34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

# Morphology-Based Wavelet Features and Multiple Mother Wavelet Strategy for Spike Classification in EEG Signals*

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

***Abstract*— New wavelet-derived features and strategies that can improve autonomous EEG classifier performance are presented. Various feature sets based on the morphological structure of wavelet subband coefficients are derived and evaluated. The performance of these new feature sets is superior to Guler's classic features in both sensitivity and specificity. In addition, the use of (scalp electrode) spatial information is also shown to improve EEG classification. Finally, a new strategy based upon concurrent use of several mother wavelets is shown to result in increased sensitivity and specificity. Various attempts at reducing feature vector dimension are shown. A non-parametric method, k-NNR, is implemented for classification and 10-fold cross-validation is used for assessment.**

## I. INTRODUCTION

Abnormal EEG activity can be separated into epileptiform and non-epileptiform activity. Between seizures, the EEG of a patient with epilepsy may be characterized by occasional epileptiform transients (ETs) which consist of spikes or sharp waves which can last for 20-200ms. Occurrence of ETs in an EEG recording indicates that a patient probably has epilepsy. ETs are difficult to detect because they have a wide range of morphologies and are similar to some normal background activities or artifacts [4].

Many approaches for machine classification of ETs have been proposed and are summarized in a recent review by Halford [4]. They include template matching, parametric methods, mimetic analysis, power spectral analysis and wavelet analysis. The wavelet transform (WT) has become popular for this task since the WT appears to represent ETs well.

There are many strategies to extract features after application of the WT. In addition, the selection of an appropriate (or perhaps optimal) mother wavelet is an open problem. Guler suggested a set of statistical features from wavelet coefficients using the Daubechies wavelet of order 2 (DB2) [1]. Indiradevi suggested that wavelet DB4 obtains the highest correlation coefficients with the epileptic spike signal among the wavelet bases available in the Matlab toolbox [3]. Other mother wavelets, e.g., DB5, DB20, bior1.3 and bior1.5 are also suggested [5] [6].

In this paper, we implement five new wavelet-derived feature sets and compare them with Guler's classic wavelet feature set. We also explore the effects of scalp location features, high frequency subband features, dataset size and the cooperative use of two mother wavelets.

## II. METHODOLOGY

### A. Wavelet-derived features

The wavelet function scales the raw EEG signal for each decomposition level and halves its bandwidth, yielding a detail subband, while the scaling function yields an approximation subband. Further processing is needed to transform the raw WT-derived coefficients from these subbands into feasible features for machine classification. Guler [1] suggested a feature set based on statistics over the coefficients. This feature set uses maximum, minimum, mean and standard deviation of the wavelet coefficients in each of the 5 subbands derived from 4-level decomposition by mother wavelet DB2 and thus has a reasonable feature vector dimension of $d = 20$.

In our work, we used a 4-level wavelet decomposition. The raw EEG signal segments were decomposed into 5 subbands (four detail subbands $D_1$-$D_4$ and one approximation subband $A_4$). Table I shows the corresponding frequency range for each subband. Features are then extracted from these 5 subbands.

In the WT subband data, ETs appear as local peaks or valleys. This is illustrated in Figure 1. Therefore, we considered the following feature sets:

- **Set #1:** The highest peak, the lowest valley, the mean of the peaks, the mean of the valleys, and the variance of the peaks and the valleys of the wavelet coefficients. This yields 5 features in each of 5 subbands, i.e., a $d = 25$ dimensional feature vector.
- **Set #2:** The highest peak, the lowest valley, the mean of the peaks, the mean of the valleys, the variance of the peaks and the variance of the valleys of the wavelet coefficients. This yields 6 features in each of 5 subbands, i.e. a $d = 30$ dimensional feature vector.
- **Set #3:** The highest peak, the lowest valley, the mean of the peaks, the mean of the valleys, the variance of the peaks and the valleys, the variance of the peaks and the variance of the valleys of the wavelet coefficients. This yields 7 features in each of 5 subbands, i.e. a $d = 35$ dimensional feature vector.
- **Set #4:** The highest peak, the lowest valley, the mean of the peaks, and the variance of the peaks and the valleys of the wavelet coefficients. This yields 4 features in each of 5 subbands, i.e. a $d = 20$ dimensional feature vector.
- **Set #5:** The highest peak, the lowest valley, and the variance of the peaks and the valleys of the wavelet co-

[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

978-1-4577-1787-1/12/$26.00 ©2012 IEEE

efficients. This yields 3 features in each of 5 subbands, i.e. a $d = 15$ dimensional feature vector.

