A Novel Application of Empirical Mode Decomposition (EMD) to Feature Extraction of Epileptic EEG

Jin Zhang, Jianhao Wei, Xiaofei Liu, Chengqian Wu and Ying Wang

Hunan Normal University, Changsha, Hunan, 410006, *China*

***Abstract* — Electroencephalogram (EEG) is the concentrated expression of physiological activity of the brain. Effective EEG feature extraction methods are key to improving different EEG recognition rates, which is a significant issue in EEG studies. A new feature extraction method is proposed in this paper based on Empirical Mode Decomposition (EMD), which can decompose non-stationary EEG into a series of Intrinsic Mode Functions (IMFs). The new method takes full advantage of characteristics of EEG to extract more effective features. The new method includes four steps: 1) EMD decomposes original EEG into a series of IMFs. 2) The correlation between each IMF and original EEG is calculated, and the energy spectrum of each IMF is calculated in θ band. 3) The IMF with maximum correlation and IMF with biggest ratio of energy are individually selected to substitute original EEG. 4) Several different features (volatility index, variability coefficient, average frequency, and variance) are extracted from IMF with maximum correction, and the phase lock feature is extracted from IMF with biggest ratio of energy in θ band. These features are composed as the feature vector of original EEGs. Support Vector Machine (SVM) is used as a classifier and the effectiveness of the new method is evaluated based on epileptic EEG. The experimental results show that the recognition rate is 94.29% and better than other EEG feature extraction methods.**

Keywords - EEG feature; input vector; support vector machines; empirical mode decomposition

I. INTRODUCTION

With the development of signal technology at present, there is a lot of EEG feature extraction methods, such as Fourier transform (FT), Wavelet transform (WT) and so on. However, Fourier transform can just handle stationary signals[1]. Wavelet transform is not self-adaptive about selection of Wavelet basis, and different wavelet basis functions will have different effects[2]. The EMD can handle time-frequency signal. EMD can not only handle stationary and linear signals, but also can handle non-stationary and nonlinear EEG signals. EMD can be self-adapting according to different signals, so it can decompose complicated signals into a series of simple signals[3].

Epileptic signal is one of the hottest issues of EEG studies. Synchronous is one of typical characteristics of epileptic EEG, and super-synchronous discharge phenomenon can be used to recognize different epileptic EEG. A noise-enhanced algorithm of EMD has been used to extract individual feature as described in[3, 4, 5]. The work in[6, 7] did not consider synchronous features for epileptic signal.

In this paper, considering the typical characteristics and synchronous feature of epileptic EEG, a feature extraction method based on EMD is proposed and typical EEG features are extracted as the input vector. In the experiments, SVM is used as classifier to classify different epileptic EEG. The results show that recognition rate is 94.29%, which is better than other EEG extraction methods.

II. EMD ALGORITHM

EMD can decompose EEG signal into a series of intrinsic mode functions (IMFs), which can represent original EEG signal. During the process to calculate IMFs, there are two basic conditions to end the process: (1) it is must be the same or at most one difference between the number of extreme points and the number of zero crossings, (2)the average value of local maxima and local minima of envelopes is zero at any time.

Assuming that the original signal is $x(t)$, the EMD algorithm is defined as follows[8]:

Step1: Determine all the maxima and minima of the signal $x(t)$.

Step2: According to the maxima and minima, obtain the upper envelopes $f_k(t)$ and lower envelopes $f_p(t)$, respectively, the cubic spline interpolation is used to connect maxima and minima. The local mean is defined as:

$$n(t) = \frac{f_k(t) + f_p(t)}{2}. \tag{1}$$

Step 3: the detail signal is extracted as:

$$a_1(t) = x(t) - n(t). \tag{2}$$

Step 4: Determine whether the detail signal satisfies the two basic conditions that are described above. If it is an IMF, let $d_1(t)$=$a_1(t)$, the detail signal $a_1(t)$ is the first IMF. Otherwise, the step1 to 4 is repeated until an IMF is obtained.

The envelopes of EMD are shown in Fig.1.

Step 5: $x(t)$ minus $d_1(t)$ generates an new signal $R(t)$ which is defined as residue $R_1(t)$. Repeating all of the above steps generates IMF $d_2(t)$, $d_3(t)$,…$d_k(t)$ until a constant function or a monotonic is obtained, the constant function or the monotonic is defined as final residue $R_K(t)$.

