2021 20th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

Recognition of power quality disturbance based on artificial bee colony algorithm to optimize kernel extreme learning machine

QU Li-ping*
College of Electrical and Information Engineering,
Beihua University
Jilin, China
E-mail: 309637683@qq.com

HE Chang-long，ZHANG Jie，GAO Tai-lu
College of Electrical and Information Engineering,
Beihua University
Jilin, China
E-mail: 13147776280@163.com

Abstract—**Aiming at the problem of classification and recognition of power quality disturbance signals, this paper proposes an artificial bee colony algorithm to optimize the classifier model of the kernel extreme learning machine. First, the wavelet transform is used to extract the features of the power quality disturbance signal simulated by Matlab, and then the traditional back-propagation (BP) neural network feedforward neural network is used for classification and recognition. Due to the inherent limitations of the algorithm, the network training calculation is large, the calculation is complicated, and the time is long. Disadvantages, and the proposal of the extreme learning machine just solves this problem, but the accuracy of the classification cannot meet the requirements, and the network needs to be optimized. Therefore, this paper proposes to use the artificial bee colony algorithm to optimize the nuclear extreme learning machine, and classify and recognize the disturbance signal under the optimal parameters of the nuclear extreme learning machine. The simulation results show that after the artificial bee colony algorithm is optimized, the correct rate of classification and recognition has increased by nearly 20%, and the misjudgment rate has dropped to about 3%.**

***Keywords-** Power quality; wavelet transform; nuclear extreme learning machine; artificial bee colony algorithm*

I. INTRODUCTION

In recent years, with the continuous advancement of the industrialization process, the increasing power consumption and the switching of a large number of non-linear and impact loads have made power quality problems more and more serious, especially for users who use high-precision instruments. Putting forward higher requirements for power quality, and providing users with high-quality power is also the development trend of the power supply sector. In order to provide users with high-quality power, first of all, it is necessary to accurately identify various disturbance signals in the power, and take corresponding compensation measures for different disturbances; therefore, accurately identifying the characteristics of power quality disturbance signals is a prerequisite for improving power quality.

At present, the commonly used feature vector extraction methods for power quality disturbance signals include Fourier Transform (FT), Short-Time Fourier Transform (STFT), Wavelet Transform (WT), Hilbert Huang Transform (HHT), etc. It is introduced that FT is integrated over the entire time period without specific time information, resulting in a large error in the spectral distribution of abrupt signals. It has good analysis capabilities for stationary signals, but for unstable electrical energy. The quality disturbance signal cannot be detected and analyzed. In order to overcome this shortcoming, STFT is introduced on the basis of it, that is, time window is added to FT, and the non-stationary signal is regarded as a series of short-term stationary signals. Because the width of the window function is fixed, it is difficult to determine the width of the window function that matches the signal for unknown signals. In order to solve this problem, WT is adopted, which is based on the idea of STFT localization, which can flexibly change the time window and frequency window. Through multi-scale analysis, it can better extract the characteristics of non-stationary signals and better retain the signals. The difference between the two, it is convenient to improve the accuracy of classification. For the method of feature recognition, neural networks are currently used more frequently. BP is used to classify and recognize disturbance signals, and the network uses gradient descent method for iterative training[7]. In the process, it is necessary to constantly adjust the weights, which leads to long training time and easy to fall into local minimums. In order to solve this problem, in 2005, Professor Huang Guangbin of Nanyang Ligong University proposed the concept of an extreme learning machine in the paper [1]. For a single hidden layer feedforward neural network, the input layer weight w and threshold b are randomly performed assignment, the output layer weight β is obtained by the generalized inverse matrix theory, the training speed is greatly reduced, and the generalization performance is good. Professor Huang Guangbin has proved its feasibility in the literature [1].

Therefore, this paper uses wavelet transform to extract the characteristic vector of the disturbance signal, and uses the extreme learning machine and the nuclear extreme learning machine optimized by the artificial bee colony algorithm to classify and recognize, carry out simulation experiments.

