35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

A study of Morphology-Based Wavelet Features and Multiple-Wavelet strategy for EEG Signal Classification: Results and Selected Statistical Analysis*

Jing Zhou[1], Robert J. Schalkoff[1], Brian C. Dean[2], Jonathan J. Halford[3]

Abstract— **Automatic detection and classification of Epileptiform transients is an open and important clinical issue. In this paper, we test 5 feature sets derived from a group of morphology-based wavelet features and compare the results with that of a Guler-suggested feature set. We also implement a multiple-mother-wavelet strategy and compare performance with the usual single-mother-wavelet strategy. The results indicate that both the derived features and the multiple-mother-wavelet strategy improved classifier performance, using a variety of performance measures. We assess the statistical significance of the performance improvement of the new feature sets/strategy. In most cases, the performance improvement is either significant or highly significant.**

I. INTRODUCTION

The electroencephalogram (EEG) is the most commonly performed clinical neurophysiology procedure. If epileptiform transients are detected in the EEG of a patient who is having seizure-like events, this suggests that the patient may be having epileptic seizures [1]. Epileptiform transients (ETs) are spikes or sharp waves with pointed peak and a duration of 20-70 ms and 70-200 ms, respectively. Sometimes ETs are followed by a slow wave [2], [3], [4].

Detecting ETs is important, yet difficult, because ETs have varied morphologies which are similar to some normal background activities (i.e. wicket spikes, exaggerated alpha activity, small sharp spikes, and sleep related activities) and artifacts (i.e. eye blink, eye movement, muscle and electrode artifacts). Detection of ETs is performed by visual inspection of EEG signals by electroencephalographers. EEGs are frequent misinterpreted by neurologists [5] so the development of automated systems for ET detection is clinically important.

Many approaches have been proposed for automatic detection and classification of ETs, including template matching, parametric methods, mimetic analysis, power spectral analysis and wavelet analysis [6]. Many researchers have demonstrated that the wavelet transform (WT) is a good feature detection strategy for ETs analysis.

The WT decomposes signals into multiple time/frequency resolutions. Particular characters, as non-stationary transient events, can be represented in various scales [7]. There are many strategies to extract features after application of the WT. Appropriate features can characterize particular traits of ETs and make the classification easier. In addition, the selection of an appropriate mother wavelet is important. A classic set of statistical features from wavelet coefficients using the Daubechies wavelet of order 2 (DB2) is widely used [8].

This paper extends on previous work [9] where we developed a group of morphology-based features based on wavelet coefficients and assembled five feature sets from them. In this study, we extend the number of feature sets tested, allow two mother wavelets to cooperate in classification, and measure whether differences in performance between different machine learning models are statistically significant. The classification method is k-nearest neighbor rule (k-NNR) with k=3 and a normalized distance measure. 10-fold cross-validation is used. The statistical significance of selected results are shown.

II. METHODOLOGY

A. Data Acquisition

The EEG data used in this research were 100 30-second EEG signal segments collected from 100 different patients. These EEG segments were selected because they contained difficult to interpret ETs as well as normal EEG events and artifacts which could be easily confused with ETs. The EEGs were recorded with a sampling rate at 256 Hz from 21 channels using the standard 10-20 electrode placement and were high-pass filtered (1 Hz), low-pass filtered (70 Hz) and notch filtered (60 Hz). 27 digitally-reformatted channels were selected for analysis: F7-T3, T3-T5, P4-O2, T4-T6, Fp1-F7, C3-P3, C4-P4, Fp2-F8, F8-T4, Fz-Cz, T5-O1, P3-O1, Fp1-F3, Cz-Pz, Fp2-F4, T6-O2, Fp2-A2, F4-C4, F3-C3, A1-avg, Fp1-A1, F7-avg, C3-avg, Fz-avg, T3-avg, P3-avg, Fp1-avg.