According to above five steps, a series of IMFs and the final residue are generated based on EMD algorithm. The original signal $x(t)$ is defined as Eq.(3). In Eq.(3), K and $d_k(t)$ represents the order of IMFs and the kth IMF respectively and the final residue is $R_M(t)$.

$$x(t)=\sum_{k=1}^{K} d_k(t)+R_K(t). \tag{3}$$

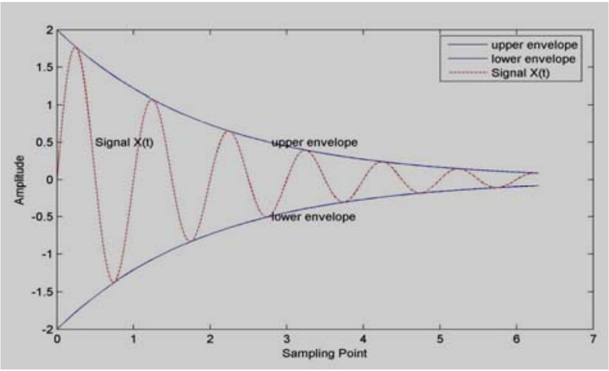

Figure 1. The envelopes of EMD method

III. FEATURE EXTRACTION METHOD BASED ON EMD

The new feature extraction method based on EMD includes three steps and each step is described in detail in section 2.1, 2.2 and 2.3.

Step1: decomposing EEG into a series of IMFs based on EMD. Different IFMs can express different characteristics of original EEG.

Step2: respectively, calculating the correlation between each IMF and EEG and the energy spectrum of each IMF in θ band. The IMF with maximum correlation and the IMF with biggest ration of energy are separately selected to represent original EEG.

Step3: extracting several different features (volatility index, variable coefficient, average frequency, variance and phase lock). The methods to calculate different features are different: calculating four features (volatility index, variable coefficient, average frequency and variance) from correlation, and calculating phase lock feature from ratio of energy.

A. Decomposing EEG

The EEG signal is susceptible to interference by EMG, ECG, EOG and other noise in the sampling signal procession, so suitable pro-process is necessary. EMD can decompose original EEG into a series of IMFs, which can remove different noises and keep main information in different IMFs.

Fig.2 shows that the IMFs generated by EMD process about subset S of seizure EEG signals, which are described in Section 3. The original EEG signal is decomposed into ten order IMFs based on EMD and the 10th IMF is residue. From Fig.2, it is obvious that epilepsy signal during seizure activity changes severely and its amplitude is larger, which is consistent with characters of epileptic EEG.

B. Calculating Correlation

Each IMF can represent the information of original EEG partly. If the correlation between some IMF and original EEG is larger, the IMF can more represent original EEG. So, the biggest correlation IMF is selected to replace original EEG. Actually, the feature of IMF with largest correlation is extracted as the feature of original EEG. The correlation coefficient is used to judge the degree of correlation between IMF and original EEG. The absolute value of the correlation coefficient is in [0, 1]. The value closer to 1 indicates stronger correlation and the value closer to 0 indicates weaker correlation.

Normal subset Z and epileptic subset S in Section 3 are decomposed based on EEG and the corresponding correlation coefficients between IMFs and Z or S are calculated, which is shown in Fig.3. In Fig.3, the lateral axis denotes different orders of IMF from 1*th* IMF to 10*th* IMF. Fig.3 shows that some IMF has the biggest correlation coefficient among all correlation coefficients between IMFs and original EEG.

There are different kinds of EEG and the kth IMF with the biggest correlation coefficient is different. How select the kth IMF is a problem. The method can be described as winner-take-all. For example, some kind of EEG includes 100 channels. After calculating, 5*th* IMF with biggest correlation coefficient in 60 channels EEG and 7*th* IMF with biggest correlation coefficient in 40 channels EEG. So, 5*th* IMF is selected to represent original EEG and the feature of 5*th* IMF is calculated as the feature of original EEG.

