Wavelet based classification of epileptic seizures in EEG signals

Katerina D. Tzimourta, Loukas G. Astrakas
Medical Physics Laboratory
University of Ioannina
GR45110, Ioannina, Greece
e-mail: katerina.tzimourta@gmail.com, astrakas@uoi.gr

Markos G. Tsipouras, Nikolaos Giannakeas, Alexandros T. Tzallas
Dept. of Computer Engineering
Technological Educational Institute of Epirus
Kostakioi, GR47100, Arta, Greece
e-mail: giannakeas@teiep.gr, tzallas@teiep.gr, tsipouras@teiep.gr

Spyridon Konitsiotis
Dept. of Medicine
University of Ioannina
GR45110, Ioannina, Greece
e-mail: skonitso@uoi.gr

***Abstract*— Epilepsy is a chronic neurological disorder characterized by recurrent, sudden discharges of cerebral neurons, called seizures. Seizures are not always clearly defined and have extremely varied morphologies. Neurophysiologists are not always able to discriminate seizures, especially in long-term EEG datasets. Affecting 1% of the world's population with 1/3 of the epileptic patients not corresponding to anti-epileptic medication, epilepsy is constantly under the microscope and systems for automated detection of seizures are thoroughly examined. In this paper, a method for automated detection of epileptic activity is presented. The Discrete Wavelet Transform (DWT) is used to decompose the EEG recordings in several subbands and five features are extracted from the wavelet coefficients creating a set of features. The extracted feature vector is used to train a Support Vector Machine (SVM) classifier. Five classification problems are addressed, reaching high levels of overall accuracy ranging from 87% to 100%.**

Keywords*— *electroencephalogram (EEG); seizure; epilepsy; Discrete Wavelet Transform (DWT)

I. INTRODUCTION

Epilepsy is a chronic brain abnormality caused by a temporary interruption or disturbance of communication between the neurons, called "seizure". Recent studies have shown that almost 1% of the world's population suffers from epileptic seizures [1] while up to 5% of any population will experience a seizure at some point in their life [2]. It is also estimated that about 1/3 of the epileptic patients do not respond to anti-epileptic medication [1].

Individuals suffering from seizures, experience a variety of clinical features and symptoms due to this malfunction of communication. The symptoms are linked with the location of the epileptic zone and the extend of the affected brain areas. Depending on the range of participation of the brain areas during the seizure and the clinical manifestations that reflect these areas, seizures are categorized in two main types. Focal (or partial) seizures, which arise from a single brain area and// remain only to this area, and generalized seizures, which affect more than one region and in many cases almost the entire brain. These two types are also subdivided creating a further list of several seizure types.

Brain activity is monitored through the electroencephalogram (EEG), which provides important information and remains an indispensable tool for evaluating the abnormal brain activity and defining seizures. Based on the location of electrodes, EEG is divided into scalp EEG (sEEG), in which the electrodes are placed in the surface, and intracranial EEG (iEEG), in which the electrodes are placed invasively inside the brain. Experienced neurophysiologists review the EEG recording, which is usually recorded between two seizures (interictal period) and seldom during the seizure (ictal period). However, visual analysis may be prone to significant mistakes, owing to the complexity of the seizures. Seizures are not always clearly defined and have extremely varied morphologies [3] displayed in the EEG as spikes, sharp waves or spike-wave complexes [4]. These abnormalities often appear as artifacts misleading the physicians' assessment. Furthermore, reviewing long-term EEG recordings may be a time-consuming and eventually an ineffective process [5]. Consequently, computerized methods and seizure detection systems have become of great scientific interest and proven to be a reliable clinical tool.