### B. Employing Multiple Mother Wavelets

Previous studies suggest that the Daubechies wavelet of order 4 (DB4) and order 2 (DB2) are effective mother wavelets for detecting ETs. This assertion is also confirmed in our study comparing DB4 and DB2 with 4 other types of mother wavelet (DB 5, DB 20, bior1.3, bior 1.5). However, in some cases DB2 detected the spikes while DB4 did not (or the inverse). To improve the classifier performance, we combined DB4 and DB2 features into one vector for classification.

### C. Using High Frequency Coefficients and Scalp Location Features

As noted, ETs are spikes or sharp waves lasting for 20∼200ms (5∼50Hz). While Table I indicates subband D1 information is above this effective frequency range, morphological differences can still be observed in D1 at the point where the ET spike occurs (See Figure 1). Based on this observation, the features of this 'high frequency subband' are found to be useful.

Experts have also indicated that the ETs usually occur in the temporal lobe. This suggests that locations of the scalp electrodes in which the ET is detected could be used as features. We employ a 2D-coordinate system by means of the 10-20 electrode placement system and use the X-Y coordinates of the midpoint of each electrode pair as spatial features. Our previous research shows that attaching the spatial information to wavelet feature vectors improves the performance in classification.

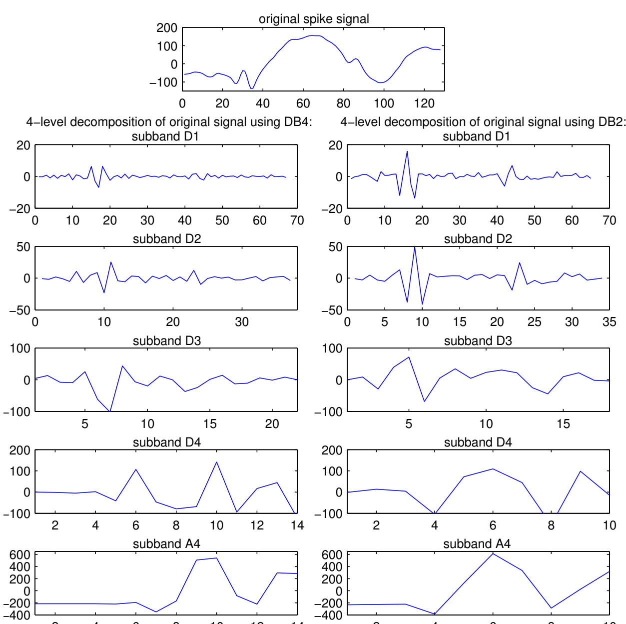


Fig. 1. Sample EEG Wavelet Decomposition Results Using DB4 and DB2

### D. Data Acquisition

The EEG data used in this study were 30 second segments from 100 patients[1]. The signals were recorded from 21 electrode channels, using the standard 10-20 electrode placement system. The sampling rate was 256Hz. Seven experts first annotated all of the paroxysmal events in the EEG segments and then went back and categorized each paroxysmal event as either an artifact, an ET, or an normal EEG event. In total, 83 ET and 2482 non-ET events (either artifacts or normal electrocortical activity) were annotated. We use a 128-length rectangular window to truncate the raw data and apply the aforementioned 4-level wavelet decomposition(s) to get the wavelet subband coefficients for each choice of mother wavelet used. Features described in Section II-A are then extracted from the subband data. For each choice of features, a single feature vector was derived from each annotation.

The dimension of the feature vector depends on the number of extracted features. The numbers of features in each set are indicated in Section II-A. If using the dual mother wavelet strategy, this dimension will be doubled. Also (x,y) scalp locations are added to each feature vector.

## III. Results

All results in Sections III and IV are based upon averaging the classification performance using 20 random selections of training and test sets ($H$ and $S_T$) from the available data. Within each selection (trial), we estimate classifier performance using k-NNR with 10-fold cross-validation [2]. This non-parametric classification design requires no assumptions about the distribution or classifier parameters. The diagonal of the covariance matrix of $H$ is used to normalize the distance measure used in k-NNR. Annotated non-ET events occur much more frequently than annotated ET events in practice and in the given dataset. The ratio in the latter is ET/non-ET=1:30. Thus, we chose $H$ to be balanced, and tested the each classifier with an unbalanced $S_T$. Extensive analysis of the statistical significance of these results using both hypothesis testing and regression analysis is currently in progress.

