A Classification Method for Epileptic Electroencephalogram Based on Wavelet Multi-scale Analysis and Particle Swarm Optimization Algorithm

Yiming Tian¹,², Wei Chen²*, Xitai Wang¹,²

ABSTRACT
The automatic classification of epileptic electroencephalogram (EEG) is important for the diagnosis and treatment of epilepsy. In this paper, an epileptic EEG classification method based on wavelet multi-scale analysis and particle swarm optimization is proposed. Firstly, the multi-scale is carried out to the original EEG to extract its sub-bands of different frequency. Secondly, the Hurst exponent and the sample entropy are used to extract the EEG signals and its sub-bands. Finally, the particle swarm optimization (PSO) algorithm is used to optimize the parameters of the extreme learning machine (ELM), and the obtained eigenvector is put to PSO-ELM to realize the purpose of classification of epileptic EEG. The proposed method in this paper achieved 99.7% classification accuracy for the discrimination between epileptic ictal and interictal EEG, which is superior to those methods in other studies.

Key Words: Epileptic Electroencephalogram, Particle Swarm Optimization, Wavelet Multi-scale Analysis, Extreme Learning Machine, Classification


Introduction
Epilepsy is a chronic brain dysfunction caused by a variety of causes. Its typical feature is the repeated over-discharge of brain’s local neurons, leading to central nervous system dysfunction (Iasemidis, 2003). There are 1%-2% world’s populations suffering from epilepsy (Fisher et al., 2005). Epileptic patients have clinical manifestations of muscle twitching, loss of consciousness and so on. Frequent epilepsies not only cause great harm to the patient’s body, but also increase the burden of patient’s family and society. Therefore, strengthening the pre-epilepsy diagnosis and post-treatment work have significant importance.

Electroencephalogram (EEG) has an irreplaceable role in the clinical diagnosis of epilepsy, lesion location and efficacy evaluation. Epileptic ictal EEG shows high synchrotron rhythms, including spike wave, sharp wave, and slow complex waves. At present, the diagnosis of epilepsy is mainly carried out by the experienced doctors to observe the long-term EEG signal. Besides, the workload is huge and subjective factors have great impacts. The results of the different
doctors may not be the same. Therefore, it is necessary to develop a method of automatic classification of epileptic EEG.

Epileptic EEG classification task can be divided into two parts: feature extraction and classification. The nonlinear dynamic correlation method has been applied in the nonlinear feature extraction of epileptic EEG. Researchers have used the nonvolatile method to extract the features of epileptic EEG, such as: Lyapunov index (Lai et al., 2008), Hurst index (Yuan et al., 2011), approximate entropy (Guo et al., 2010), fractal dimension (Yuan et al., 2012), sample entropy (Jouny et al., 2012), power spectrum entropy (Kumar et al., 2010), Lelpel-Ziv complexity (Jouny et al., 2012) and so on. At present, most researches use K-nearest neighbor classifier (Vidyasagar et al., 2015), bayesian classifier (Sood et al., 2014), decision tree (Kovacs et al., 2014), artificial neural network (Sood et al., 2014), support vector machine (Chisci et al., 2010; Kumar et al., 2014) in the classification of epileptic EEG and these researches have achieved some results. Paper (Kumar et al., 2014) compared the classification effects of SVM, Bayesian classifier, K-nearest neighbor classifier, ANN and other classification methods in the same dataset.

This paper presents an epileptic EEG classification method based on wavelet multi-scale analysis and particle swarm optimization algorithm (Kennedy, 2011; Solihin et al., 2011). First, the multi-scale was carried out to the original EEG to extract its sub-bands of different frequency. Then, Hurst exponent and sample entropy were used to extract the nonlinear features. Finally, the particle swarm optimization algorithm had been applied to optimize the parameters of extreme learning machine which was used to classify the epileptic EEG. Lastly, the performance of the classifier was tested by the University of Bonn’s Epileptic EEG database.

**Feature Extraction Method based on Wavelet Multi-scale Analysis**

**Wavelet Multi-scale Analysis**

Wavelet multi-scale analysis is a method of decomposing signal into subspaces on multiple scales. The decomposed signal has the resolution of time domain and frequency domain in each subspace and we can easily analyze the signal, which has a very good application in the biomedical signal processing.

