EPILEPTIC FOCUS LOCALIZATION USING EEG BASED ON DISCRETE WAVELET TRANSFORM THROUGH FULL-LEVEL DECOMPOSITION

Duo Chen⋆ Suiren Wan⋆ Forrest Sheng Bao‡

⋆ School of Biological Science & Medical Engineering, Southeast University, Nanjing, Jiangsu, China
‡ Dept. of Electrical & Computer Engineering, University of Akron, Akron, OH 44325, USA

Emails: {230139426, srwan}@seu.edu.cn, forrest.bao@gmail.com

ABSTRACT
Electroencephalogram (EEG) is a gold standard in epilepsy diagnosis and has been widely studied for epilepsy-related signal classification, such as seizure detection or focus localization. In the past few years, discrete wavelet transform (DWT) has been widely used to analyze epileptic EEG. However, one practical question unanswered is the optimal levels of wavelet decomposition. Deeper DWT can yield a more detailed depiction of signals but it requires substantially more computational time. In this paper, we study this problem, using the most difficult epileptic EEG classification task, focus localization, as an example. The results show that decomposition level effects the localization accuracy more significantly than mother wavelets. For all wavelets, decomposition beyond level 7 improves accuracy limitedly and even decreases accuracy. We further study what are the most effective bands and features for focus localization. An interpretation of our results is that focal and non-focal epileptic EEGs differ the most at high frequencies of EEG rhythms. The best accuracy of epileptic focus localization achieved in this research is 83.07% using sym6 from levels 1 to 7.

Index Terms— EEG, epileptic focus localization, DWT

1. INTRODUCTION
Affecting approximately 60 million people worldwide, epilepsy is the second most common neurological disorder. Epilepsy is characterized by recurring seizures caused by abnormal discharges in the brain [1]. Directly recording the neuroelectric activities, electroencephalogram (EEG) is a gold standard in epilepsy diagnosis. Diagnostic tasks for epilepsy, such as seizure detection [2, 3], spike detection [4, 5] and focus localization [6, 7], usually require long-term EEG recording up to a few days. Therefore, many computer-aided solutions have been developed to assist neurologists. Combining signal processing and machine learning, most of those approaches model the problem as classification of signals, such as epileptic vs. healthy for epilepsy diagnosis [8, 9], ictal (on seizure) vs. interictal for seizure onset detection [10, 11], etc. The most difficult type of classification is seizure focus localization [5] where the source (anatomical or in terms of EEG channels) of seizure activities need to be identified [12].

Applying Discrete Wavelet Transform (DWT) on epilepsy-related EEG signal classification is gaining ground in recent years. The main advantage of DWT is that the resolution of time and frequency in DWT can be adapted to the frequency content of the examined patterns, thus leading to an optimal time-frequency resolution in all frequency ranges [12, 13]. This makes DWT specially suitable for the analysis of non-stationary signals such as EEG [5, 14].

However, among all applications of DWT in epileptic EEG research, one question still unclear is how many levels of decomposition are sufficient. More levels of decomposition provide a more detailed depiction to the signals, but increase the computational cost, sometimes exponentially (e.g., RBF kernel SVM [15]).

This paper aims at finding the trade-off between decomposition level and EEG signal classification accuracy. For space sake, we pick seizure focus localization, the most difficult epileptic EEG classification problem [5], as an example to study. The machine learning approach to seizure focus localization is to classify focal and non-focal EEGs recorded simultaneously from multiple channels [6, 7]. Another reason we pick focus localization is because DWT has delivered promising results on other epileptic EEG classification problems [10, 16] and therefore may be useful to apply to focus localization.