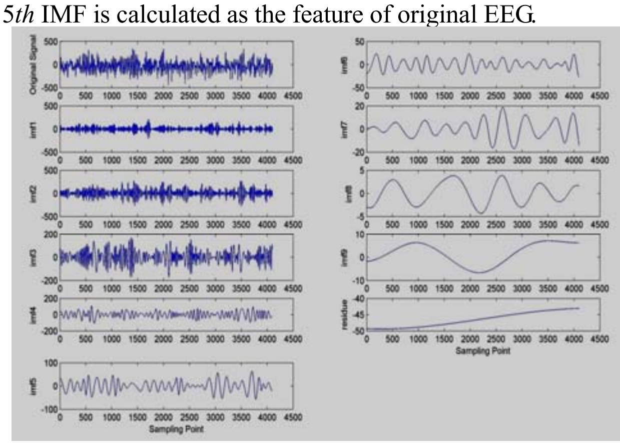

Figure 2. The result of seizure signal based on EMD algorithm

According to this strategy, the EEG dataset in Section 3 has 500 channels, and each channel is decomposed by EMD to get 10 order IMFs. So, 5000 correlation coefficients are calculated and the results of correlation analysis of five kinds of EEGs in Section 3 are gotten as Table 1. In section 2.4.1, IMF2 of EEG in Z, O and S are used to represent original EEG, IMF5 of EEG in N and IMF3 in F are used to represent original EEG.

C. *Calculating Ratio of Energy*

The typical wave of epileptic EEGs includes spike wave, sharp wave, spines slow wave, tip slow wave, many spines composite wave and peak rhythm disorders, etc. Spines slow wave is the most typical waveform of epileptic EEG [9]. The frequency of spines slow wave is distributed in 3Hz to 8Hz, which is defined as θ band. The energy spectrum of epileptic EEG is high in the corresponding θ band, and the θ band of the energy spectrum changes obviously when the EEG changes. So, θ band is selected to calculate energy of EEG.