### A. Feature Set Performance Comparison

The average results of various feature sets are listed in Table II. To achieve a single numerical measure of performance combining sensitivity and specificity, we use the distance between the result of a feature set to the coordinate (0,1) in the Receiver Operating Characteristic (ROC).

Employing Guler's feature set only, the resulting classifier sensitivity is 79.88% and the specificity is 69.53%. By combining the DB4 and DB2 features, the sensitivity improves to 83.50% and the specificity improves to 72.05%. Using amended feature Set #3, the sensitivity can be pushed to 85.63% with a 75.64% specificity. Using amended feature Set #1, the specificity increases to 76.29% with a 84.69% sensitivity.

[1] http://eegnet.clemson.edu/

TABLE I

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

### B. Max vs All

In Guler's method, 5 subbands result from the 4-level wavelet decomposition and 4 features are extracted from each subband. Thus, there are 22 features in one vector (20 wavelet-derived features and 2 spatial features which are X and Y scalp coordinates). The vector size increases to 42 while using the dual mother wavelet cooperation strategy (20 wavelet derived by DB4 and DB2 respectively plus 2 X-Y coordinates). To reduce the computational complexity and to increase the efficiency, we tried using only the maximum of the coefficients in each subband (since a spike usually creates higher coefficients where it occurs than the background signal). This results in a $d = 7$ dimensional feature vector. The average results are listed in Table III by mother wavelet. Note the performance of $d = 22$ dimensional feature vectors is superior to the $d = 7$ case in sensitivity, specificity and distance to (0,1) except when using the mother wavelet of DB20 and bior1.3.

### C. Effects of Electrode Pair Scalp Location Features

Incorporating spatial information in the feature vector generally helps to improve classification performance. The average results are listed in Table IV by mother wavelet. Using the ROC distance to (0,1), we observe that the results with spatial information are closer to point (0,1), except in the case of mother wavelet of DB5. However, comparing the resulting changes of sensitivity and specificity shows a more complex situation. The sensitivity improves while the specificity decreases when using the mother wavelet of DB20. The specificity improves while the sensitivity decreases when using the mother wavelet of DB2, DB4 and DB5. Both the sensitivity and the specificity improve when using the mother wavelet of bior1.3, bior1.5 and multiple-wavelet combined feature sets (DB4+DB2 set and DB4+DB2+bior1.5 set). By evaluating the sensitivity only, the best case is that the sensitivity is improved by 4.31% after adding location features when using DB20 feature set. By evaluating the specificity only, the best case is that the specificity is improved by 2.04% after adding location features when using DB2 feature set.

### D. Effects of the Dataset Size

The dataset used in this study provides a limited number of spike events (83 samples total). It is suggested that increasing the size of the dataset would achieve better results. The effect of the dataset size was studied. Three subsets of the available data were used:

1) **2480-set:** 80 ET and 2400 non-ET samples.
2) **1860-set:** 60 ET and 1800 non-ET samples.

TABLE II
K-NNR(K=3) COMPARATIVE CLASSIFICATION RESULTS OF NEW FEATURE SETS

| | | Sensitivity | Specificity | Distance to (0,1) |
|---|---|---|---|---|
| Guler | DB4 | 80.38% | 70.47% | 0.3546 |
| | DB2 | 79.88% | 69.53% | 0.3652 |
| | DB4+DB2 | 83.50% | 72.05% | 0.3246 |
| Set #1 | DB4 | 82.50% | 74.23% | 0.3115 |
| | DB2 | 82.06% | 73.41% | 0.3207 |
| | DB4+DB2 | 84.69% | 76.29% | 0.2823 |
| Set #2 | DB4 | 82.69% | 73.92% | 0.3130 |
| | DB2 | 81.56% | 73.58% | 0.3222 |
| | DB4+DB2 | 83.19% | 75.94% | 0.2935 |
| Set #3 | DB4 | 82.19% | 73.61% | 0.3184 |
| | DB2 | 79.38% | 73.48% | 0.3360 |
| | DB4+DB2 | 85.63% | 75.64% | 0.2828 |
| Set #4 | DB4 | 82.13% | 73.68% | 0.3182 |
| | DB2 | 79.44% | 73.03% | 0.3392 |
| | DB4+DB2 | 83.88% | 75.90% | 0.2899 |
| Set #5 | DB4 | 82.63% | 72.82% | 0.3226 |
| | DB2 | 79.13% | 72.09% | 0.3485 |
| | DB4+DB2 | 81.13% | 74.84% | 0.3145 |