Seven families of 54 total mother wavelets are used in this research. For each mother wavelet, we decompose the signal to the maximum allowed levels, i.e., full-level decomposition. Not only do we study the relationship between decomposition level and accuracy, we also perform feature selection [3] and wavelet band selection. Choosing suitable features that can best represent the characteristics of the EEG signals is important for EEG classification [10]. Features used in this research are those well-known in wavelet-based EEG signal classification [13, 17]. Our results show that given deep enough decomposition, all mother wavelets deliver similar results. Furthermore, consistently across all mother wavelets, decomposition beyond certain level provides little accuracy improvement and even sometimes decreases performance instead. Our explana-
tion to the results is that focal and non-focal EEG differs the most at high frequencies of EEG rhythms. The best accuracy of 83.07% is achieved by RBF-kernel SVM [18] when using sym6 as mother wavelets and its features from levels 1-7.

2. PROBLEM FORMULATION AND DATASET

We formulate the problem of epileptic focus localization as classifying focal and non-focal EEG channels from multi-channel EEG recordings. “Focal EEG channels” are defined as channels that first show ictal EEG signal changes as visually judged by neurologists. All other channels are “non-focal.” We call EEG recordings from focal and non-focal channels as focal and non-focal EEG signals, respectively. Therefore, epileptic focus localization can be further transformed into a signal classification problem: classifying focal and non-focal EEG signals from simultaneously-recorded multi-channel EEG signals.

In this research, we use the dataset from Department of Neurology, University of Bern, Barcelona, Spain [19] to develop and test our classifier. To our best knowledge, this is the only open dataset that provides clear annotation on focal and non-focal EEG signals during ictal periods. The dataset contains intracranial EEG (iEEG) signals of five patients who had long-standing pharmacoresistant temporal lobe epilepsy (TLE) and were candidates for epilepsy surgery. For each patient, the iEEG signals were determined by at least two neurologists who are also board-certified electroencephalographers. The iEEG signals were sampled at 512Hz and digitally band-pass filtered between 0.5 and 150Hz using a 4th-order Butterworth filter. For each subject, 750 focal and 750 non-focal EEG segments are included, resulting in 7,500 total segments. Each segment lasts 20 seconds. More details about the dataset can be found in [19].

3. METHOD

3.1. Discrete Wavelet Transform

Here we briefly go over concepts of discrete wavelet transform. A wavelet is a quickly vanishing oscillating function localized both in frequency and in time. In continuous wavelet analysis, the signal is decomposed into scaled and translated versions \( \psi_{a,b}(t) \) of a single function \( \psi(t) \) called mother wavelet:

\[
\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right)
\]

where \( a \) and \( b \) are the scale and translation parameters, respectively, with \( a,b \in \mathbb{R} \) and \( a \neq 0 \). The discrete wavelet transform (DWT) [20] is obtained by discretizing the parameters \( a \) and \( b \). In its most common form, the DWT employs a dyadic sampling with parameters \( a \) and \( b \) based on powers of two: \( a = 2^j \) and \( b = k2^j \), with \( j,k \in \mathbb{Z} \). By substituting in Eq. 1, we obtain the dyadic wavelets:

\[
\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k)
\]

The DWT can be written as

\[
d_{j,k}(t) = \int_{-\infty}^{+\infty} s(t)2^{-j/2}\psi^*(2^{-j}t - k) \, dt = \langle s(t), \psi_{j,k}(t) \rangle
\]

where \( d_{j,k} \) are known as wavelet (or detail) coefficients at level \( j \) and location \( k \) [21].

3.2. Wavelet Families

In this paper, we test 7 most commonly used wavelet families’ performance on epileptic focus localization using EEG signal [5, 14]. The 7 wavelet families are: Coiflets, Daubechies, DMeyer, Haar, ReverserBior and Symlets [5]. They include 54 family members (mother wavelets) in total as shown in Table 1.

3.3. Decomposition bands

In this research, each wavelet will be tested through full-level decomposition. The maximum level \( L \) of decomposition is jointly determined by the signal and the mother wavelet to satisfy the condition:

\[
L \leq \log_2 \left( \frac{N}{F} - 1 \right) + 1
\]

where \( N \) is the signal size and \( F \) is the filter size [22]. Each trial in the Bern-Barcelona dataset has 10240 samples and hence each wavelet has a maximum decomposition level as shown in Table 1.