Since the early 1970's scientists have developed a variety of methodologies and signal processing techniques for detecting epileptic activity. In the literature, there is an abundance of different methodologies, the majority of them concur with a two-stage procedure: feature extraction and classification. A method based on mimetic techniques is proposed in [6], wherein scientists compare spike attributes with values provided by the neurophysiologists. Additionally, authors in [7] proposed a method based on morphological analysis in which the basic idea is to separate spikes from background activity using morphological operators. Several methods [8,9] have been developed using a template matching approach

2372-9198/17 $31.00 © 2017 IEEE
DOI 10.1109/CBMS.2017.116

wherein spikes are visually selected from a set of test signals creating a template for identifying abnormal activity. A new technique based on parametric approaches is presenting in [5], where researchers make use of the nonstationarities of the EEG and use a time-varying autoregressive model to enhance spikes and separate them from background activity. Methods based on independent component analysis have also been proposed for artifact elimination and spike detection [10, 11] and separation of different spikes [12]. Furthermore, methods based on artificial neural networks [13, 14] have been developed; in the later either raw data or extracted features from EEG segments are fed into a neural network in order to detect spikes. Subasi [15] presented a method for epileptic seizure detection based on a modular neural network. Discrete wavelet transform (DWT) was applied in the EEG recordings and the extracted statistical features from the resulting subbands were fed into an optimized neural network. Data mining techniques have also been proposed to create automatic detection systems, wherein scientists in study [16] used algorithms based on decision trees and Bayesian classifiers whereas in studies [17-19] scientists achieved good classification results using a Support Vector Machine based classifier.

In this paper, a wavelet-based method wherein there is no need for a clear definition of spike morphology is presented. The method utilizes the Discrete Wavelet Transform in order to divide EEG recordings to subbands and extract several features. Subsequently, these features are given as an input to train a Support Vector Machines classifier. The method has been tested in 5 different classification problems and results are presented.

II. MATERIALS AND METHOD

The proposed method consists of three stages: wavelet analysis, feature extraction and classification. In the first stage, a 5-level wavelet-decomposition is applied in raw EEG recordings dividing each signal into several frequency subbands. In the next stage, 5 features are calculated from every subband creating a feature vector. Finally, all feature vectors are used to train a Support Vector Machine (SVM) classifier. In Fig. 1, a concise diagram of the proposed method is presented.

A. Data Selection

The EEG dataset used for this work is from the University of Bonn. The dataset consists of the sets A, B, C, D and E, each containing 100 single-channel EEG segments specified as Z, O, N, F and S respectively. These segments were selected and cut out from continuous multichannel EEG recordings [20]. The duration of each segment is 23.6 sec and the signal length is 4096 samples. The sampling frequency of the data was 173.61Hz and any artifacts were visually removed.

The recordings from the five subsets were acquired from five healthy volunteers and five patients suffering from seizures. Sets A and B consist of sEEG segments taken from five healthy volunteers. Specifically, set A contains signals marked with the letter Z, taken from five healthy volunteers who were relaxed in an awake state with eyes open. The second set of healthy sEEG (set B) includes signals marked with the letter O, which were taken from volunteers who were relaxed and awaked with eyes closed.

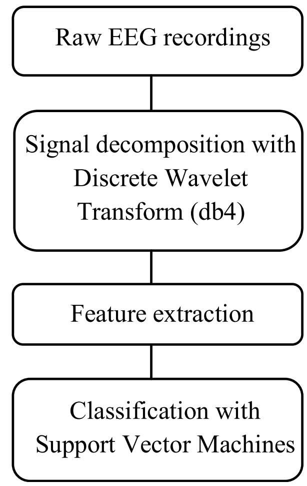

Fig. 1. A concise diagram of the proposed method

Sets C, D and E consist of iEEG segments that were collected intracranially from five epileptic patients during presurgical evaluation.

The epileptogenic zone was the hippocampal formation and the iEEG segments, which are originated from the hippocampal formation of the opposite hemisphere, are included in the set C and marked with the letter N. Set D encompasses iEEG segments derived from the epileptogenic zone during seizure-free intervals and the letter ‘F’ is used to describe these particular signals. Both sets C and D contain interictal activity, whilst ictal activity is demonstrated only in set E, which comprises the iEEG segments marked with the letter S. In this study, 500 segments were employed, and carried out five different classification problems.

B. Discrete Wavelet Transform

Discrete Wavelet Transform (or DWT) has gained significant ground in the feature extraction process and the analysis of nonstationary signals with transient occurrences like EEG recordings. The DWT is widely applied in numerous spike analysis studies, as it provides information in both frequency and time domain.