A group of 7 clinical neurophysiologists used a web-based EEG annotation system to mark small segments containing all of the paroxysmal events in these EEG segments, then 11 clinical neurophysiologists annotated each event as one of the following categories:

1) Artifact
2) Abnormal epileptiform
3) Normal electrocortical activity

Further details about how this annotation was performed are given elsewhere [10]. The 3 categories can be divided into

[1]J. Zhou and R. J. Schalkoff are with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC 29631, USA
[2]B. C. Dean is with the Department of Computer Science, Clemson University, Clemson, SC 29631, USA
[3]J. J. Halford is with the Department of Neurosciences, Medical University of South Carolina, Charleston, SC 29425, USA

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

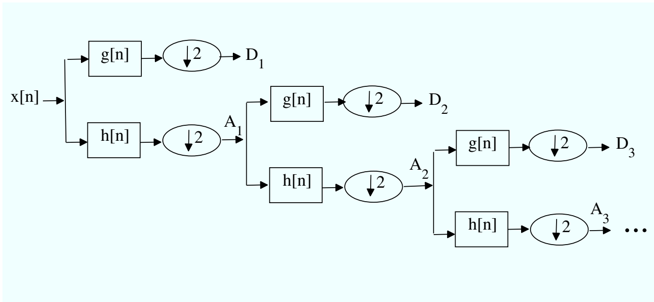

Fig. 1. wavelet decomposition tree

TABLE I

Subband	Frequency Range
D1	64Hz ∼ 128Hz
D2	32Hz ∼ 64Hz
D3	16Hz ∼ 32Hz
D4	8Hz ∼ 16Hz
A4	0Hz ∼ 8Hz

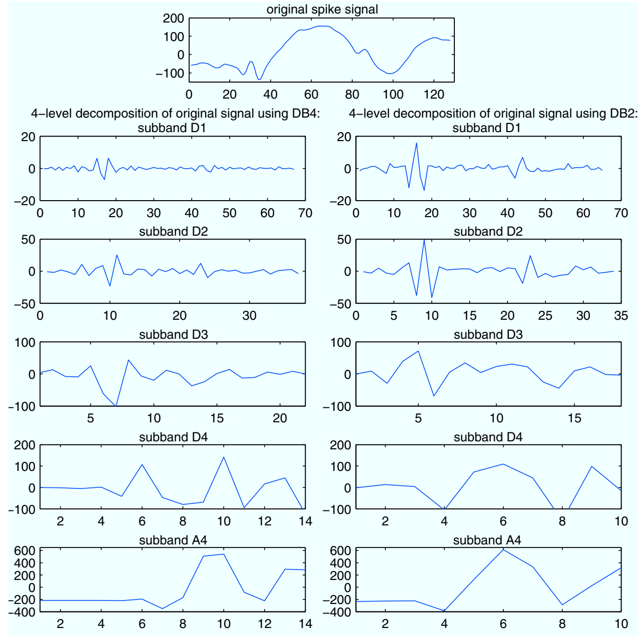

Fig. 2. Sample EEG Wavelet Decomposition Results Using DB4 and DB2

2 classes: The 'abnormal epileptiform' is ET class and the other two belong to non-ET class.

For this study, we used only the results annotated by the 7 neurophysiologists with the best inter-rater correlation. To derive a single annotation for each event, we considered the 7 neurophysiologists' opinions as votes. The category that received most votes from the 7 neurophysiologists was regarded as the consensus annotation of that event.

In total, we derived 83 ETs annotations and 2482 non-ETs annotations. A single feature vector was derived from each annotation.

B. Discrete Wavelet Transform

We windowed the bipolar montage with a 128-sample (500 ms) rectangular window, whose length is long enough to include the paroxysmal events. We used a 4-level wavelet decomposition. The 128-sample segment was decomposed into 5 subbands (four detail subbands D_1-D_4 and one approximation subband A_4). Table I shows the corresponding approximate frequency range of each subband.

In a previous study we found that the Daubechies wavelet of order 4 (DB4) and order 2 (DB2) mother wavelets are more useful for ET detection than certain other mother wavelets (DB 5, DB 20, bior1.3, bior 1.5) which have been used in ET detection research [9]. Another study has suggested that the DB4 mother wavelet is particularly useful for ET detection, since it obtains the highest correlation coefficients with the epileptic spike signal among the wavelet bases available in the Matlab toolbox [11]. So for this study, we have used only wavelet transforms with the DB2 and DB4 mother wavelet.