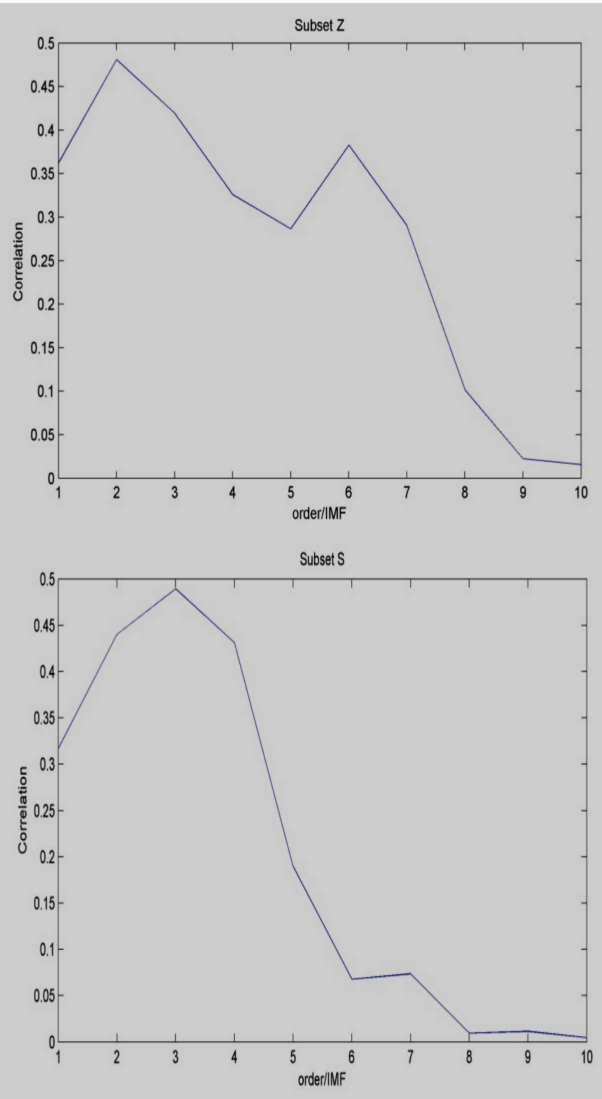

Figure 3. the result of correlation between each IMF and original EEG

TABLE 1. THE CORRELATION OF FIVE SUBSETS SIGNAL

Signal subset	Z	O	N	F	S
Maximum correlation of IMF	IMF2	IMF2	IMF5	IMF3	IMF2

A series of IMFs can be generated based on EMD from original EEG. The power spectrum of IMFs of the epileptic EEG is shown in Fig.4, the part energy spectrum of each IMF is distributed in the θ band. First, the energy of the kth IMF is calculated, which is defined as M_i. Then the energy of the kth IMF is calculated in θ band, which is defined as N_i. When the ratio of N_i and M_i is the maximum, the kth IMF is selected to calculated phase lock as the phase lock of original EEG. The ratio is defined as ratio of energy. From the Fig.4, the ratio of energy of third order IMF is biggest in the θ band. So, the IMF3 is selected to calculate the value of phase lock as one of the features of original EEG.

The IMF with the biggest ratio of energy is different in θ band for each channel EEG. The method is similar to selection of biggest correlation. For example, each kind of EEG includes 100 channels. After calculating, IMF3 with biggest ratio of energy in 70 channels EEG and IMF4 with biggest ratio of energy in 30 channels EEG for θ band. So, IMF3 is selected to represent original EEG and the feature of IMF3 is calculated as the feature of original EEG.

According to this method, the datasets in Section 3 need to calculate 500 energy spectrums. So the energy spectrum of each kind of EEG is obtained: the fourth order IMF according to Z is selected, the third order IMF according to O is selected, the fourth order IMF according to N are selected, the third order IMF according to F is selected, the third order IMF according to S is selected. In section 2.4.2, IMF4 of EEG in Z and N are used to represent original EEG, IMF3 of EEG in O, F and S are used to represent original EEG.

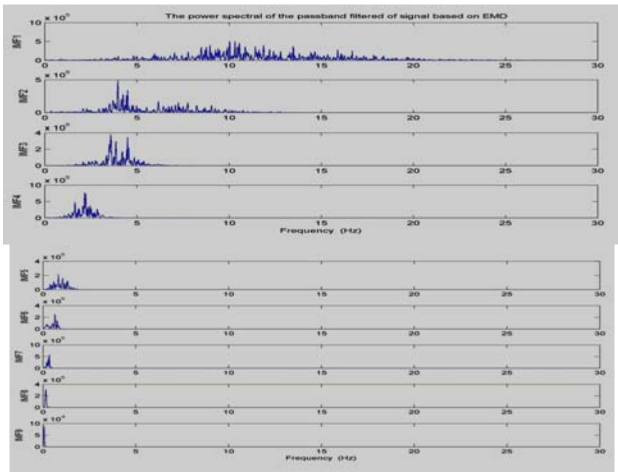

Figure 4. The energy spectrum of the epileptic EEG

D. *Extracting Features*

EEG is nonlinear, non-stationary and strong noise physiological signal, which contains too much complex information. So, effective EEG features needs to be extracted to represent original EEG. The aim of feature extraction is to find out the most effective features to recognize target. Based on the research about the performance of EEG features [10], five kinds of typical features (volatility index, relative coefficient, average frequency, variance and phase lock) of the epileptic EEG dataset are extracted and regarded as the input vector of the classifier. These five features can be classified two kinds: features (volatility index, relative coefficient, average frequency and variance) calculated from just correlation and features calculated from ratio of energy, such as phase lock.

1) Features Calculated from Correlation

According to the results of the Table 1, four features (volatility index, variable coefficient, average frequency and variance) are calculated by MATLAB. Each channel EEG has 4 features, so two thousand features will be calculated for 500 channels EEG. Each kind of EEG with three examples is shown in Table 2.