The basic idea underlying the DWT is that a signal can be expressed as a linear combination of a particular set of functions, obtained by shifting and dilating one single function called a mother wavelet [15]. The mother wavelet uses wavelet and scaling function as band-pass filters to decompose the signal into high and low frequency subband. The latter is further divided into high and low frequency subbands until the entire signal is decomposed. The coefficients of low frequencies are called Approximation coefficients (cA) and those of high frequencies are called Detail coefficients (cD).

Choosing the number of decomposition levels and the appropriate mother wavelet is essential in DWT. The number of decomposition levels is chosen based on the dominant frequency. The best mother wavelet function was selected after manual testing of different mother wavelets, mainly considering the Daubechies wavelets. This particular family of wavelet functions was chosen after visual inspection, based on the similarity between the epileptic EEG segments (set E) and the mother wavelets. In this study, a 5-level-decomposition transform is used in order to acquire valuable information and the family of Daubechies wavelets of order 4 (db4) is selected as the most appropriate mother wavelet to decompose the signal. Table I shows the corresponding frequencies to different levels of decomposition.

TABLE I. WAVELET DECOMPOSITIONS WITH THE CORRESPONDING FREQUENCY RANGES FOR SAMPLING FREQUENCY 173.6 HZ

Decomposed signal	Frequency range (Hz)
D1	43.4-86.8
D2	21.7-43.4
D3	10.8-21.7
D4	5.4-10.8
D5	2.7-5.4
A5	0-2.7

C. *Feature Extraction*

In the literature, a variety of statistical and non-statistical features have been suggested to represent the time-frequency distribution of EEG waveforms. In the present study, five features were computed in each decomposition level to create the complete feature vector. For clarity's sake, no more features were extracted from the signals to keep the feature set clear from noise and sustain the quality of predicting:

- Energy of the wavelet coefficients in each subband,
- Entropy of the wavelet coefficients in each subband based on the signal histogram and the Probability Density Function (PDF)
- Mean of the absolute values of the wavelet coefficients in each subband
- Standard deviation of the wavelet coefficients in each subband
- Variance of the wavelet coefficients in each subband.

All calculations implemented in Matlab environment. The entire set of features including the class attribute was given as an input feature vector to train an SVM.

D. *Classification*

Support Vector Machines (SVM) is a machine learning technique and it is widely used in biomedical application for binary classifications problems. It applies either hard-margin or soft-margin algorithm to separate linear or non-linear data, respectively. Non-linear instances that need to be classified are mapped to a high-dimension feature space. In this feature space instances are separated by a very clear gap, named hyperplane. The vectors that lie on the margin are called support vectors and they carry all the information about the classification problem. The idea of this technique is to locate an Optimal Separating Hyperplane, which maximizes the distance between the margin and the support vectors and minimizes the classification error. The transformation from the original space to the higher dimensional is performed by the kernel function. There are several kernel functions that have been used widely, such as linear, RBF, polynomial, and sigmoid kernels. In this study, RBF kernel function was used.

In machine learning the SVM algorithm can be optimized by choosing the best values for two parameters: the parameter C, which inserts a penalty for the misclassified instances and affects the margin boundaries, and the parameter gamma, which is originated from the RBF kernel function. For the purpose of optimizing the algorithm and therefore the classification results, we limited the range of the values and performed a grid search on these parameters using cross-validation. A separate search for various pairs of the parameters C and gamma was performed in each training fold and the pair with the best cross-validation accuracy was chosen.