One focus of this study is to measure if combining the mother wavelets DB2 and DB4 in machine learning feature sets produces better results than using these mother wavelets alone. We have noticed that sometimes the response of mother wavelet DB2 to spikes is much larger than that of DB4 (as shown in Figure 2).

C. Feature Extraction

Guler suggested a feature set based on statistics over the WT coefficients (A 4-level decomposition using DB2 in Guler's research) [8]. This feature set uses maximum, minimum, mean and standard deviation of the wavelet coefficients in each subband. Therefore, for a 4-level decomposition (5 subbands), the feature vector dimension is 20.

The derivation of the wavelet-based features is an open problem, requiring considerable judgment, computational resources and trial-and-error [12]. Following Guler's methods and elaborating on them, in this study we have used the following features in each subband:

- **Feature #1:** the highest peak of the wavelet coefficients
- **Feature #2:** the lowest valley of the wavelet coefficients
- **Feature #3:** the mean of the peaks of the wavelet coefficients
- **Feature #4:** the mean of the valleys of the wavelet coefficients
- **Feature #5:** the variance of the peaks and the valleys of the wavelet coefficients
- **Feature #6:** the variance of the peaks of the wavelet coefficients
- **Feature #7:** the variance of the valleys of the wavelet coefficients

In order to achieve high performance with relatively low vector dimension, we assembled 5 combinations and their feature choices and dimensions are shown in Table II. When a single feature was used for classification, Feature #4 yields the lowest sensitivity and specificity result and Feature #3 yields the second lowest sensitivity and specificity. To reduce feature vector dimension with the lowest decrease in performance, we discarded Feature #4 in Set #4 and then Feature #3 in Set #5.

TABLE II

FEATURE CHOICES AND DIMENSIONS OF NEW FEATURE SETS

Set \ Selected Features	#1	#2	#3	#4	#5	#6	#7	DIM
Set #1	×	×	×	×	×			25
Set #2	×	×	×	×		×	×	30
Set #3	×	×	×	×	×	×	×	35
Set #4	×	×	×		×			20
Set #5	×	×	×					15

D. Employing Multiple Mother Wavelets

In Figure 2, in decompositions D_1 and D_2, both DB2 and DB4 present the ET event at corresponding x-coordinates where the ET occurs in the time domain (starting at x-coordinate 30 in 'original spike signal' plot). Observe the peak value of DB2 is twice of that of DB4; this is an example of how a feature can be more evident using a WT with one mother wavelet than another. To improve the classifier performance, we combined features using both DB4 and DB2 into one vector for classification. By using this dual mother wavelet strategy, the vector dimension is doubled.

E. Scalp Location Features

Experts have indicated that the ETs usually occur in the temporal lobe, suggesting the locations of the scalp electrodes in which the signal is recorded could be useful additional features. We employed a 2D-coordinate system (the 10-20 electrode placement system) and used the X, Y coordinates of the midpoint of each bipolar pair as spatial features. Our previous research shows that attaching the 2 spatial features to wavelet feature vectors help improve the classification performance in some cases [9].

III. RESULTS AND DISCUSSION

A. Feature Set Performance Comparison

To test the classification ability of different features and different mother wavelets, 18 datasets are built using the 6 feature sets in Section II-C and 3 choices of mother wavelets: DB2, DB4 and DB4+DB2.

Within EEG recordings, non-ET events occur more frequently than ET events. In our dataset, there are a total of 83 ET feature vectors and 2482 non-ET feature vectors derived from the annotations and the ratio of ET/non-ET is 1:30. The annotations indicated that all 100 patients provided non-ET events while only 31 patients provided ET events. To avoid prejudice in classification, we chose to balance the training set (H); we kept the ET/non-ET ratio as 1:30 in test set (S_T) to imitate the unbalanced practical situation.