TABLE III
K-NNR(K=3) CLASSIFICATION RESULTS OF USING OVERALL FEATURES BASED ON GULER'S FEATURES VS. USING ONLY MAXIMA

| | | Sensitivity | Specificity | Distance to (0,1) |
|---|---|---|---|---|
| DB2 | All | 79.88% | 69.53% | 0.3652 |
| | Max | 75.00% | 68.25% | 0.4041 |
| DB4 | All | 80.38% | 70.47% | 0.3546 |
| | Max | 78.00% | 68.14% | 0.3872 |
| DB5 | All | 73.00% | 69.98% | 0.4037 |
| | Max | 72.94% | 69.71% | 0.4062 |
| DB20 | All | 76.81% | 68.83% | 0.3885 |
| | Max | 77.13% | 67.51% | 0.3974 |
| bior1.3 | All | 76.69% | 67.38% | 0.4009 |
| | Max | 75.13% | 68.55% | 0.4009 |
| bior1.5 | All | 77.81% | 69.63% | 0.3761 |
| | Max | 72.00% | 66.79% | 0.4344 |
| DB4+DB2 | All | 83.50% | 72.05% | 0.3246 |
| | Max | 77.56% | 71.26% | 0.3646 |
| DB4+DB2 +bior1.5 | All | 82.19% | 72.13% | 0.3308 |
| | Max | 76.63% | 71.36% | 0.3696 |

3) **1240-set:** 40 ET and 1200 non-ET samples.

The three subsets are formed on the principle that the ratio of ET/non-ET is 1:30, as in the original dataset. The average classification results are listed in Table V.

Table V shows that performance (measured as distance to (0,1) in the ROC) increases with increasing in the size of the dataset. The specificity is definitely improved as the dataset gets larger. However, the sensitivities do not monotonically increase in all cases. The exceptions occur when the features are derived using DB5, DB20 or DB4+DB2, where the sensitivity of 1860-set is the highest in each case respectively. By evaluating the sensitivity only, the best case is that the sensitivity is improved by 6.44% from the 1240-set to the 2480-set when using DB4+DB2+bior1.5 feature set. By evaluating the specificity only, the best case is that the specificity is improved by 4.73% from the 1240-set to the 2480-set when using DB4+DB2 feature set.

TABLE IV

K-NNR(K=3) CLASSIFICATION RESULTS WITH/WITHOUT LOCATION FEATURES BASED ON GULER'S FEATURES

| | | Sensitivity | Specificity | Distance to (0,1) |
|---|---|---|---|---|
| DB2 | with XY | 79.88% | 69.53% | 0.3652 |
| | no XY | 81.31% | 67.49% | 0.3750 |
| DB4 | with XY | 80.38% | 70.47% | 0.3546 |
| | no XY | 81.06% | 69.53% | 0.3588 |
| DB5 | with XY | 73.00% | 69.98% | 0.4037 |
| | no XY | 77.38% | 69.49% | 0.3799 |
| DB20 | with XY | 76.81% | 68.83% | 0.3885 |
| | no XY | 72.50% | 69.36% | 0.4117 |
| bior1.3 | with XY | 76.69% | 67.38% | 0.4009 |
| | no XY | 76.25% | 67.01% | 0.4065 |
| bior1.5 | with XY | 77.81% | 69.63% | 0.3761 |
| | no XY | 75.50% | 68.40% | 0.3998 |
| DB4+DB2 | with XY | 83.50% | 72.05% | 0.3246 |
| | no XY | 82.31% | 70.54% | 0.3437 |
| DB4+DB2 | with XY | 82.19% | 72.13% | 0.3308 |
| +bior1.5 | no XY | 78.63% | 70.68% | 0.3628 |

