

Article

Power Quality Disturbance Classification Based on DWT and Multilayer Perceptron Extreme Learning Machine

Jidong Wang , Zhilin Xu and Yanbo Che *

Key Laboratory of Smart Grid of Ministry of Education, Tianjin University, Tianjin 300072, China;
jidongwang@tju.edu.cn (J.W.); zhilinxu@tju.edu.cn (Z.X.)

* Correspondence: ybche@tju.edu.cn; Tel.: +86-138-2067-2692

Received: 16 April 2019; Accepted: 4 June 2019; Published: 5 June 2019



Abstract: In order to effectively identify complex power quality disturbances, a power quality disturbance classification method based on empirical wavelet transform and a multi-layer perceptron extreme learning machine (ELM) is proposed. The model uses the discrete wavelet transform (DWT) multi-resolution method to extract classification features. Combined with hierarchical ELM (H-ELM) characteristics, the particle swarm optimization (PSO) single-object feature selection method is used to select the optimal feature set. The hidden layer of the H-ELM classifier in the model is trained by forward training. Once the previous layer is established, the weight of the current layer can be fixed without fine-tuning. Therefore, the training speed can be accelerated, the recognition accuracy is almost independent of the parameter adjustment, and the model has strong robustness. In order to solve the problem of data imbalance in the actual power system, a data enhancement method is proposed to reduce the impact of data imbalance and enhance the generalization performance of the network. The simulation results showed that the proposed method can identify 16 disturbances efficiently and accurately under different noise conditions, and the robustness of the proposed method is verified by the measured data.

Keywords: classification; extreme learning machine; feature extraction; optimal feature selection; power quality

1. Introduction

Due to the large-scale use of power electronic devices, there has been an increase in distributed power supply grid-connected non-linear loads. Concurrently the proliferation of reactive power devices and solid-state switches cause the power grid to frequently suffer from various interferences. All these factors result in the emergence of various power quality disturbances [1]. Accurate positioning and identification of power quality disturbances is the premise of power quality analysis and governance. Therefore, pattern recognition of power quality disturbances has become a top priority [2]. The classification study of power quality disturbance (PQD) is divided into three stages, feature extraction, feature selection [3] and classifier design.

At present, power quality disturbance feature extraction [4–8] is mainly based on experience and statistics. In the literature [9], the S transform is used for feature extraction, but the S-transformed Gaussian window changes in the same direction as the frequency, which hinders its adaptability for different signals analysis. In reference [10], the authors further applied the multi-resolution S transform to extract the features, but the analysis of the signal was cumbersome. In another reference [6], the feature extraction was performed by using the short-time Fourier transform, but the Fourier transform window function is fixed, the time-frequency resolution is single, and the extracted features lack the

multi-resolution features. Researchers [11] have proposed using discrete wavelet transform (DWT) to overcome the fixed-resolution problem of short-time Fourier transform to analyze PQD signals. DWT is especially suitable for automatic detection and feature extraction of PQD, especially in terms of transient interference. Moreover, DWT has multiple resolutions, which can determine the initial feature set more accurately.

In recent years, the vast development of data mining, machine learning algorithms and hardware computing capabilities has offered powerful tools to various fields. From the perspective of classifier design, decision trees (DT), probabilistic neural networks (PNN), support vector machines (SVM) and deep neural networks (DNN) have achieved good results in PQD classification. However, the DT [12] classification threshold setting depends on training samples, simultaneously, DT classification has poor generalization ability; compared with DT, PNN [13] is faster and more accurate in general, but it is slower in classifying new cases, PNN requires more storage space than DT; SVM [14,15] needs to set many parameters and is prone to overfitting. Although the neural network model [16,17] has a high classification accuracy, its training and classification speed is slower, and the training network requires a large amount of data. In this study, we extend the extreme learning machine (ELM) and propose a hierarchical ELM (H-ELM) framework for ELM-based multilayer perceptron. H-ELM has both the ability to classify small samples and the high accuracy of deep learning classification. At the same time, the framework has a high classification performance. When processing large amounts of data, the classification speed is fast, and its self-learning feature extraction module can greatly prevent the model over-fitting.