II. FEATURE EXTRACTION OF POWER QUALITY DISTURBANCE SIGNAL

A. *Typical power quality disturbance model*

According to the relevant international power quality disturbance standards, the power quality disturbances in the actual power system are concentratedly reflected as transient voltage quality disturbances: voltage swells, voltage sags, voltage interruptions, transient oscillations, notches, flicker, and harmonics, these seven kinds of waves [4]. These disturbances are short in duration and random in nature. Other types of power quality disturbances are less likely to occur and have little harm.

2473-3636/21/$31.00 ©2021 IEEE
DOI 10.1109/DCABES52998.2021.00040

Therefore, this article takes these seven power quality disturbances as examples for analysis and research.

Perform Matlab simulation on the above 7 kinds of disturbance signals and a standard signal, as shown in Figure 1 below, where f=50HZ, T=0.02s, 100 points are sampled in each cycle, 10 cycles, 1000 sampling points are sampled.

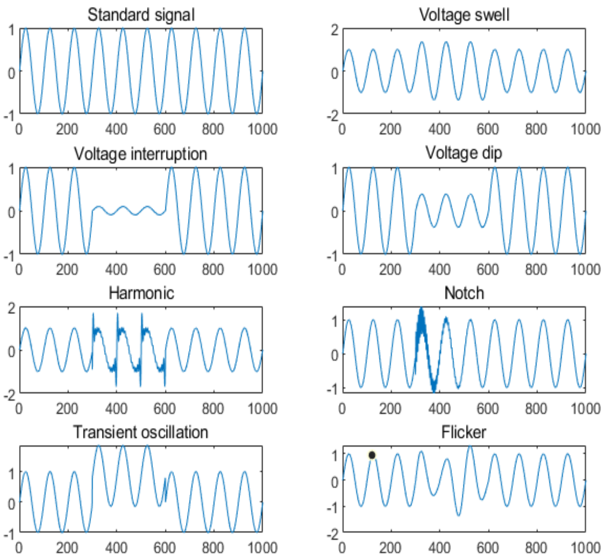

Fig. 1 Disturbance signal waveform

B. *Principle of wavelet transform*

Wavelet transform is a powerful mathematical transform that analyzes the time-frequency characteristics of a signal, and its principle is similar to that of Fourier. Wavelet transform is a wave function of a specific small area. It is translated, stretched, and superimposed to replace the original signal. Its essence is to measure the local similarity between the analyzed signal waveform and the wavelet waveform used. The higher the similarity, the larger the corresponding wavelet coefficient. The essence of wavelet transform is to express a signal function using wavelet function and wavelet transform coefficients. For the time domain signal $f(x)$, its wavelet transform can be expressed as:

$$f(x) = \sum s_{a,b}\Psi_{a,b} \quad (1)$$

Among them, $s_{a,b}$ is the wavelet coefficient and $\Psi_{a,b}$ is the wavelet function. For function $\Psi(t) \in L^2(R)$, if its Fourier transform satisfies:

$$C_\Psi = \int_{-\infty}^{+\infty} \frac{|\Psi(w)|^2}{|w|} dw < \infty \quad (2)$$

It is said that $\Psi(t)$is a basic wavelet. Common wavelets include Haar, Biorthogonal, Coiflet, Daubechies, etc. After the mother wavelet is translated and stretched, a wavelet sequence family $\{\Psi_{a,b}\}$ is obtained:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right), a > 0, b \in R \quad (3)$$

In the formula: a——scale factor b——translation factor;

For the continuous wavelet transform formula of any function $f(x)$:

$$W(a,b)_f = \langle f, \Psi_{a,b}\rangle = \frac{1}{\sqrt{|a|}}\int f(t)\,\Psi^*\left(\frac{t-b}{a}\right)dt \quad (4)$$