Within a single trial, 80 ET vectors and 2400 non-ET vectors were randomly selected from the available data. We estimated classification performance using k-NNR (k=3) with 10-fold cross-validation. We chose the non-parametric classification method, i.e. k-NNR, since it requires no assumptions about the distribution of the data or classifier parameters. Therefore, the results reflect the properties of the feature data, not the classifier. Ordinary k-NNR measures the Euclidean distance between 2 vectors. In practice, however, the entry values in one vector could be different by several orders of magnitude due to the distribution and the range of the data they represent. To normalize the entry values in a single vector, we computed the distance as following

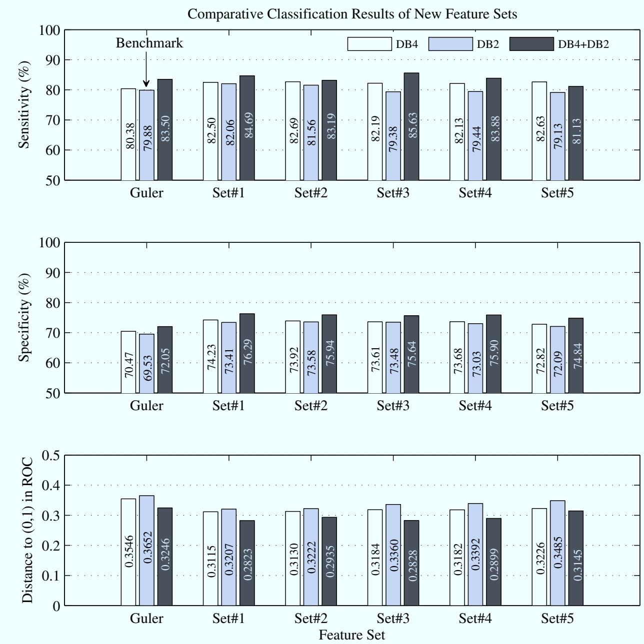

Fig. 3. k-NNR (k=3) comparative classification results of new feature sets

$$d(\vec{v}_1,\vec{v}_2) = \sqrt{(\vec{v}_1-\vec{v}_2)^T\Sigma^{-1}(\vec{v}_1-\vec{v}_2)} \tag{1}$$

where Σ is the diagonal of the covariance matrix of the randomly selected single-trial dataset. In 10-fold cross-validation, the available dataset, D, is partitioned into 10 mutually exclusive subsets of equal size by random sampling. It is trained on $D\backslash D_t$ and tested on D_t with $t = 1,2,...10$. Predicted performance estimates for various measures (sensitivity, specificity, etc.) are thus obtained. 10-fold cross-validation is a recommended performance prediction method with less bias and variance [13].

The test performance is assessed by sensitivity and specificity, defined as:

Sensitivity = TP/(TP + FN), capacity to recognize positive events;

Specificity = TN/(TN + FP), capacity to recognize negative activity.

To achieve a single numerical measure of performance combining sensitivity and specificity, we also measured the distance between our classifier result and the coordinate (0,1) in the Receiver Operating Characteristic (ROC) space:

$$distance = \sqrt{(1-sensitivity)^2+(1-specificity)^2} \tag{2}$$

In an ideal case with 100% TP and 0% FP, this distance to (0,1) is 0.

Considering the uncertainty and variation of the random selection of the data in a single trial, 20 trials were done for each feature set/wavelet choice. The average performance of various feature sets are listed in Figure 3.

B. Statistic Significance of Detection Improvement: One-Tailed t-Test

To assess statistical significance, a one tailed t-test is used [14]. We assess whether the mean performance results from two feature sets are statistically different.

First, we test if the mean of sensitivities/specificities of a feature set/wavelet combination, is higher than that of benchmark Guler-suggested feature set/wavelet choice, with significance levels α (weakly significant: α=0.1; significant level: α=0.05; highly significant: α=0.01) and 20 observations. The hypotheses are:

$$H_0 : \mu_g = \mu,$$
$$H_1 : \mu_g < \mu,$$

μ (μ_g) is the mean of the observed performance of a feature set/wavelet choice (Guler-suggested feature set/wavelet choice). The standard deviation of the observations in each set are assumed unequal and unknown. The critical region to reject H_0 is $t < -t_\alpha$.

To test if the distance-to-(0,1) decreased significantly, we revise the hypotheses:

$$H_0 : \mu_g = \mu,$$
$$H_1 : \mu_g > \mu,$$

where μ (μ_g) is the mean of the observed distance (mean of the benchmark Guler-suggested feature set distance). The critical region to reject H_0 is $t > t_\alpha$.