III. RESULTS

To validate the efficiency of the proposed method, 5 different classification problems were created, as suggested in studies [5] and [21]. In the literature, the vast majority of the studies focus on the classification between normal and epileptic activity. However, this classification lacks medical interest from the neurophysiologists' point of view. Therefore, five different classification problems were addressed, including the classification in two, three and five classes, to comprise the most common discriminations in the medical field related to epilepsy. More specifically:

- In the first problem (ZONF-S) the entire dataset was used and the EEG recordings were divided in two classes: sets Z, O, N and F were merged to a single class forming the class "normal" and set S the class "seizure".
- In the second problem (Z-S) only the EEG recordings from the sets Z and S were used. Similar to the first problem, the EEG recordings from set Z formed the class "normal" whereas the ones from the set S formed the class "seizure".
- The third problem (ZO-NF-S) is a problem of three classes. EEG recordings from the sets Z and O were combined together forming the class "normal", sets N and F were combined forming the class "seizure-free" and finally set S formed the class "seizure".
- In the fourth problem (Z-F-S) only the EEG recordings from sets Z, F and S were used. Similar to the third problem, EEG recordings from set Z formed the class "normal", the ones from set F formed the class "seizure-free" whilst the set S

TABLE II. RESULTS FOR THE FIVE CLASSIFICATION PROBLEMS IN TERMS OF ACCURACY, SENSITIVITY, SPECIFICITY AND KAPPA STATISTIC. ALL WAVELET COEFFICIENTS ARE INCLUDED IN THE CLASSIFICATION.

Classification problem	***Accuracy* (%)**	***Sensitivity* (%)**	***Specificity* (%)**	***Kappa statistic***
ZONF-S	99.20	99.50	97.00	0.969
Z-S	99.50	99.00	100	0.990
ZO-NF-S	98.00	98.77	98.33	0.969
Z-F-S	98.33	98.33	99.17	0.975
Z-O-N-F-S	87.00	87.00	96.44	0.837

TABLE III. RESULTS FOR THE FIVE CLASSIFICATION PROBLEMS IN TERMS OF ACCURACY, SENSITIVITY, SPECIFICITY AND KAPPA STATISTIC. ONLY D3,D4,D5 AND A5 COEFFICIENTS ARE USED.

Classification problem	***Accuracy***	***Sensitivity***	***Specificity***	***Kappa statistic***
ZONF-S	98.40%	99.00%	96.00%	0.95
Z-S	100%	100%	100%	1
ZO-NF-S	97.40%	94.30%	98.65%	0.959
Z-F-S	97.33%	96.72%	98.66%	0.960
Z-O-N-F-S	82.80%	54.85%	95.37%	0.785

formed the class “seizure”. Most of the seizure detection methods were evaluated with these sets, excluding set O since it consists of normal EEG recordings taken from subjects with closed eyes, a fact that may induce alpha rhythms. Set N is also excluded, because even though recordings have been taken intracranially from the opposite side of the epileptic zone, a seizure in temporal lobe is regarded as a focal seizure [22].

- The fifth problem (Z-O-N-F-S) represents a problem of five classes, wherein each set is a class. This problem is more challenging to classify, because in many studies the obtained results are lower than the ones obtained from the binary problems.

In the experiments, the 10-fold cross-validation technique was employed. In the later, the entire dataset is divided in 10 equal folds. In the first division 9 of these folds are used to train the classifier and the rest 1 to test it. The procedure is repeated 10 times and each time a different fold is used as test set. In the final stage, the procedure is repeated on the entire dataset for a last time and the reported results are an aggregation of the results of the 10 models.

Sensitivity, specificity and accuracy were calculated for evaluating the prediction performance of the proposed method. These metrics are defined by the statistical measures True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN), which are related to the number of classified instances.

In the evaluation, Kappa statistic was considered, even though is rarely shown in automated seizure detection studies. Kappa statistic is based on the comparison between the correctly classified instances and those that classified correctly by chance due to uncertainty. The higher the obtained value of the kappa, the greater the expected chance agreement [23].

The obtained results for the five classification problems are shown in Table II. For the multiclass problems average values of accuracy, sensitivity and specificity were calculated. The best accuracy is achieved at the second classification problem (Z-S), while the lowest accuracy is 87% obtained through the fifth problem (Z-O-N-F-S).

IV. DISCUSSION AND CONCLUSIONS

In this work, a wavelet-based method for automated detection of epileptic activity is presented. The DWT was used and a 5-level decomposition was performed, decomposing the EEG recordings in 6 frequency subbands. Tests were performed with different types of mother wavelet and the fourth order Daubechies (db4) was the one that gave the best efficiency and selected for this method. Energy, entropy, mean of the absolute values of the wavelet coefficients, standard deviation and variance were calculated in each subband of interest, forming a vector of features. The feature vector was then utilized by a Support Vector classifier to create the best decision boundary between seizure and non-seizure EEG segments. Five different classification problems were conducted for the evaluation of the proposed method.