In the formula, when a and b take continuous values, it is continuous wavelet transform, which is mainly used in theoretical research. In practical applications, a and b are discretized, and the discrete wavelet transform is:

$$W(a,b)_f = a_0^{-m/2}\int_{-\infty}^{+\infty} f(t)\Psi_{m,n}^*(a_0^{-m}t - nb_0)dt \quad (5)$$

In the formula, $a = a_0^m$; $b = nb_0a_0^m$; m, n are integers; a_0 is a scaling step greater than 1; $b_0 > 0$ and is related to the mother wavelet form.

C. *Equations*

By performing wavelet transform on the power quality disturbance signal, the wavelet transform coefficients obtained have the effective characteristics of the disturbance signal. Because the wavelet coefficients are long and complex and cannot intuitively characterize the characteristics of the signal, it is possible to perform certain operations on these coefficients. On the premise of keeping the original signal features, try to reduce the dimensionality of the feature vector, so as to reduce the size of the input vector when these features are applied to the classification neural network.

Using MATLAB to generate 200 samples of 7 common power quality disturbance signals and a standard signal, 1600 samples in total are abtained simulate the actual situation and ensure the reliability of the analysis results, the parameters of each disturbance (such as the start and end time, amplitude, duration, etc.) of each disturbance change randomly within an allowable range. The sample data is obtained, and the appropriate wavelet basis function is selected for wavelet transformation. Comprehensively considering the characteristics of the disturbance signal, this paper uses the db4 wavelet to decompose the disturbance signal in 8 layers, and the sum of the squares of the wavelet coefficients of each layer will be obtained as its energy representation [5]:

$$E_j = \sum |x_j(k)|^2 \quad (j = 1,2, \ldots 10) \quad (6)$$

Select the standardized energy feature P as the new feature input, and the standardized formula is as follows:

$$P = [p_1, p_2, \ldots p_{10}] \quad (7)$$

$$p_i = \sqrt{\frac{E_j}{k}} \quad (8)$$

In the formula, $j = 1,2,\ldots,10$; k is the number of sampling points. Then the data is normalized between [0,1] to improve the accuracy of distinguishing between data. Figure 2 shows the energy characteristics corresponding to 10 scales of normal signals and 7 kinds of disturbance signals. It can be seen intuitively that the energy of wavelet coefficients is mainly concentrated on scales 4, 5, and 6, when it reaches scale 1and10, it almost drops to 0, and then The next decomposition loses the distinguishing degree of features. If the energies of scales 1 and 10 are also used as the eigenvectors of the disturbance signal, it will not only increase the amount of calculation, but at the same time have little effect on the discrimination between the data, and the gain is not worth the loss. Therefore, this paper chooses scale 2 to scale 9 as the 8-dimensional feature vector of the disturbance signal, which not only preserves the signal characteristics, but also reduces the computational complexity.