![Fig. 1. An example of 8-level decomposition and corresponding frequency bands](image-url)

More levels of decomposition can provide more details for signals. Fig. 1 illustrates the frequency bands covered by each level of decomposition, given the frequency range from 0.5Hz to 150Hz. As to be discussed in Section 3.4, wavelet coefficients at each band can be used to construct feature vectors for
### Table 1. Wavelet members, maximum decomposition level, best accuracy and best level of decomposition

<table>
<thead>
<tr>
<th>Wavelet</th>
<th>Max level</th>
<th>Best (Accuracy / Level)</th>
<th>Wavelet</th>
<th>Max level</th>
<th>Best (Accuracy / Level)</th>
<th>Wavelet</th>
<th>Max level</th>
<th>Best (Accuracy / Level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bior1.1</td>
<td>13</td>
<td>80.45% / 9</td>
<td>cof4</td>
<td>8</td>
<td>81.01% / 7</td>
<td>rbio2.8</td>
<td>9</td>
<td>81.43% / 9</td>
</tr>
<tr>
<td>bior1.3</td>
<td>11</td>
<td>80.57% / 7</td>
<td>cof5</td>
<td>8</td>
<td>81.21% / 7</td>
<td>rbio3.1</td>
<td>11</td>
<td>81.33% / 6</td>
</tr>
<tr>
<td>bior1.5</td>
<td>10</td>
<td>80.72% / 7</td>
<td>db1</td>
<td>13</td>
<td>80.45% / 9</td>
<td>rbio3.3</td>
<td>10</td>
<td>81.11% / 7</td>
</tr>
<tr>
<td>bior2.2</td>
<td>11</td>
<td>81.05% / 8</td>
<td>db2</td>
<td>11</td>
<td>81.26% / 7</td>
<td>rbio3.5</td>
<td>9</td>
<td>81.16% / 7</td>
</tr>
<tr>
<td>bior2.4</td>
<td>10</td>
<td>81.18% / 9</td>
<td>db3</td>
<td>11</td>
<td>81.17% / 8</td>
<td>rbio3.7</td>
<td>9</td>
<td>80.97% / 9</td>
</tr>
<tr>
<td>bior2.6</td>
<td>9</td>
<td>81.32% / 8</td>
<td>db4</td>
<td>10</td>
<td>80.92% / 7</td>
<td>rbio3.9</td>
<td>9</td>
<td>81.29% / 7</td>
</tr>
<tr>
<td>bior2.8</td>
<td>9</td>
<td>81.41% / 8</td>
<td>db5</td>
<td>10</td>
<td>80.86% / 7</td>
<td>rbio4.4</td>
<td>10</td>
<td>81.49% / 7</td>
</tr>
<tr>
<td>bior3.1</td>
<td>11</td>
<td>80.33% / 7</td>
<td>db6</td>
<td>9</td>
<td>81.13% / 7</td>
<td>rbio5.5</td>
<td>9</td>
<td>81.28% / 9</td>
</tr>
<tr>
<td>bior3.3</td>
<td>10</td>
<td>80.77% / 7</td>
<td>db7</td>
<td>9</td>
<td>81.24% / 9</td>
<td>rbio6.8</td>
<td>9</td>
<td>81.67% / 9</td>
</tr>
<tr>
<td>bior3.5</td>
<td>9</td>
<td>80.57% / 7</td>
<td>db8</td>
<td>9</td>
<td>81.08% / 7</td>
<td>sym2</td>
<td>11</td>
<td>81.27% / 7</td>
</tr>
<tr>
<td>bior3.7</td>
<td>9</td>
<td>80.98% / 8</td>
<td>db9</td>
<td>9</td>
<td>81.12% / 7</td>
<td>sym3</td>
<td>11</td>
<td>81.17% / 8</td>
</tr>
<tr>
<td>bior3.9</td>
<td>9</td>
<td>81.08% / 8</td>
<td>db10</td>
<td>9</td>
<td>80.78% / 9</td>
<td>sym4</td>
<td>10</td>
<td>81.51% / 7</td>
</tr>
<tr>
<td>bior4.4</td>
<td>10</td>
<td>81.44% / 10</td>
<td>rbio1.1</td>
<td>13</td>
<td>80.45% / 9</td>
<td>sym5</td>
<td>10</td>
<td>81.47% / 10</td>
</tr>
<tr>
<td>bior5.5</td>
<td>9</td>
<td>81.66% / 7</td>
<td>rbio1.3</td>
<td>11</td>
<td>80.92% / 7</td>
<td>sym6</td>
<td>9</td>
<td>81.59% / 7</td>
</tr>
<tr>
<td>bior6.8</td>
<td>9</td>
<td>81.37% / 7</td>
<td>rbio1.5</td>
<td>10</td>
<td>81.41% / 7</td>
<td>sym7</td>
<td>9</td>
<td>81.00% / 7</td>
</tr>
<tr>
<td>coif1</td>
<td>11</td>
<td>80.93% / 7</td>
<td>rbio2.2</td>
<td>11</td>
<td>80.97% / 7</td>
<td>sym8</td>
<td>9</td>
<td>81.44% / 7</td>
</tr>
<tr>
<td>coif2</td>
<td>9</td>
<td>81.45% / 8</td>
<td>rbio2.4</td>
<td>10</td>
<td>81.41% / 7</td>
<td>dmey</td>
<td>6</td>
<td>79.67% / 5</td>
</tr>
<tr>
<td>coif3</td>
<td>9</td>
<td>81.65% / 7</td>
<td>rbio2.6</td>
<td>9</td>
<td>81.40% / 7</td>
<td>haar</td>
<td>13</td>
<td>80.45% / 9</td>
</tr>
</tbody>
</table>