Table III shows the obtained results for the five classification problems described above. With regard to the 2-class problem, the accuracy is 99.2% and 99.5%, for the first (ZONF-S) and the second (Z-S) problem respectively. The vast majority of the studies focus on these classification problems. For the 3-class problem, the average accuracy is 98% and 98.33%, for the third (ZO-NF-S) and the fourth (Z-S) problem respectively. The most challenging classification is the one occurring in the fifth problem (Z-O-N-F-S). The accuracy of this problem is 87% and is the lowest value of the proposed method. This result is attributed to the misclassification between sets C and D and consequently to the single sensitivity of these classes, which is 83% and 70% respectively.

It is beyond doubt that the frequency range of interest, wherein any epileptic activity occurs, is [0.5-30] Hz. In the study [15], the author included only the calculated features from the low frequency bands (D3, D4, D5 and A5), which correspond to this spectrum. However, we conducted our method in two directions. In the first case, we included all the detail coefficients (D1-D5) and the approximation coefficient (A5). In the second case, we excluded from the classification the high frequency bands (D1 and D2), as suggested in [15]. In the first case we obtained better classification results for the

TABLE IV. A COMPARISON OF PERFORMANCES OF THE VARIOUS METHODS PROPOSED IN THE LITERATURE FOR THE DETECTION OF SEIZURES

Authors	*Method*	*Dataset*	*Acc*
Subasi [15]	Discrete Wavelet Transform (DWT), mixture of expert model	Z-S	95.00
Tzallas et al. [21]	Time Frequency (TF) analysis, Artificial Neural Network (ANN)	Z-S Z-F-S Z-O-N-F-S	100 100 89.00
Güler et al. [24]	Discrete Wavelet Transform (DWT), Adaptive neuro-fuzzy inference system	Z-O-N-F-S	98.68
This study	Discrete Wavelet Transform (DWT), Support Vector Machines (SVM)	ZONF-S Z-S ZO-NF-S Z-F-S Z-O-N-F-S	99.20 99.50 98.00 98.33 87.00

multiclass problem (3, 4 and 5). On the other hand, in the second case we obtained better classification results for the 2-class problem, which is the first and the second, and we even reached high accuracy (100%) for the same classification problem as described in [15]. Table IV presents a comparison between our method and other proposed methods in the literature. Only methods evaluated in the same dataset are included so that a comparison between the results is feasible.

Epilepsy is a persistent brain disorder which restricts the ailing individual. Scientific research has been focused on the detection of seizures, targeting to predict these occurrences, and so to improve the patient's quality of life. The proposed method is a tool, aiming to assist the neurophysiologists in the seizure detection task. In the future, it can be tested in long-term EEG datasets, which are closer to clinical EEG recordings, for evaluating its robustness. Alternative approaches, including different classifiers and feature combinations, must be examined to improve the method's statistical results.

ACKNOWLEDGMENT

This work has been partly funded by the Research Committee of Technological Educational Institute of Epirus.

REFERENCES

[1] Tsiouris, K. M., Tzallas, A. T., Markoula, S., Koutsouris, D., Konitsiotis, S., & Fotiadis, D. I. (2015). A Review of Automated Methodologies for the Detection of Epileptic Episodes Using Long-Term EEG Signals. *Handbook of Research on Trends in the Diagnosis and Treatment of Chronic Conditions*, 231.

[2] Bell, G. S., Neligan, A., & Sander, J. W. (2014). An unknown quantity—the worldwide prevalence of epilepsy. *Epilepsia*, *55*(7), 958-962.

[3] Gotman, J. (1999). Automatic detection of seizures and spikes. *Journal of Clinical Neurophysiology*, *16*(2), 130-140.

[4] Tzallas, A. T., Tsalikakis, D. G., Karvounis, E. C., Astrakas, L., Tzaphlidou, M., Tsipouras, M. G., & Konitsiotis, S. (2012). *Automated epileptic seizure detection methods: a review study*. INTECH Open Access Publisher.