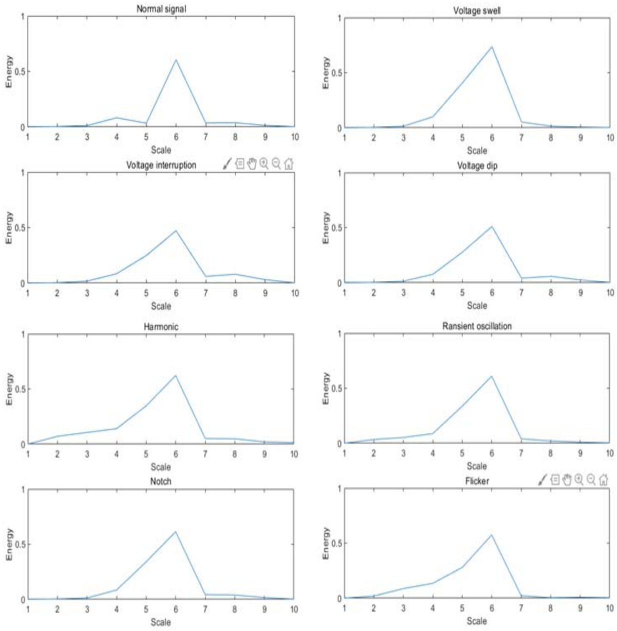

Fig. 2 Disturbance signal energy distribution diagram

III. CLASSIFICATION AND RECOGNITION OF DISTURBANCE SIGNALS

A. BP neural network

The BP (Back Propagation) network was proposed in 1986 by a scientific group headed by Rumelahart and McCelland. It is a multi-layer feedforward neural network trained on the error back propagation algorithm. It is currently one of the most widely used neural network models.

The 1600 sets of sample signals extracted above are divided into training set and test set, with 1200 as training and 400 as testing. There are 400 test samples, 50 samples for each type of signal, and the classification accuracy of each type of signal obtained through simulation is shown in Table 1 below. It can be seen that the training time of the BP neural network is relatively long and the convergence speed is slow. In the actual power system, for the massive detection data, due to the inability to quickly feedback, it may cause a major accident in the power grid.

TABLE 1 BP CLASSIFICATION RESULT

Disturbance type	test samples	Correct numbers	Correct rate(%)	Total correct rate (%)	Training time (s)
Standard signal	50	50	100	76.3	23
Voltage swell	50	47	94		
Voltage dip	50	50	100		
Voltage interruption	50	11	22		
Harmonic	50	46	92		
Transient oscillation	50	6	12		
Notch	50	50	100		
Flicker	50	45	90		

B. Kernel Extreme Learning Machine

In 2005, Professor Huang Guangbin proposed the ELM concept for a single hidden layer feedforward neural network, which is to randomly generate the connection weights between the input layer and the hidden layer. The training process does not need to be adjusted, and the optimal solution can be obtained by only setting the number of neurons in the hidden layer [6].

A single hidden layer neural network with L neurons in the hidden layer can be expressed as:

$$\sum_{i=1}^{L}\beta_i \cdot g(w_i \times X_i + b_i) = O_j \,, j = 1,2,\dots,n; \quad (9)$$

When the number of samples is equal to the number of hidden nodes, the output can be approximated with zero error. Given N different samples (X_i, Y_i), where X is an arbitrarily differentiable activation function $g(x)$, for an Single hidden layer feedforward neural network, w_i and b_i are arbitrarily assigned, and the output matrix H of its hidden layer is invertible , And$\|H\beta - Y\| = 0$[7].

In this regard, Professor Huang Guangbin, based on the extreme learning machine, draws on the concept of the kernel function in the support vector machine, and proposed the kernel extreme learning machine algorithm, which greatly improves the generalization ability and stability of ELM.

Randomly select samples of length N, $(X_i, t_i) \in R^N \times R^M, i = 1, 2, \cdots, N$, $X_i \in R^N$as input. Its expression is:

$$f(X_i) = h(X_i)H^T(HH^T + I/C)^{-1}T \quad (10)$$

In the formula, $h(X_i) = G(w_i X_i + b_i)$ is the output row vector of the hidden layer, G is the activation function, HH^T is the inner product form of $h(X_i)$, which can be replaced by the kernel matrix Ω_{KELM}:

$$\Omega_{KELM} = h(X_i)h(X_i)^T = K(X_i X_j) \quad (11)$$

Therefore, the solution formula of KELM is as follows:

$$f(X) = \begin{bmatrix} K(X, X_1) \\ \vdots \\ K(X, X_N) \end{bmatrix}^T (I/C + \Omega_{KELM})^{-1}T \quad (12)$$

Among them, $\beta = (I/C + \Omega_{KELM})^{-1}T$ is the output weight. There are a variety of kernel functions to choose from in KELM, such as linear kernel function (linear), polynomial kernel function (polynomial), Gaussian radial basis kernel function (RBF), etc. This article selects the RBF kernel function and defines it as follows:

$$K(X_i, X_j) = exp(-\frac{\|x_i - x_j\|}{\gamma^2}) \quad (13)$$

Although KELM introduces a kernel function and a penalty factor C to solve the problem of random initialization of ELM's input weights and obtains a stable predictive output. But its performance is easily affected by the penalty coefficient C and the kernel parameter γ.