the EEG signal [12]. Therefore, deeper decomposition means more frequency bands and thus longer feature vectors. An optimal level of decomposition should provide a good balance between computation complexity and accuracy.

### 3.4. Classification

Wavelet decomposition results in a sequence of coefficients for each band. Because of their high dimensionality, features are often extracted from them to be used by classifiers [5, 9]. The commonly used features for wavelet coefficients in each band include: the maximum coefficient (denoted as $Max$), the minimum coefficient (denoted as $Min$), the mean of coefficients, the standard deviation (denoted as $STD$) of coefficients, the skewness of coefficients, the kurtosis of coefficients, the squared sum of all coefficients (Energy), the normalized standard deviation $\frac{STD}{Max-Min}$ and the normalized energy. The normalized energy is the ratio between energy and the size of the band. Features from all bands form the final feature vector.

The Bern-Barcelona iEEG dataset [19] provides two kinds of EEGs: focal or non-focal. As mentioned earlier in Section 2, the problem of seizure focus localization can be formulated into a binary classification problem on those two kinds of EEGs. We use SVM with RBF kernel as the classifier here. Grid search on SVM parameters is performed on values $10^{-4}, 10^{-3}, \ldots, 10^{3}, 10^{4}$ for both $C$ and $\gamma$. The best $C$ and $\gamma$ of each wavelet family and corresponding wavelet member will be used later in band selection and feature selection.

In order to study the performance of classifiers, especially its ability to overcome individual differences, we use leave-one-subject-out cross validation (CV). In each time, only one subject’s data is used as test set while all others’ data as training set. If we mix the same subject’s data in both training and test sets, the classifier can learn prior knowledge about this subject’s EEG. Hence, leave-one-subject-out CV can truly reveal the robustness of the classifier on overcoming individual difference.

### 3.5. Band Selection & Feature Selection

In this paper, not only do we want to know the optimal depth of wavelet transformation, but also what bands and features are most discriminative in classifying focal and non-focal EEGs. Hence, we performed band selection and feature selection on features extracted from all bands.