For the signal \( S \), it is decomposed into the approximate part of \( A_1 \) and the detail part of \( D_1 \), and then do the same for the \( A_1 \), getting the approximate part \( A_2 \) and the detail part \( D_2 \), then do the same for \( A_2 \). Here the approximate part corresponds to the low frequency of the signal and the detail part corresponds to the high frequency of the signal. Unlike the Fourier analysis, the high frequency part is hierarchical and gradually generated at different resolutions. Figure 1 shows the signal \( S \) decomposed with the wavelet 3-layer decomposition.

![Wavelet Multi-scale Analysis](image)

**Figure 1. Scheme of 3 layer wavelet decomposition**

**Hurst Index**

Hurst index is proposed by British scientist H. E. Hurst in the middle of the 20th century to determine whether the time series has time-dependent parameters and now it is also used to study the prediction problem of chaotic time series. In this paper, the method of rescaled range analysis \((R/S)\) is used, the steps are as follows:

For a given time series \( x_i (i=0, 1, 2, \cdots, N) \), it is assumed that the length of calculation is \( n \) and the time series is divided into \( M \) subsequences of length \( n \). The \( th \) data of the \( mth \) subsequence is denoted as \( x_{i,m} (i=1, 2, \cdots, n, m=1, 2, \cdots, M) \) and the ratio of rescaled range of the \( mth \) subsequence is:

\[
(R/S)_{n,m} = \frac{R_{n,m}}{S_{n,m}}
\]

(1)

Where \( R_{n,m} \) is the extreme deviation of the \( mth \) subsequence, \( S_{n,m} \) is the standard deviation of the \( mth \) sub-sequence:

\[
R_{n,m} = \max_{1 \leq k \leq n} \sum_{i=1}^{k} (x_{i,m} - \bar{x}_{n,m})
\]

\[
\min_{1 \leq k \leq n} \sum_{i=1}^{k} (x_{i,m} - \bar{x}_{n,m})
\]

(2)
Where $\bar{x}_{n,m}$ is the $m$th sub-sequence. The time series $x_i$ corresponds to the extreme deviation which has calculated length of $n$:

$$ (R/S)_n = \frac{1}{M} \sum_{m=1}^{M} (R/S)_{n,m} $$

(4)

Change the length of the calculation $n$ and recalculate the ratio of rescaled range, we can get a sequence of calculated length and rescaled ranges were got. Hurst argues that an exponential relationship between the sequence of rescaled range ($R/S$) and the calculated length $n$ is as follows:

$$ \left( \frac{R}{S} \right)_n = C \times n^H $$

(5)

Where $C$ is the constant, $H$ is the Hurst index.

Using the method of least squares to fit formula (6) then we can estimate the Hurst index.

### 1.3 Sample Entropy

The sample entropy is a method of measuring the complexity of the signal. According to the definition of sample entropy in (Xing et al, 2013), the sample entropy formula is:

$$ SamEn(m, r) = \lim_{N \to \infty} \left\{ - \ln \left( B^{n+1}(r) / B(r) \right) \right\} $$

(7)

When $N$ is a finite value, the above equation is expressed as:

$$ SamEn(m, r) = - \ln \left( B^{n+1}(r) / B^m(r) \right) $$

(8)

In order to ensure good statistical properties, this paper chooses $m = 2$, $r = 0.2$ as the sample entropy’s parameters.

### Epileptic EEG Classification based on Extreme Learning Machine Optimized by Particle Swarm Algorithm

#### Extreme Learning Machine

As a single hidden layer feed-forward neuron networks (SLFNs), extreme learning machine was formally put forward in 2004. Different from the machine learning method of traditional neuron network, ELM randomly selects the parameters (input weight, hidden element bias) of SLFNs and calculates the output weight by generalized inverse analysis. ELM can be used as a classifier to train neural networks through the activation function and ELM has been widely used to solve the problem of local minimum, slow convergence and complex iterative calculation.

For any $N$ different samples $(x, t)$, $j = 1, 2, \ldots, N$, where $x_j = [x_{j1}, x_{j2}, \ldots, x_{jn}]^T$ is the $j$th sample, each sample contains $n$-dimensional features, and $t_j = [t_{j1}, t_{j2}, \ldots, t_{jm}]^T$ is the encoded class label. All samples belong to $m$ different classes, and the ELM mathematical model with $L$ hidden neurons can be expressed as:

$$ \sum_{i=1}^{L} \beta_i g(w_i \cdot x_j + b_i) = t_j, \ j = 1, \ldots, N $$

(9)