Most existing studies have aimed to optimize the classifiers and feature extraction, but have lacked consideration of the actual operating conditions of power quality disturbance data. Artificial intelligence applications, however, should be able to consider actual grid data characteristics. There is a data imbalance in the power quality disturbance data collected by the power grid, and there are great differences even with the same type of interference. Most of the existing research is focused on the balanced data of simulation, while they do not pay attention to the above problem. In this paper, the data enhancement method is used to deal with the imbalance of data and to achieve data balance. For the disturbance difference, the H-ELM classifier with strong generalization ability is used for classification. For the first time, the method is applied to the classification of power quality disturbances. We propose to use a power quality disturbance recognition method based on DWT and a multilayer perceptron extreme learning machine. The major contributions of this paper are as follows:

(1) For the first time, the H-ELM algorithm is applied to the PQD classification. A comprehensive experimental exploration of H-ELM for PQD classification is performed.

(2) The feature selection algorithm is combined with the H-ELM algorithm to improve the classification accuracy and speed. The simulation results show that the method is more accurate than the traditional methods in classification accuracy. Both speed and the ability to process big data have improved significantly.

(3) In this paper, we consider the problem of PQD data imbalance, and utilize data enhancement to solve the data imbalance. Simultaneously, the paper also uses the data enhancement method to expand the data set to solve the problem of insufficient measured tag data.

The rest of the paper is organized in the following sequence. Section 2 describes power quality disturbance feature extraction. Section 3 discusses the classification of power quality disturbance based on H-ELM. In Section 4, a simulation experiment verifies the feasibility of the algorithm. Section 5 uses the measurement data to verify the feasibility of the algorithm. Section 6 concludes the paper.

2. Power Quality Disturbance Feature Extraction

2.1. Feature Extraction Based on Discrete Wavelet Transform

Wavelet transform [18] is used to analyze stationary and non-stationary signals in various scenes, which can analyze local discontinuities in the signal. Mathematically, continuous wavelet transform (CWT) for a continuous signal with respect to the wavelet function $\psi(t)$ is given by (1).

$$f_{\text{CWT}}(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt \quad a, b \in \mathbb{R}, \quad a \neq 0 \quad (1)$$

Parameters a and b represent scale and conversion parameters, $f(t)$ represents the original signal. In the practical application of CWT, there is redundant information that is not suitable for computer analysis. The study [11] found that DWT is more suitable for the analysis of PQDs, as shown in Equation (2).

$$g_{\text{DWT}}(m, n) = \frac{1}{\sqrt{a_0^m}} \sum_k f(k) \psi\left(\frac{n - kb_0 a_0^m}{a_0^m}\right) \quad (2)$$

The scaling and translation parameters are replaced by functions of m and n integers, i.e., $a = a_0^m$ and $b = kb_0 a_0^m$, respectively, whereas $f(k)$ is the sequence of discrete points of the continuous time signal $f(t)$.

In the feature extraction process, the PQD signal is decomposed using a discrete wavelet transform. This wavelet analysis is actually a measure of the similarity between the mother wavelet and the input signal. The correct choice of wavelet master function is one of the main problems in the execution of DWT applications. In this paper, the widely used Daubechies4 wavelet filter is used as a mother wavelet [19]. The number of decomposition levels l is also very important. Choosing a higher l will bring more information into the system. In the text, the PQD signal is decomposed into eight levels for feature extraction.

The statistical parameters that are used for the feature extraction were obtained from the literature [17]. The seven statistical features are entropy (Ent), standard deviation (σ), mean (μ), kurtosis (KT), skewness (SK), root mean square (RMS) and range (RG). These statistical feature values are calculated using the mathematical formulas of Equations (3) through (9). The power quality waveform is decomposed into eight levels, providing eight detail coefficients and one approximation coefficient, and the total features obtained is 63, which is done for each PQD in the 4500×1280 signal matrix. Finally, a 4500×63 feature matrix is obtained and normalized for classification. The original feature set according to DWT statistics is shown in Table 1.

Table 1. Original feature set. Standard deviation (σ), mean (μ), root mean square (RMS), kurtosis (KT), entropy (Ent), skewness (SK) and range (RG).