TABLE 2. THE FOUR FEATURE VALUES OF FIVE KINDS OF EEGS

Signal channel	volatility index	variable coefficient	variance	average frequency
Z026	22.4016	8654.1000	466.6570	0.1458
Z037	15.6852	3445.9000	457.3950	0.1418
Z072	20.3085	37964.000	535.3909	0.3921
O030	20.7079	1429.0000	661.6733	0.1816
O052	15.6735	1514700.0	514.7955	0.2298
O078	29.5870	1317.7000	550.5281	0.2059
N024	21.2323	315.74520	1573.000	0.3191
N042	21.9200	18066.000	1784.000	0.3617
N082	13.9933	5516.8000	362.2022	0.3831
F012	10.2091	4440.6503	650.5241	0.1060
F044	12.7693	888.05310	1129.000	0.0007
F085	7.59370	63587.000	1914.700	0.0634
S013	225.5507	1394.2000	189840.0	0.6671
S044	109.9035	1422.3000	78195.00	0.4934
S090	133.3365	221470.00	31610.00	0.5703

There is obvious difference between normal EEGs in Z, O or seizure-free EEGs in N, F and seizure EEG in S about four feature values (volatility index, variable coefficient, average frequency and variance) from the Table 2. The volatility index values of EEG in S are more than 100, however, the volatility index value of EEGs in Z, O, N and F are less than 100. In addition, the variance values and the average frequency values of seizure EEG in S are significantly higher than normal EEGs in Z, O or seizure-free EEGs in N, F.

2) Feature Calculated from Ratio of Energy

The method to calculate phase lock feature is different from the method of first four features in Section 2.4.1, so it is will be calculated separately.

According to the results in Section 2.3, in Fig.5, the phase lock values are randomly selected from the normal EEG in Z and epileptic EEG in S as an example which is shown. It is obvious that the phase lock values of epileptic EEG are bigger than normal EEG and normal EEG and epileptic EEG can be separate totally based on the phase lock values. The results are consistent with super synchronization phenomenon of epileptic EEG.

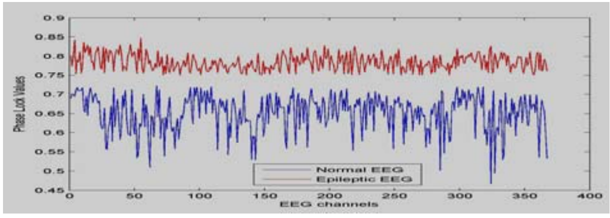

Figure 5. The values of phase lock of EEG in Z and S

IV. PERFORMANCE EVALUATION

A. EEG Dataset

The dataset used in this paper is a real EEG data collected by R.G.A. Andrzejak et al [11].The dataset consists of five subsets, which are represented as Z, O, N, F and S. Each data set is composed of five volunteers tested. Each subset containing 100 channels EEG. Each channel containing 4097 sampling points. The sampling frequency of signals is 173.61Hz and the sampling signal having duration 23.6s. The detailed datasets are described below in Table 3.

TABLE 3. THE EXPLANATION OF DATASET

Signal subsets	Z	O	N	F	S
Recording position	Extracr-anial	Extracr-anial	Intracra-nial	Intracra-nial	Intracra-nial
Signal electrodes	All brain areas	All brain areas	Epilept-ogenic zone	Opposite to epilept-ogenic zone	Epilept-ogenic zone
Testers statement	Eyes open and health	Eyes closed and health	Seizure-free intervals	Seizure-free intervals	Seizure activity

B. Process of EEG Recognition

Five typical features (volatility index, variable coefficient, average frequency, variance and phase lock) of EEG are extracted based on EMD and composed as the feature vector of original EEGs. Support vector machine (SVM) is used as classifier to recognize different kind of epileptic EEGs. Fig.6 shows that the flowchart of feature extraction and classification.