TABLE V

K-NNR(K=3) CLASSIFICATION RESULTS WITH DATASETS OF DIFFERENT SIZE

| | DS size | Sensitivity | Specificity | Distance to (0,1) |
|---|---|---|---|---|
| DB2 | 1240 | 74.63% | 65.69% | 0.4268 |
| | 1860 | 78.25% | 67.37% | 0.3922 |
| | 2480 | 79.88% | 69.53% | 0.3652 |
| DB4 | 1240 | 77.63% | 65.78% | 0.4088 |
| | 1860 | 78.75% | 68.16% | 0.3828 |
| | 2480 | 80.38% | 70.47% | 0.3546 |
| DB5 | 1240 | 71.50% | 65.71% | 0.4459 |
| | 1860 | 73.58% | 68.08% | 0.4144 |
| | 2480 | 73.00% | 69.98% | 0.4037 |
| DB20 | 1240 | 73.38% | 63.65% | 0.4506 |
| | 1860 | 77.58% | 66.83% | 0.4004 |
| | 2480 | 76.81% | 68.83% | 0.3885 |
| bior1.3 | 1240 | 71.25% | 63.87% | 0.4617 |
| | 1860 | 74.67% | 66.44% | 0.4205 |
| | 2480 | 76.69% | 67.38% | 0.4009 |
| bior1.5 | 1240 | 73.25% | 65.27% | 0.4384 |
| | 1860 | 75.08% | 67.65% | 0.4083 |
| | 2480 | 77.81% | 69.63% | 0.3761 |
| DB4+DB2 | 1240 | 81.13% | 67.32% | 0.3774 |
| | 1860 | 83.58% | 69.83% | 0.3434 |
| | 2480 | 83.50% | 72.05% | 0.3246 |
| DB4+DB2 | 1240 | 75.75% | 68.58% | 0.3969 |
| +bior1.5 | 1860 | 79.92% | 70.34% | 0.3582 |
| | 2480 | 82.19% | 72.13% | 0.3308 |

## IV. CONCLUSIONS

We presented and compared several new feature extraction strategies for ET discrimination. These are summarized in Figure 2. Many feature choices were observed to have a positive effect on ET spike classification performance, including new wavelet features, use of multiple mother wavelets and the inclusion of scalp location features. We also discussed the feasibility of reducing feature vector dimension, the necessity of keeping wavelet features in the high frequency range and the effect of increasing dataset cardinality (size).

Our results indicate that by combining features derived from two types of mother wavelet, we can improve the performance over using features from a single mother wavelet. By using feature set #1 with cooperation of 2 wavelets, the sensitivity has been increased 4.81% and the specificity has been increased 6.76%, compared to Guler's choice of features (derived using only DB2).

The addition of location features improves the classification results in all but one case. Using only wavelet subband maxima as features degrades the classification results somewhat.

Finally, our results indicate a larger set of AEP training samples improves classification performance. Since ETs have varying morphologies, a larger dataset can perhaps provide more examples of ETs features for machine learning. We do not know how many training datasets would be needed to provide optimal performance in this ET classification task, although we suspect it would be much larger than the dataset we used here.

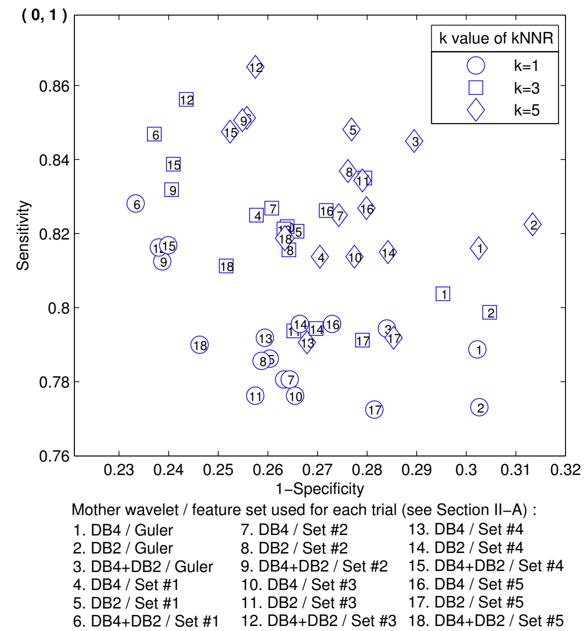


Fig. 2. Composite Summary of Feature Set Evaluations

## REFERENCES


[1] 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.

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

[3] 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.

[4] J. J. Halford, Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized EEG interpretation, Clinical Neurophysiology, 2009.

[5] A. Berdakh and S. H. Don, Epileptic Seizures Detection using Continuous Time Wavelet Based Artificial Neural Networks, IEEE Computer Society Sixth International Conference on Information Technology: New Generations, 2009.

[6] T. Kalayci and O. Ozdamar, Wavelet preprocessing for automated neural network detection of EEG spikes, IEEE Engineering in Medicine and Biology, Mar./Apr., 1995.