Band selection and feature selection are done as follows. Given a mother wavelet, the maximum decomposition level of a single trial EEG is $j$. DWT will give $j+1$ frequency bands ($j$ details and 1 approximation). There are $\sum_{i=1}^{j+1} \binom{j+1}{i}$ combinations of bands. Then, feature selection is performed in each band. Since each band has 9 features, there are $\sum_{n=1}^{9} \binom{9}{n}$ combinations of features. Finally, for each wavelet, we have a total of

$$
\sum_{i=1}^{j+1} \binom{j+1}{i} \cdot \sum_{n=1}^{9} \binom{9}{n}
$$

combinations of bands and features. For each of the combinations, a cross validation is performed. Because of the high time complexity of band and feature selections, we only perform them on mother wavelets that exhibit the best performance in their respective families.
4. RESULTS AND DISCUSSION

The DWT-based approach delivers accurate classification performance for seizure focus localization. Table 1 shows consistent accuracy above 80% across all 54 wavelets. Because we use leave-one-subject-out CV, this DWT-based approach can overcome individual difference and build robust models.

4.1. Accuracy and Decomposition Level

The first goal of this research is to find out an optimal level of wavelet decomposition that yields decent accuracy. The 9 features used in each subband are: 1. the maximum coefficient (denoted as Max), 2. the minimum coefficient (denoted as Min), 3. the mean of coefficients (denoted as Mean), 4. the standard deviation (denoted as Std) of coefficients, 5. the skewness of coefficients, 6. the kurtosis of coefficients, 7. the squared sum of all coefficients (Energy), 8. normalized standard deviation, and 9. normalized energy.

Results show a consistent pattern across all mother wavelets that more levels of decomposition brings little accuracy improvement beyond a certain point. Fig. 3 shows how accuracy changes as decomposition level increases from 1 to maximum levels on 7 mother wavelets that perform the best in their families. When levels are low, deeper decomposition brings significant accuracy improvement. But, after the levels are about half of its maximum level, the accuracy improvement becomes very limited. In many cases, the accuracy even drops. For example, for db2 and bior5.5, the accuracy peaks at level 7. Hence, pursuing more levels of decomposition, which increases computational cost, will not necessarily improve performance in turn.

Table 1 shows that given the optimal level of decomposition, all mother wavelets can yield similar accuracy. Most wavelets obtain accuracy above 80% when being decomposed to levels 7, 8 or 9. The highest classification accuracy is 81.67% when using rbio6.8 decomposed to level 9. This provides us a guideline that the accuracy is not sensitive to the choice of mother wavelet.

We further quantitatively study the relationship between decomposition level and accuracy. Such relationship follows
two-phase exponential association function very well with $R^2 \geq 0.99$ in 95% confidence bounds, as shown by solid curves in Fig. 3. In the setting of this paper, the two-phase exponential association function is

$$y = y_0 + A_1(1 - e^{-\frac{t}{\theta}}) + A_2(1 - e^{-\frac{t}{\gamma}}),$$

where $y$ is accuracy, $x$ is decomposition level, $y_0$, $A_1$, $A_2$, $t_1$, and $t_2$ are 5 regression parameters.

### 4.2. Band Selection & Feature Selection

The band and feature selections are done on the 7 wavelets that perform the best in experiments above in 7 respective families. The best bands and features, and accuracies achieved using them are given in Table 2.

Among all wavelets, sym6 achieves the highest accuracy of 83.07% using 7 features from 6 bands. Its coefficients are given in Fig. 2. Considering the computation complexity, coif3 is the best choice. coif3 has 82.7% accuracy when using only 5 features from 6-bands—that is a feature vector of dimension 30 (6 levels, 25 features from detail coefficients and 5 from approximation coefficients).