Wavelet Coefficients	Statistical Features						
	μ	σ	RMS	KT	Ent	SK	RG
D1	μ_1	σ_1	RMS_1	KT_1	Ent_1	SK_1	RG_1
D2	μ_2	σ_2	RMS_2	KT_2	Ent_2	SK_2	RG_2
D3	μ_3	σ_3	RMS_3	KT_3	Ent_3	SK_3	RG_3
D4	μ_4	σ_4	RMS_4	KT_4	Ent_4	SK_4	RG_4
D5	μ_5	σ_5	RMS_5	KT_5	Ent_5	SK_5	RG_5
D6	μ_6	σ_6	RMS_6	KT_6	Ent_6	SK_6	RG_6
D7	μ_7	σ_7	RMS_7	KT_7	Ent_7	SK_7	RG_7
D8	μ_8	σ_8	RMS_8	KT_8	Ent_8	SK_8	RG_8
A8	μ_9	σ_9	RMS_9	KT_9	Ent_9	SK_9	RG_9

μ_1 represents the mean feature provided by the first layer of detail coefficients, σ_1 represents the standard deviation feature provided by the first layer of detail coefficients, RMS_1 represents the root mean square feature provided by the first layer of detail coefficients, KT_1 represents the kurtosis feature provided by the first layer of detail coefficients, Ent_1 represents the entropy feature provided by the first layer of detail coefficients, SK_1 represents the skewness feature provided by the first layer of detail coefficients and RG_1 represents the range feature provided by the first layer of detail coefficients.

Rang:

$$RG_i = \text{Max}(A_{ij}, D_{ij}) - \text{Min}(A_{ij}, D_{ij}) \quad (3)$$

Entropy:

$$Ent_i = -\sum_{j=1}^N \{A_{ij}^2 \log(A_{ij}^2), D_{ij}^2 \log(D_{ij}^2)\} \quad (4)$$

Standard deviation:

$$\sigma_i = \left(\frac{1}{N-1} \sum_{j=1}^N \{(A_{ij} - \mu_i)^2, (D_{ij} - \mu_i)^2\} \right)^{\frac{1}{2}} \quad (5)$$

Mean:

$$\mu_i = \frac{1}{N} \sum_{j=1}^N (A_{ij}, D_{ij}) \quad (6)$$

Kurtosis:

$$KT_i = \left(\frac{E(A_{ij} - \mu_i)^4}{\sigma_i^4}, \frac{E(D_{ij} - \mu_i)^4}{\sigma_i^4} \right) \quad (7)$$

Skewness:

$$SK_i = \left(\frac{E(A_{ij} - \mu_i)^3}{\sigma_i^3}, \frac{E(D_{ij} - \mu_i)^3}{\sigma_i^3} \right) \quad (8)$$

Root mean square:

$$RMS_i = \left(\frac{1}{N} \sum_{j=1}^N (A_{ij}^2), (D_{ij}^2) \right)^{\frac{1}{2}} \quad (9)$$

where $i = 1, 2, \dots, l$ represents the number of wavelet decomposition at level l . Here N is the number of coefficients in each decomposed data. The PQD waveforms are decomposed into up to eight levels which provide eight detail coefficients (D1, D2, D3, D4, D5, D6, D7, D8) and one approximation coefficient (A8).

2.2. Feature Selection to Select the Best Feature

The combination of different features has different effects on the classifier. In order to verify that the proposed feature is an effective feature and find the best feature combination, the existing research adopts the multi-objective feature selection method. The H-ELM classifier proposed in this paper has a high calculation speed, and the difference in the number of classification features is less than 10, which has little effect on the classification speed of the classifier.

In this paper, the particle swarm optimization algorithm is used to optimize the error of the classifier to select the best feature combination. The features selected by PSO-SVM meet the classification accuracy requirements of H-ELM. The main idea of PSO is to select the subset as a search optimization problem, generate different combinations, evaluate the combination and compare with other combinations. This makes the selection of the best feature subset an optimization problem. The algorithm used in this paper uses SVM classification accuracy as the objective function as shown in Equation (10), where a_i is the SVM classification accuracy.

$$Q_{fitness} = a_i \quad (10)$$

For the fitness definition, the classification accuracy a , representing the percentage of the example of the correct classification, is evaluated by Equation (11). The number of examples of correct and incorrect classifications is represented by c and u , respectively.

$$a = \frac{c}{c + u} \times 100\% \quad (11)$$

The classification accuracy of SVM is the objective function, the classification error is the lowest, and the most suitable feature combination is selected. The best feature combination selected by the PSO single-objective optimization feature selection is shown in Table 2.

Table 2. Particle swarm optimization (PSO) selected best feature set.