Where $g(x)$ is the excitation function, $w_i$, $b_i$, and $\beta_i$ are the input weight, hidden element offset and output weights of the $i$th hidden neuron node respectively. Equation (9) can be written in matrix form:

$$ H \beta = T $$

(10)

Where the $\beta$ represents the output weight, $T$ is the corresponding coding class label, and $H$ is the hidden layer output matrix:

$$ H = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_1 \cdot x_n + b_1) \\ \vdots & \ddots & \vdots \\ g(w_L \cdot x_1 + b_L) & \cdots & g(w_L \cdot x_n + b_L) \end{bmatrix}_{n \times L} $$

(11)

Since the neural network system is linear, the $\beta$ output weight is obtained by the following equation:

$$ \beta = H^\dagger T $$

(12)

Where $H^\dagger$ is the generalized inverse matrix of $H$.

#### Particle Swarm Optimization Algorithm

Particle Swarm Optimization (PSO) was proposed by E. Eberhart and Kennedy in 1995. The PSO algorithm initially simulates bird’s predation behavior, and each bird does not know the exact
The PSO algorithm is designed to generate a global optimal solution, which can be attributed to the advantages of the location vector, the location of the closest bird from the valley, and all the birds are searching for the nearest bird around the valley. PSO algorithm as a stochastic optimization method is widely used in various fields, such as design of feedback control system, problems about function optimization.

Firstly, the PSO is initialized to generate the particle population. The initial population is evenly distributed in the solution space. All the particles are calculated by an optimized function. Each particle also has a velocity to determine its flight direction and distance. Assuming that the solution space dimension is $d$, the fitness function is $f()$, and each particle is expressed by its own position vector $p = (p_1, p_2, \cdots, p_d)$ and the velocity vector $v = (v_1, v_2, \cdots, v_d)$. These particles can remember their local optimum position $p^*_{i}$ and optimal position $p^*_g$ of entire particle group in the process of moving. Then the particle velocity and location of next generation are as follows:

$$v(t+1) = \omega v(t) + c_1 r_1 (p^*_{i} - p(t)) + c_2 r_2 (p^*_g - p(t))$$

(13)

$$p(t+1) = p(t) + v(t+1)$$

(14)

Where $\omega$ is the inertia factor, which decreases gradually with the growth of generation:

$$\omega(t) = \omega_{\text{max}} - \frac{\omega_{\text{max}} - \omega_{\text{min}}}{\text{iter}_{\text{max}}} \cdot t$$

(15)

Where $\omega_{\text{max}}$ and $\omega_{\text{min}}$ represent the upper and lower boundary values of $\omega$ respectively, $\text{iter}_{\text{max}}$ is the upper limit of the iteration number of the algorithm; $t$ is the number of the current iteration. $c_1$ is used to adjust the experience of the particles themselves, $c_2$ is used to adjust the social experience of the particles, both of which are positive numbers. $r_1, r_2$ is the random number of $[0, 1]$, which are used to keep the diversity of particles. The particle determines the next moving speed and position by (13) and (14). Each iteration produces a range of speeds $[v_{\text{min}}, v_{\text{max}}]$ and positions $[p_{\text{min}}, p_{\text{max}}]$. After the new generation of particles generated, it is necessary to calculate and update the fitness function $f(x)$ of the particle. If the fitness value is better than the current optimal value of the particle, the local optimal position $p^*_{i}(t+1)$ of the particle is updated and the optimal value of the particle is updated:

$$p^*_{i}(t+1) = \begin{cases} p^*_{i}(t) & \text{if } f(p(t+1)) \leq f(p^*_{i}(t)) \\ p(t+1) & \text{if } f(p(t+1)) > f(p^*_{i}(t)) \end{cases}$$

(16)

At the same time, it is necessary to consider updating the global optimal position $p^*_g$ of the whole particle swarm:

$$p^*_g = \max_{p_i}(f(p^*_{i}), f(p^*_2), \cdots f(p^*_m))$$

(17)

**Extreme Learning Machine based on Particle Swarm Optimization Algorithm**

The optimal input weight and the hidden element bias of ELM are selected by PSO under the premise of determining the number of hidden neurons. The classifier is called the PSO-ELM based on the particle swarm optimization algorithm. The concrete steps are as follows:

1) Firstly, the initial population is randomly created, which contains $m$ particles, each particle represents an ELM classifier model. Assuming that the input dimension of the classifier is $m$ then the network structure of the ELM classifier represented by each particle $p_i$ can be expressed by the second-order real matrix:

$$p_i = \begin{bmatrix} W_{i1} & W_{i2} & \cdots & W_{in} & b_1 \\ W_{i1} & W_{i2} & \cdots & W_{in} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ W_{i1} & W_{i2} & \cdots & W_{in} & b_L \end{bmatrix}$$

(18)

Where $L$ is the number of rows of particles $p_i$, $i=1, 2, \cdots, m$, representing the number of hidden neurons, and each row $(w_{1i}, w_{2i}, \cdots, w_{ni}, b_i)$ in the matrix represents the input weight and the hidden element bias of the $f_{ji}$ hidden layer neurons. The number of hidden neurons is already determined, and the size of the particles in the population is the same;

2) Calculate the fitness function of each particle;

3) Use the PSO algorithm to update the rule (13)-(17) to get the next generation of particles;

4) Return 2) until the maximum number of iterations is reached.

5) The optimal ELM classifier model which has the optimal input weight and the hidden element bias is obtained by getting the optimal particle.
Method of Epileptic EEG Classification

In this paper, an epileptic EEG classification method based on wavelet multi-scale analysis and particle swarm optimization is proposed. Firstly, wavelet transform is used to decompose the original EEG signals into five layers, and four different frequency bands are extracted from the EEG signals. Secondly, the frame error and the sample entropy were used to extract the EEG signals from the original EEG signals and the four different frequency bands, and the 10-dimensional eigenvector was obtained by using the two nonlinear methods. Finally, the 10-dimensional eigenvector is used as the input of the PSO-ELM to realize the purpose of epileptic EEG classification. Detailed flow chart is shown in Figure 2.

Feature Analysis

Hurst and sample entropy were used to extract the features from primitive EEG signals and its sub-bands of different frequency of ictal and interictal period. In order to facilitate the comparison of two features of ictal and interictal EEG signals S, we show the distribution of features of primitive EEG signals of ictal and interictal period in Figure 3 and Figure 4. From Fig. 3 (a) we can know that for most samples, the Hurst eigenvalue of the interictal period is significantly greater than the Hurst eigenvalue of the ictal period. From Fig. 3 (b) we can know that the mean value of interictal period is also greater than that of ictal period. Comparison results of Hurst eigenvalue show that epileptic EEG of ictal and interictal period are both persistent and stable sequences, and the stability of interictal period is stronger. Fig. 4 (a) shows the feature distribution of SamEn-S in epileptic ictal and interictal period. For most samples, the value of sample entropy of the interictal period is significantly higher than that of ictal period. Fig. 4 (b) shows the mean value of the sample entropy in interictal period is also significantly higher than the mean value of the sample entropy in ictal period. The comparison of the sample entropy shows that the epileptic EEG of interictal period is more irregular than epileptic EEG of ictal period and the probability of generating new models is higher.

Experimental Results and Analysis

Experimental Data Source

The experimental data comes from the EEG database of epileptic research center of Bonn University. The sampling frequency of EEG signal acquisition system is 12Hz and filtering bandwidth is 0.53~40Hz. The dataset contains five subsets named as Z, O, N, F, and S. The data subsets Z and O are the normal EEG signals of healthy people when eyes open and eyes close; the data subsets N and F contain the EEG signal of outside the lesion area and the lesion area when patients are in interictal epilepsy; the data subset S has the epileptic ictal EEG signal. Each subset contains 100 EEG signals with 4096 points. Each EEG signal is divided into four segments, and calculate features of each segment, and the mean value of the four features is taken as the eigenvalue of the segment data. In this paper, db3 wavelet basis function is used, the parameters of sample entropy are $m=2$, $r = 0.2$. Table 1 is an abbreviation and its meaning in this paper.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>EEG</td>
<td>Hurst-S</td>
<td>Sample entropy of S</td>
</tr>
<tr>
<td>A5</td>
<td>0~5.4Hz signal</td>
<td>Hurst-A5</td>
<td>Sample entropy of A5</td>
</tr>
<tr>
<td>D5</td>
<td>5.5~10.85Hz signal</td>
<td>Hurst-D5</td>
<td>Sample entropy of D5</td>
</tr>
<tr>
<td>D4</td>
<td>10.86~21.7Hz signal</td>
<td>Hurst-D4</td>
<td>Sample entropy of D4</td>
</tr>
<tr>
<td>D3</td>
<td>21.8~40Hz signal</td>
<td>Hurst-D3</td>
<td>Sample entropy of D3</td>
</tr>
</tbody>
</table>