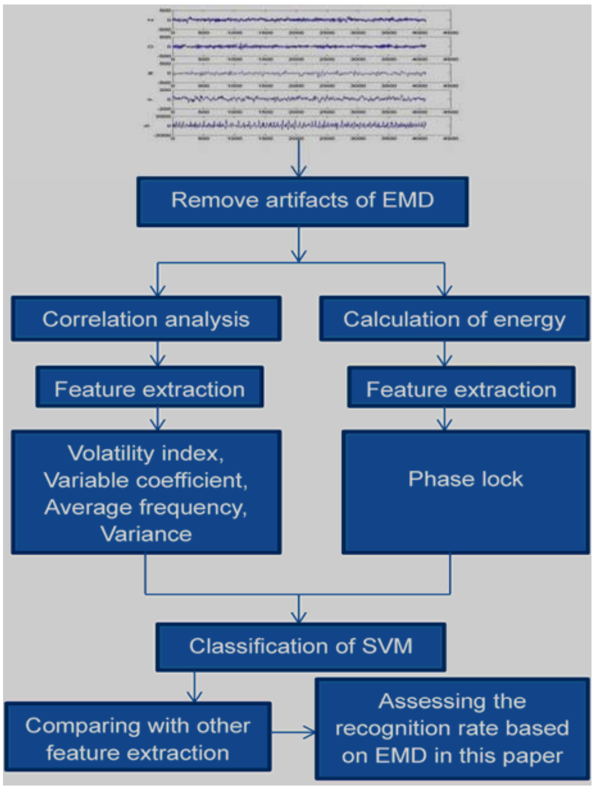

Figure 6. The flowchart of Feature Extraction and classification based on EMD.

C. *The Result of Experiment*

The experiment is divided into three groups, and each group of experiments is performed based on a 10-fold cross validation method. The final recognition results of each group are averaged.

(1) The first group of experiments:

According to the normal signal in Z with healthy and the epileptic EEG in S with seizure, and five typical features of EEG in Section 2 are extracted as the input vector. SVM is used to recognize these two kinds of EEGs. Its purpose is to assess recognition effect of normal EEG and epileptic EEG based on EMD, which is the basic purpose of this study.

(2) The second group of experiments:

The EEG in F with seizure-free and the EEG in S will be classified in this experiment. On the basis of the first group of experiment, its purpose is to assess whether the recognition effect is also good for epilepsy seizure of EEGs and non-seizure of EEGs.

(3) The third group of experiments:

Five kinds of EEGs are classified in this experiment, and the aim of this experiment is to test whether it is more reliable to divide normal EEGs in Z, O, N and F into one class and the epilepsy EEG in S into another class based on the EMD.

The results of recognition rate are shown in Table 4.

The results are obtained from Table 4: the recognition rate to recognize normal EEG from epileptic EEG is 94.29% in the first experiment, and the recognition effect is good for epileptic EEG and normal EEG. The recognition rate is 90% in the second experiment, and this method has good effective recognition for epilepsy seizure of EEGs and non-seizure of EEGs. In addition, the third experiment can be effective to identify five kinds of EEGs based on EMD. The effect of recognition all are fully obvious and the recognition rates all are more than 90% in these three experiments. In order to further verify the effectiveness of the proposed method, the recognition results of our method and other methods based on the same dataset are compared in Table 5. It is shown that the proposed feature extraction method based on EMD has higher recognition rate.

TABLE 4. THE RESULTS OF RECOGNITION RATE

Subject and Recognition rate[%] / Feature Vector	First Experiment: Subset Z and S	Second Experiment: Subset F and S	Third Experiment: Z、O、N、F、S
Volatility Index, Variable Coefficient, Average Frequency, Variance and Phase Lock	94.29	90	92.50

TABLE5. THE RECOGNITION RATE OF DIFFERENT FEATURE EXTRACTION

EEG feature extraction method	Recognition Rate [%]
Wavelets[12]	89.2
Adaptive Neuro-Fuzzy Inference System [13]	86.66
Multilayer perceptron neural network [14]	91.6
Wavelet Analysis and Teager Energy[15]	82.94
Higher Order Statistics[16]	80
The EMD algorithm of this paper	94.29

V. SUMMARY

A new feature extraction method of EEG based on EMD is proposed in this paper. According to real five kinds of epileptic EEGs, the original EEG is decomposed into a series of IMFs based on EMD. Respectively, the correlation between each IMF and original EEG is calculated and the energy spectrum of each IMF is calculated in θ band. The IMF with biggest correlation and the IMF with biggest ratio of energy are separately selected to represent original EEG. Finally, several different features (volatility index, variable coefficient, average frequency, variance and phase lock) are extracted in different methods and are composed as the feature vector of original EEGs. SVM is used to classify different kind of epileptic EEG. The results show that the proposed feature extracting method based on EMD is very effective and can give the recognition rate of 94.29%. In addition, comparing with other methods, the proposed feature extracting method based on EMD has higher recognition rate.

This article took the IMF with biggest correlation as an object of feature extraction of original EEG. The effectiveness of correlation analysis is not researched in this paper, which is future works and will be discussed in detail in another paper.

