Phys. Rev. E.* 64 (6) (2001) 61907.

[7] R. C. Gonzalez, R. E. Woods, Digital image processing (2nd ed.). Prentice Hall. (2002).

[8] S. Robin C. Duby, Analyse en composante principales, Institut National Agronomique Paris, *www.agroparistech.fr/IMG/pdf/ACP2006.pdf* (2006), accessed: 21 December 2012.

[9] XLSTAT Software, Running a Principal Component Analysis (PCA) with XLSTAT, *http://www.xlstat.com/en/learning-center/tutorials/running-a-principal-component-analysis-pca-with-xlstat.html*, accessed: 24 November 2012.

[10] U. R. Acharya, F. Molinari, S. V. Sree, S. Chattopadhyay, K-H. Ng, J. S. Suri. Automated diagnosis of epileptic EEG using entropies, *BIOMED. SIGNAL. PROCES.* 7 (4) (2012) 401-408.

[11] U. R. Acharya, S. V. Sree, A. P. C. Alvin, J. S. Suri. Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework, *EXPERT. Syst. Appl.* 39 (10) (2012) 9072-9078.