Table 1: Abbreviations and its meanings in this paper

Figure 2: Flowchart of classification algorithm for epileptic EEG

PSO Optimizes ELM

In this paper, the number of hidden neurons of the extreme learning machine is 80 and other parameters are set according to the experience, which has little influences to the result. We set the maximum inertia weight $w_{max}=1.2$, the minimum inertia weight $w_{min}=0.73$, the maximum iteration number $iter_{max}=200$, the particle self-learning factor $c_1=1.5$, the social empirical learning factor $c_2=1.5$. 
Chen W., A Classification Method for Epileptic Electroencephalogram Based on Wavelet Multi-scale Analysis and Particle Swarm Optimization Algorithm

**Classification Result of Single Feature**
Table 2 shows the classification ability of the single feature of epileptic EEG. The following conclusions can be drawn from the Table 2: (1) The accuracy of classification of epileptic EEG for any one feature is higher than 50%. (2) Hurst-S obtains the highest classification accuracy 89.6% of all the single feature. (3) The classification accuracy of the feature extracted from primitive EEG is higher than that extracted from its sub-bands of different frequency. In general, the classification performance of Hurst eigenvalues of EEG is better than that of sample entropy.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Threshold</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hurst-S</td>
<td>0.69</td>
<td>96</td>
<td>84</td>
<td>89.6%</td>
</tr>
<tr>
<td>Hurst-A5</td>
<td>0.84</td>
<td>86</td>
<td>21</td>
<td>82.4%</td>
</tr>
<tr>
<td>Hurst-D5</td>
<td>0.58</td>
<td>82</td>
<td>31</td>
<td>79</td>
</tr>
<tr>
<td>Hurst-D4</td>
<td>0.45</td>
<td>80</td>
<td>58</td>
<td>58.9</td>
</tr>
<tr>
<td>Hurst-D3</td>
<td>0.39</td>
<td>91</td>
<td>17</td>
<td>68</td>
</tr>
<tr>
<td>SamEn-S</td>
<td>0.64</td>
<td>61</td>
<td>75</td>
<td>87</td>
</tr>
<tr>
<td>SamEn-A5</td>
<td>0.26</td>
<td>79</td>
<td>12</td>
<td>57.5</td>
</tr>
<tr>
<td>SamEn-D5</td>
<td>0.37</td>
<td>86</td>
<td>9</td>
<td>59.4</td>
</tr>
<tr>
<td>SamEn-D4</td>
<td>0.68</td>
<td>85</td>
<td>35</td>
<td>67.3</td>
</tr>
<tr>
<td>SamEn-D3</td>
<td>0.83</td>
<td>58</td>
<td>74</td>
<td>74.3</td>
</tr>
</tbody>
</table>

Table 3 shows the comparison results of method proposed by this paper and several other methods on classification for epileptic ictal and interictal EEG. It indicates that the use of wavelet multi-scale analysis combined with nonlinear feature extraction and PSO-ELM is better than other non-linear features and other classifiers.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Accuracy</th>
<th>Classification problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aarabi et al., 2009</td>
<td>99.73%</td>
<td>F-S</td>
</tr>
<tr>
<td>Yuan et al., 2012</td>
<td>99.65%</td>
<td>F-S</td>
</tr>
<tr>
<td>Chua et al., 2011</td>
<td>99.82%</td>
<td>F-S</td>
</tr>
<tr>
<td>Proposed method</td>
<td>99.87%</td>
<td>F-S</td>
</tr>
</tbody>
</table>

**Conclusions**
The theory of nonlinear dynamics has a high value in EEG analysis. Hurst and sample entropy can be used to analyze smaller data with good anti-noise ability and can be used for physiological signal analysis. In this paper, we use hurst and sample entropy to extract the features from EEG signal and its sub-bands. The results show that all the single nonlinear feature can achieve a classification accuracy over 50%, it is shown that the non-linear feature extraction method used in this paper can reflect the nonlinear dynamics of epileptic ictal and interictal EEG signals. This
paper uses the extreme learning machine optimized by particle swarm algorithm as classifier and the classification accuracy of the test set is 99.87%. The results obtained in this paper are superior to those in other studies, and it is shown that the POS-ELM constructed by wavelet multi-scale analysis combined with nonlinear characteristics can accomplish epileptic EEG classification with a satisfactory result.

Acknowledgements
This work was supported by National Key Technology Research and Development Program (2015BAI06B00).

References