ACKNOWLEDGMENTS

The work is supported by the National Nature Science Foundation of China (60901080), the Open Platform of Innovation Foundation Project of Hunan(15k082) and Innovation Project of Master of Hunan (CX2016B202).

REFERENCES

[1] Pushpendra Singh, S. D. Joshi1, and R. K. Patney, *et al*. Fourier-Based Feature Extraction for Classification of EEG Signals Using EEG Rhythms. Circuits Syst Signal Process, 2016, vol.35, pp.3700-3715.

[2] Jin Hui. Research of Voice Activity Detection and Speech Enhancement Based on Empirical Mode Decomposition. Harbin: Harbin Engineering University, 2012.

[3] A. Tabrizi, L. Garibaldi, and A. Fasana, *et al*. Early damage detection of roller bearings using wavelet packet decomposition, ensemble empirical mode decomposition and support vector machine. Meccanica, 2015, vol.4, no.50, pp.865-874.

[4] Yanli Yang, Jiahao Deng , and Dali Kan. An improved empirical mode decomposition by using dyadic masking signals. SIViP, Springer, Verlag London, 2015, vol.9, no.10, pp 1259–1263.

[5] Tao Zhang, Peng Xu1, and Lanjin Guo1. Multivariate empirical mode decomposition based sub-frequency bands analysis of the default mode network: a resting-state fMRI data study. Springer Applied Informatics, 2015, vol.1, no.2, pp.1-11.

[6] Xian Du , Sumeet Dua , and Rajendra U. Acharya. Classification of Epilepsy Using High-Order Spectra Features and Principle Component Analysis. Springer, J Med Syst, 2012, vol.36, pp.1731–1743.

[7] Shengkun Xie and Sridhar Krishnan. Wavelet-based sparse functional linear model with applications to EEGs seizure detection and epilepsy diagnosis. Springer, Med Biol Eng Comput, 2013, vol. 51, pp.49-60.

[8] Varun Bajaj and Ram Bilas Pachori. Classification of Seizure and Nonseizure EEG Signals Using Empirical Mode Decomposition. IEEE, IEEE Transactions On Information Technolgy In Biomedicine, 2012, vol.6, no.16, pp.1135-1142.

[9] Ding Shangwen, Qian Zhiyu, and Li Weitao, *et al*. Approximate Entropy Evaluation of EEG θ Rhythm Based on Low-Frequency Photon Stimuli. Laser and Optoelectronics Progress, 2012, vol.49, no.31701, pp.1-5.

[10] Wei Jianhao, Zhang Jin, and Ma Xiang-feng, *et al*. Effectiveness Analysis of EEG Feature Based on Kruskal-Waills Test. Journal of Electronics & Information Technology, 2015, vol.12, pp.93-95.

[11] R.G.Andrzejak, K.Lehnertz, F.Mormann, *et al*. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. phy. Rev.E, 2001, vol.64, no.6, pp.64-71.

[12] Thiago M. Nunes, Andre L.V and Clodoaldo A.M.Lina. EEG signal classification for epilepsy diagnosis via optimum path forest A systemic assessment. Neurocomputing, 2014, vol.20, no.1, pp.10-16.

[13] Shereen A.El-aal, Rabie A.Ramadam, and Neveen I.Ghali. EEG Signals of Motor Imagery Classification Using Adaptive Neuro-Fuzzy Inference System. Springer International Publishing Switzerland, 2016, pp.105-111.

[14] Seyfettin Noyan Oğulata, Cenk Şahin, and Rızvan Erol. Neural Network-Based Computer-Aided Diagnosis in Classification of Primary Generalized Epilepsy by EEG Signals. J Med Syst, 2009, vol.33, pp.107-112.

[15] Chen Jinyan, Zou Junzhong, and Zhang Jian. Automatic Classification of Epileption Activity in EEG Based on Wavelet Analysis and Teager Energy Operator. Advances in Biological Engineering, 2012, vol.33, no.3, pp.150-151.

[16] S.M.Shaficul Alam and M.I.H.Bhuiyan. Detection of Seizure and Epilepsy Using Higher Order Statistics in the EMD Domain. IEEE JOUNRCAL OF BIOMEDICAL AND HETICS, 2013, vol.17, no.2, pp.218-312.