[5] Tzallas, A. T., Oikonomou, V. P., & Fotiadis, D. I. (2006, August). Epileptic spike detection using a Kalman filter based approach. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual Int. Conference of the IEEE* (pp. 501-504). IEEE.

[6] Dingle, A. A., Jones, R. D., Carroll, G. J., & Fright, W. R. (1993). A multistage system to detect epileptiform activity in the EEG. *IEEE Transactions on Biomedical Engineering*, *40*(12), 1260-1268.

[7] Xu, G., Wang, J., Zhang, Q., Zhang, S., & Zhu, J. (2007). A spike detection method in EEG based on improved morphological filter. *Computers in biology and medicine*, *37*(11), 1647-1652.

[8] El-Gohary, M., McNames, J., & Elsas, S. (2008, August). User-guided interictal spike detection. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 821-824). IEEE.

[9] Vijayalakshmi, K., & Abhishek, A. M. (2010). Spike detection in epileptic patients EEG data using template matching technique. *Int. Journal of Computer Applications*, *2*(6), 5-8

[10] Iriarte, J., Urrestarazu, E., Valencia, M., Alegre, M., Malanda, A., Viteri, C., & Artieda, J. (2003). Independent component analysis as a tool to eliminate artifacts in EEG: a quantitative study. *Journal of clinical neurophysiology*, *20*(4), 249-257.

[11] Kobayashi, K., James, C. J., Nakahori, T., Akiyama, T., & Gotman, J. (1999). Isolation of epileptiform discharges from unaveraged EEG by independent component analysis. *Clinical neurophysiology*, *110*(10), 1755-1763.

[12] Urrestarazu, E., Iriarte, J., Artieda, J., Alegre, M., Valencia, M., & Viteri, C. (2006). Independent component analysis separates spikes of different origin in the EEG. *Journal of clinical neurophysiology*, *23*(1), 72-78.

[13] Tzallas, A. T., Tsipouras, M. G., & Fotiadis, D. I. (2007). Automatic seizure detection based on time-frequency analysis and artificial neural networks. *Computational Intelligence and Neuroscience*, *2007*.

[14] Zainuddin, Z., Huong, L. K., & Pauline, O. (2013). Reliable epileptic seizure detection using an improved wavelet neural network. *The Australasian medical journal*, *6*(5), 308-314.

[15] Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, *32*(4), 1084-1093.

[16] Valenti, P., Cazamajou, E., Scarpettini, M., Aizemberg, A., Silva, W., & Kochen, S. (2006). Automatic detection of interictal spikes using data mining models. *Journal of neuroscience methods*, *150*(1), 105-110.

[17] Meier, R., Dittrich, H., Schulze-Bonhage, A., & Aertsen, A. (2008). Detecting epileptic seizures in long-term human EEG: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns. *Journal of Clinical Neurophysiology*, *25*(3), 119-131.

[18] Temko, A., Thomas, E., Marnane, W., Lightbody, G., & Boylan, G. (2011). EEG-based neonatal seizure detection with support vector machines. *Clinical Neurophysiology*, *122*(3), 464-473.

[19] Acır, N., & Güzeliş, C. (2004). Automatic spike detection in EEG by a two-stage procedure based on support vector machines. *Computers in Biology and Medicine*, *34*(7), 561-575.

[20] Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, *64*(6), 061907.

[21] Tzallas, A. T., Tsipouras, M. G., & Fotiadis, D. I. (2009). Epileptic seizure detection in EEGs using time–frequency analysis. *IEEE trans. on information technology in biomedicine*, *13*(5), 703-710

[22] Liang, S. F., Wang, H. C., & Chang, W. L. (2010). Combination of EEG complexity and spectral analysis for epilepsy diagnosis and seizure detection. *EURASIP Journal on Advances in Signal Processing*, *2010*(1), 1.

[23] McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

[24] Güler, I., & Übeyli, E. D. (2005). Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. *Journal of neuroscience methods*, *148*(2), 113-1