<table>
<thead>
<tr>
<th>Wavelet (best in family)</th>
<th>Highest accuracy</th>
<th>Best features</th>
<th>Best bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>bio5.5</td>
<td>82.69%</td>
<td>1, 2, 4, 5, 7, 8, 9</td>
<td>2-7</td>
</tr>
<tr>
<td>coif3</td>
<td>82.7%</td>
<td>1, 2, 4, 5, 8</td>
<td>2-7</td>
</tr>
<tr>
<td>db2</td>
<td>82.77%</td>
<td>1, 2, 4, 5, 7, 8, 9</td>
<td>1-7</td>
</tr>
<tr>
<td>dmey</td>
<td>80.38%</td>
<td>1, 2, 4, 5, 7, 8, 9</td>
<td>1-5</td>
</tr>
<tr>
<td>haar</td>
<td>81.43%</td>
<td>1, 2, 4, 5, 7, 8, 9</td>
<td>1-9</td>
</tr>
<tr>
<td>rbio6.8</td>
<td>82.72%</td>
<td>1, 4, 5, 7, 8, 9</td>
<td>2-9</td>
</tr>
<tr>
<td>sym6</td>
<td>83.07%</td>
<td>1, 2, 4, 5, 6, 8, 9</td>
<td>2-7</td>
</tr>
</tbody>
</table>

Column 3 of Table 2 shows feature combination that yield the highest accuracy for each wavelet. Max, Std, skewness, normalized Std and normalized energy are always chosen. Mean is not chosen in any case. It might be because it hides the details of coefficients in each band.

According to column 4 of Table 2, bands at lower levels yield the best performance. A pattern can be found from the correspondence between those bands and EEG rhythms which are $\delta(<4\text{Hz})$, $\theta(4-7\text{Hz})$, $\alpha(8-15\text{Hz})$, $\beta(16-31\text{Hz})$ and $\gamma(>31\text{Hz})$. The best accuracy is achieved when there are always two DWT bands covering one EEG rhythm. For example, DWT bands 2 to 7 are the best bands for wavelet sym6. The correspondence between the DWT bands of sym6 and EEG rhythms is illustrated in Fig. 4. Such pattern still holds for other wavelets.

The reason why deep DWT bands (e.g., bands 8 and 9 for sym6) do not contribute to classification accuracy can be explained from the property of wavelet transform. Deep DWT bands close to maximum decomposition levels correspond to low frequencies of EEG such as $\delta$ rhythm. Using more features from deep DWT bands increases the length of feature vectors and the portion of features from low-frequency signals increases. One explanation is that focal and non-focal signals do not differ in low frequencies as much as they differ in higher frequencies. More features from lower frequencies actually “confuse” the classifier. This also explains why more levels of decomposition earlier actually decreases the accuracy. We will further investigate this as future work.

Results in Table 2 also show that for many wavelets, band 1 (75-150Hz) does not contribute to accuracy. Because the upper bound of $\gamma$ rhythm is usually considered between 80Hz and 100Hz, we can hypothesize that focal and non-focal EEGs exhibit their major difference in conventional EEG rhythms. Because the Bern-Barcelona cuts off signals at 150Hz, it will be interesting to study whether it is still the case in high-frequency oscillation (HFO) [23] in the future.

### 5. CONCLUSION

Because of its adaptive resolution property for different frequencies, wavelet transform is gaining its ground in epilepsy-related EEG classification. However, a long unanswered question is what the optimal level of decomposition is. In this paper, we investigate this question on the most difficult epileptic EEG classification problem, seizure focus localization, for 7 families of commonly used wavelets. Our results show that across all wavelets, the optimal level of decomposition is around levels 7, 8 and 9. More levels of decomposition actually decreases the accuracy. Band and feature selections are further performed on the best-performing wavelet of each family. From the results, we hypothesize that focal and non-focal EEGs differ the most at the high-frequency end of conventional EEG rhythms. We will further investigate this point in future work.

### 6. REFERENCES