C. Artificial bee colony algorithm optimizes nuclear extreme learning machine

Artificial bee colony (ABC) was proposed by Turkish scholar Karaboga in 2005. It simulates the honey-collecting behavior of bees to solve some multi-dimensional and multi-mode optimization problems in life. It was originally applied to numerical optimization problems. Since its proposal, it has received great attention from many scholars, and has been widely used in many fields such as neural networks, data mining, and image recognition.

In this paper, the ABC algorithm and the kernel extreme learning machine are combined, and the ABC algorithm is used to find the optimal parameters of the kernel function to the network to achieve better generalization mapping ability [8]. The kernel function in this article selects the Gaussian kernel function, because it not only has strong learning ability, but also has better generalization, which can achieve good results in practical applications [10].

1) Create a KELM network.

2) Initialize the parameters of the ABC algorithm. Including the size of the colony (N_c), the number of bees (N_e), the number of following bees (N_o), the number of solutions (N_s), the limit value (limit), the maximum number of cycles (MNC) and the initial D dimension Solution $X_i(i = 1,2 \cdots N_s)$. Among them, N_c , N_e , N_o , N_s satisfy the following relationship:

$$N_c = 2N_s = N_e + N_o, N_e = N_o \qquad (14)$$

The D-dimensional solution vector X_i represents the two parameters of the Gaussian kernel function, and the initial solution is a randomly generated value between (-1, 1).

3) Calculate the fitness of each solution according to the following formula.

$$f(X_i) = \begin{cases} 1 & MSE_i = 0 \\ \dfrac{1}{MSE_i + 1} & MSE_i > 0 \end{cases} \qquad (15)$$

In the formula, MSE_i represents the network mean square error of the i-th solution.

4) The honeybee searches for a new solution based on the current memory solution.

$$V_{ij} = X_{ij} + rand(-1,1)(X_{ij} - X_{kj}) \qquad (16)$$

5) Calculate the possible value P_i of each solution.

$$P_i = \frac{f(X_i)}{\sum_{n=1}^{N_s} f(X_i)} \qquad (17)$$

In the formula, $f(X_i)$ is the fitness value of the i-th solution, and the follower bee searches for a new solution from the neighborhood of the existing solution according to these possible values.

6) If the number of failures to update the solution X_i exceeds the preset limit value, then this solution cannot be optimized, it needs to be discarded, and then a new solution is generated to replace it.

$$X_i = X_{min} + \mathrm{rand}(0,1)(X_{max} - X_{min}) \qquad (18)$$

7) If the number of iterations is greater than the maximum number of cycles MCN, the training ends, otherwise, return to step 4.

8) Input the optimized optimal solution into the network, and use the data to simulate and test the network.

D. *Experimental results and analysis*

The performance of ELM and ABC-KELM was tested and compared on the Matlab 2018b simulation platform, and the experimental results were used to verify whether the ABC optimized nuclear extreme learning can improve the accuracy of classification. The experiment has a total of 1600 sets of sample data, 1200 sets are used as training samples, and 400 sets are test samples. The test samples include: seven power quality disturbance signals and normal signals, each with 50 samples. The parameters of the ABC algorithm are set in the experiment: the colony size is 200, the maximum limit number is 50, and the maximum number of cycles (MNC) is 100. According to the modeling steps in 2.3, the number of bees are both 100, that is, $N_s = N_e = N_O = 100$, with the mean square error (MSE) as the performance standard, the search range of the penalty factor C and the kernel parameter γ are [10,1000] and [0.01,1], respectively. In order to avoid the impact of sample differences on the experiment, the same training sample and test sample are used as the training set and test set of ELM and ABC-KELM. The simulation results are as follows:

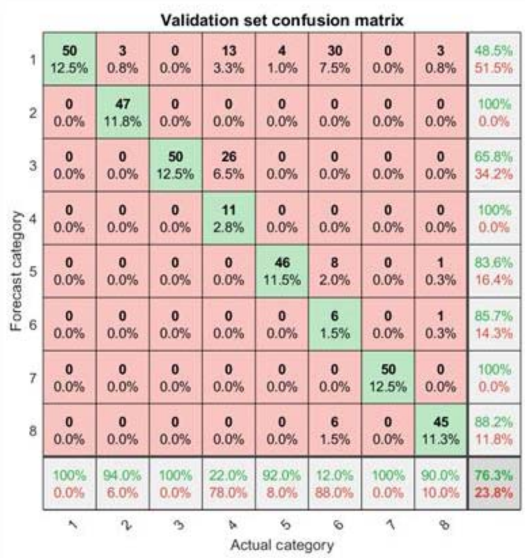

Fig. 4 Validation set of ELM

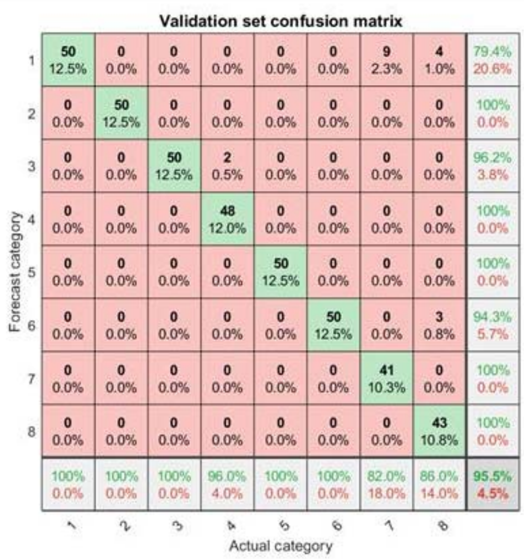

Fig. 5 Validation set of ABC-KELM

It can be clearly seen from the above results that inputting the same sample and using the extreme learning machine for classification and recognition, the poorer classification is the fourth type (voltage sag), 50 test samples, only 11 samples are accurately identified. The correct rate is only 22%. In addition, 13 samples were incorrectly classified into the first category (standard signal), 26 samples were incorrectly classified into the third category (voltage interruption), and in addition, The sixth type (transient oscillation) does not work well. Of the 50 samples, only 14 were accurately identified, the accuracy rate was less than 30%, 26 samples were incorrectly classified into the first class (standard signal), and 10 samples were incorrectly classified. To the fifth category (harmonic), this classification effect is far from reaching the requirements of engineering applications. In order to solve this problem, an artificial bee colony algorithm is proposed to optimize the kernel extreme learning machine to improve the classification accuracy. Through the iterative optimization of the bee colony, the optimal fitness solution of the network is found, and an optimal solution based on the mean square error is obtained. It can be seen intuitively from Figure 5 that the accuracy rate of the fourth category (voltage sag) is 96%, and only 2 samples are incorrectly classified into the third category (voltage interruption) and the sixth category. It has achieved the full recognition effect, and the total accuracy rate has reached 95.5%.

From the table, it can be seen that the artificial bee colony algorithm optimizes the kernel extreme learning machine very well. The accuracy of classification and recognition has increased by about 20%, and the classification accuracy has been greatly improved. However, in actual engineering applications, not only the classification accuracy must be considered, but also the classification speed must be considered to meet actual needs.