Table III and Table IV shows the highest level at which H_0 can be rejected in these tests.

From Table III, comparing the results based upon the same wavelet choice, we observed: (1) For sensitivity, H_0 is rejected at a significant level ($\alpha = 0.05$) in 2 cases (DB4 and DB2) of Set#1 & Set#2, 1 case (DB4+DB2) of Set#3 and 1 case (DB4) of Set#5; H_0 is rejected at a weakly significant level ($\alpha = 0.1$) in 1 cases (DB4) of Set#3 & Set#4; (2) For specificity, H_0 is rejected at a highly significant level ($\alpha = 0.01$) in all cases, indicating the improvement in specificity is both universal and profound; (3) Influenced by specificity, H_0 is also rejected at a highly significant level in distance-to-(0,1) except 2 cases (DB2 and DB4+DB2) of Set#5, where the H_0 is still rejected at a significant level. Comparing the results with benchmark feature set/wavelet choice, we observed: (1) By simply using DB4 instead of DB2,, the sensitivity can be significantly improved and the specificity and distance-to-(0,1) can be highly significantly improved; (2) By employing the dual-wavelet strategy, sensitivities are highly increased; The exception is Set#5, in which dual-wavelet degrades the sensitivity.

Table IV compares the performance of single-wavelet with dual-wavelet within feature set. In Table IV, we observed H_0 is rejected at a highly significant level ($\alpha = 0.01$) in most cases, especially in half of the sensitivities, indicating dual-wavelet is a powerful strategy, since Table III has shown that it is difficult to make improvement in sensitivity. Only

[1]Benchmark tests use Guler's feature set and mother wavelet DB2

TABLE III

THE HIGHEST LEVEL AT WHICH H_0 CAN BE REJECTED WITH DIFFERENT FEATURE SET/WAVELET CHOICES

Comparison between μ_g with Different Feature Sets μ with Same Wavelet Choice						
μ_g \ μ	Set1	Set2	Set3	Set4	Set5	mother wavelet
			Sensitivity			
Guler	0.05	0.05	0.1	0.1	0.05	DB4
	0.05	0.05	fail	fail	fail	DB2
	fail	fail	0.05	fail	fail	DB4+DB2
			Specificity			
	0.01	0.01	0.01	0.01	0.01	DB4
	0.01	0.01	0.01	0.01	0.01	DB2
	0.01	0.01	0.01	0.01	0.01	DB4+DB2
			Distance to (0,1)			
	0.01	0.01	0.01	0.01	0.01	DB4
	0.01	0.01	0.01	0.01	0.05	DB2
	0.01	0.01	0.01	0.01	0.05	DB4+DB2
Comparison between Benchmark μ_g and Different Feature-Set + Wavelet-Choice μ						
			Sensitivity			
Benchmark [1]	0.05	0.05	0.05	0.05	0.01	DB4
	0.01	0.01	0.01	0.01	0.1	DB4+DB2
			Specificity			
	0.01	0.01	0.01	0.01	0.01	DB4
	0.01	0.01	0.01	0.01	0.01	DB4+DB2
			Distance to (0,1)			
	0.01	0.01	0.01	0.01	0.01	DB4
	0.01	0.01	0.01	0.01	0.01	DB4+DB2

TABLE IV

THE HIGHEST LEVEL AT WHICH H_0 CAN BE REJECTED OF SINGLE VS. DOUBLE MOTHER WAVELETS WITHIN FEATURE SET

Comparison between Dual-Wavelet μ and Single-Wavelet within Feature Set μ_g						
μ_g \ μ	Guler	Set1	Set2 DB4+DB2	Set3	Set4	Set5
			Sensitivity			
DB4	0.01	0.05	fail	0.01	0.05	fail
DB2	0.01	0.01	0.1	0.01	0.01	0.05
			Specificity			
DB4	0.01	0.01	0.01	0.01	0.01	0.01
DB2	0.01	0.01	0.01	0.01	0.01	0.01
			Distance to (0,1)			
DB4	0.01	0.01	0.01	0.01	0.01	0.1
DB2	0.01	0.01	0.01	0.01	0.01	0.01

2 cases failed to reject H_0 at a weakly significant level. If comparing to the benchmark feature set using single-wavelet, all cases of feature sets using dual-wavelet reject H_0 at a highly significant level except the sensitivities of Set#5.