TABLE 4 COMPARISON OF ELM AND ABC-KELM CLASSIFICATION

Classification	Training samples	Test sample	Average accuracy
ELM	1200	400	76. 7%
ABC-KELM	1200	400	97. 2%

IV. CONCLUSION

This paper proposes a power quality disturbance classification method based on the ABC algorithm to optimize the kernel extreme learning machine. First, db4 wavelet is used to decompose common power quality disturbance signals in 8 layers, and the square sum of wavelet coefficients is used to characterize its energy characteristics as the characteristic vector of the disturbance signal; then the traditional BP feedforward neural network is used for classification and recognition, due to its algorithm Inherent limitations, the network training calculation is large, the calculation is complicated, the time is long and other shortcomings, and the proposal of the extreme learning machine just solves this problem, but the accuracy of the classification does not meet the requirements, and the network needs to be optimized. Therefore, the ABC algorithm is used to optimize the nuclear extreme learning machine, and the kernel extreme learning machine under the optimal parameters is used to classify the disturbance signal. After simulation experiments, it is verified that this method has a great improvement in the classification effect of the traditional extreme learning machine. It has certain engineering application value.

REFERENCES:

[1] Das Soumya Ranjan,Ray Prakash K.,Sahoo Arun K.,Singh Krishna Kant,Dhiman Gaurav,Singh Akansha. Artificial intelligence based grid connected inverters for power quality improvement in smart grid applications[J]. Computers and Electrical Engineering,2021,93.

[2] Kumar Ashish,Tomar Harshit,Mehla Virender Kumar,Komaragiri Rama,Kumar Manjeet. Stationary wavelet transform based ECG signal denoising method.[J]. ISA transactions,2021,114.

[3] Gong Jie,Li Dayi,Wang Tingkang,Pan Wenhao,Ding Xinzhi. A comprehensive review of improving power quality using active power filters[J]. Electric Power Systems Research,2021,199.

[4] Guo Shifeng,Feng Haowen,Fenga Wei,Lv Gaolong,Chen Dan,Liu Yanjun,Wu Xinyu. Automatic Quantification Of Subsurface Defects By Analyzing Laser Ultrasonic Signals Using Convolutional Neural Networks And Wavelet Transform.[J]. IEEE transactions on ultrasonics, ferroelectrics, and frequency control,2021,PP.

[5] S. Syed Rafiammal,D. Najumnissa Jamal,S. Kaja Mohideen. Detection of Epilepsy Seizure in Adults Using Discrete Wavelet Transform and Cluster Nearest Neighborhood Classifier[J]. Iranian Journal of Science and Technology, Transactions of Electrical Engineering,2021(prepublish).

[6] Ji Xiu,Wang Hui,He Lin. New energy grid-connected power quality management system based on internet of things[J]. Sustainable Computing: Informatics and Systems,2021,30.

[7] Zhou Feifei,Yang Hao,Wang Jiahe,Sun Xi'an,Wu Xu. Prediction of the Quantity of Boxed Meals Preparation by High-Speed Train Based on ELM[J]. Journal of Physics: Conference Series,2021,1910(1).

[8] Mrutyunjaya Sahani,P.K. Dash,Debashisa Samal. A real-time power quality events recognition using variational mode decomposition and online-sequential extreme learning machine[J]. Measurement,2020,157.

[9] Junchang Xin,Zhongyang Wang,Shuo Tian,Zhiqiong Wang. NMR image segmentation based on Unsupervised Extreme Learning Machine[J]. Multidimensional Systems and Signal Processing,2017,28(3).

[10] Linqi Zhu,Chong Zhang,Yang Wei,Xueqing Zhou,Yuyang Huang,Chaomo Zhang. Inversion of the permeability of a tight gas reservoir with the combination of a deep Boltzmann kernel extreme learning machine and nuclear magnetic resonance logging transverse relaxation time spectrum data[J]. Society of Exploration Geophysicists and American Association of Petroleum Geologists,2017,5(3).