IV. CONCLUSIONS

In this paper we have experimented with new strategies for improving the performance of machine classifiers used to detect segments of EEGs containing ETs. From a group of 7 potential wavelet features, we derived and tested 5 distinct feature sets. We assessed classifier performance by combining features derived from two different mother wavelets. We also ran tests to determine if any improvement in model performance was statistically significant. Our results showed that our new wavelet features improve the classification ability (in the best case, +5.75% in sensitivity and +6.76% in specificity at highly significant level: α=0.01) and that the use of two dual-mother-wavelets in a classifier may be better than using a single-mother-wavelet under the condition that both wavelets are able to detect the events of interest. We think that the dual mother wavelet strategy improves performance of the machine learning classifiers because it may be difficult to represent these various ET signal patterns using feature sets based on a single mother wavelet.

REFERENCES

[1] N. B. Fountain and J. M. Freeman, EEG is an essential clinical tool: pro and con, Epilepsia, vol. 47 (Suppl 1), pp. 23-25, 2006.

[2] B. F. Westmoreland, Epileptiform electroencephalographic patterns, Mayo Clin. Proc., vol. 71, pp. 501-511, 1996.

[3] IFSECN, A glossary of terms commonly used by clinical electroencephalographers, Electroencephalogr. Clin. Neurophysiol., vol. 37, pp. 538-548, 1974.

[4] S. Noachtar, C. Binnie, J. Ebersole, F. Mauguiere, A. Sakamoto and B. Westmoreland, A glossary of terms most commonly used by clinical electroencephalographers and proposal for the report form for the EEG findings, in: G. Deuschl, A. Eisen (Eds.), Recommendation for the Practice of Clinical Neurophysiology: Guidelines of the International Federation of Clinical Physiology (EEG Suppl. 52), 1999.

[5] S. R. Benbadis, Errors in EEGs and the misdiagnosis of epilepsy: importance, causes, consequences, and proposed remedies, Epilepsy & Behavior, vol. 11, pp. 257-262, 2007.

[6] J. J. Halford, Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized EEG interpretation, Clinical Neurophysiology, vol. 120, pp. 1909-1915, 2009.

[7] S. G. Mallat, A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, Pattern Analysis and Machine Intelligence, vol. 11, pp. 674-693, July 1989.

[8] I. Guler and E. D. Ubeyli, Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients, Journal of Neuroscience Methods, vol. 148, pp. 113-121, Apr. 2005.

[9] J. Zhou, R. J. Schalkoff, B. C. Dean and J. J. Halford, Morphology-Based Wavelet Features and Multiple Mother Wavelet Strategy for Spike Classification in EEG Signals, Engineering in Medicine and Biology Society (EMBS), 2012 Annual International Conference of the IEEE, pp. 3959-3962, 2012.

[10] J. J. Halford, R. J. Schalkoff, J. Zhou, S. R. Benbadis, W. O. Tatum, R. P. Turner, S. R. Sinha, N. B. Fountain, A. Arain, P. B. Pritchard, E. Kutluay, G. Martz, J. C. Edwards, C. Waters and B. C. Dean, Standardized Database Development for EEG Epileptiform Transient Detection: EEGnet Scoring System and Machine Learning Analysis, Journal of Neuroscience Methods, vol. 212(2), pp. 308-316, Jan. 2013.

[11] K. P. Indiradevi, E. Elias, P. S. Sathidevi, S. D. Nayak and K. Radhakrishnan, A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram, Computers in Biology and Medicine, vol. 38, pp. 805-816, Apr. 2008.

[12] R. Battiti, Using Mutual Information for Selecting Features in Supervised Neural Net Learning, Neural Networks, vol. 5, pp. 537-550, July 1994.

[13] R. Kohavi, A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, International Joint Conference on Artificial Intelligence(IJCAI), 1995.

[14] R. E. Walpole and R. H. Myers, Probability and statistics for engineers and scientists, Macmillan Publishing Co., Inc, 1978.