Best Feature Set	
20 dB	$\mu_1, \mu_3, \mu_4, \mu_6, \sigma_1, \sigma_4, \sigma_6, \sigma_7,$ RMS_3,RMS_5,RMS_6,RMS_7,RMS_8,KT_7,Ent2,Ent_7,SK_4,RG_2,RG_3
30 dB	$\mu_1, \mu_3, \sigma_2, \sigma_8, \sigma_9, \text{RMS}_1,$ RMS_8,RMS_9,KT_5,Ent_1,Ent_5,Ent_9,SK_9,RG_1,RG_3,RG_4,RG_5,RG_6,RG_9
40 dB	$\mu_4, \mu_5, \text{RMS}_1, \text{RMS}_4, \text{RMS}_6, \text{RMS}_7, \text{RMS}_8, \text{RMS}_9, \text{KT}_4, \text{Ent}_8, \text{RG}_2, \text{RG}_6, \text{RG}_8$
50 dB	$\mu_2, \mu_3, \mu_5, \mu_8, \sigma_1, \sigma_3, \sigma_4, \sigma_6, \sigma_8, \sigma_9, \text{RMS}_7, \text{RMS}_9, \text{KT}_5, \text{KT}_7, \text{KT}_8, \text{Ent}_7, \text{RG}_2$

The feature selection algorithm used in this paper is offline. The feature selection is first performed to find the best combination of features. Then, the feature extraction is performed for the feature quantity to be extracted with the selected best feature, which reduces the computational complexity of the feature extraction and the running time of the algorithm. The selected features are related to the data set and related to the parameter selection of the feature selection, but a common result is that the feature combinations selected by this method have better classification accuracy.

3. Classification of Power Quality Disturbance Based on H-ELM

The essence of machine learning is to establish a network with specific weights and deviations, by using given input data and target values, and to classify them. After the arrival of new data, the output category can be judged through the trained network. The H-ELM framework [20] is a multilayer perceptron extreme learning machine that consists of two independent phases: (1) an unsupervised hierarchical feature representation that automatically extracts features from the input data and the original input features are converted to a higher latitude representation, and (2) the supervised feature classification. Because ELM combines the entire network through feature extraction and classification, it does not need to fine-tune the parameters, and can adapt to the network through the sparse self-encoder, so it has the advantages of fast training speed and high classification accuracy.

3.1. ELM Learning Algorithm

ELM can be built using randomly initialized hidden layer nodes, given power quality disturbance data $\{(x_i, t_i) | x_i \in R^d, t_i \in R^m, i = 1, \dots, N\}$, where x_i is the training data vector, t_i represents the type of each power quality disturbance data, and L represents the number of hidden layer nodes. ELM theory seeks minimal training errors as shown in Equation (12).

$$\text{Minimize : } \|\beta\|_u^{\sigma_1} + \lambda \|H\beta - T\|_v^{\sigma_2} \quad (12)$$

where $\sigma_1 > 0, \sigma_2 > 0, u, v = 0, (\frac{1}{2}), 1, 2, \dots, +\infty$, H is the output matrix of the hidden layer as shown in Equation (13) and β is the output weight. λ is a user-specified parameter and provides a tradeoff between the distance of the separating margin and the training error [21].

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \cdots & h_L(x_1) \\ \vdots & \vdots & \vdots \\ h_1(x_N) & \cdots & h_L(x_N) \end{bmatrix} \quad (13)$$

T is the training data tag matrix, as shown in Equation (14):

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & \vdots & \vdots \\ t_{N1} & \cdots & t_{Nm} \end{bmatrix} \quad (14)$$

The ELM training algorithm can be divided into the following three steps:

(1) Randomly assign hidden layer node parameters;

- (2) Calculate the hidden layer output matrix H ;
- (3) Obtain an output weight vector such as Equation (15).

$$\beta = H^{\dagger}T \quad (15)$$

where $T = [t_1, \dots, t_N]^T$, H^{\dagger} is the generalized inverse matrix of Moore–Penrose of matrix H . According to the ridge regression theory, it was suggested that a positive value $(1/\lambda)$ is added to the diagonal of HH^T in the calculation of the output weights β . To improve the robustness of ELM, the output weight vector can be obtained using Equation (16).

$$\beta = \left(\frac{1}{\lambda} + HH^T\right)^{-1} H^T T \quad (16)$$

The output function of ELM is shown in Equation (17):

$$f(x) = h(x)\beta = h(x)\left(\frac{1}{\lambda} + HH^T\right)^{-1} H^T T \quad (17)$$

3.2. ELM-Based Sparse Autoencoder

The ELM universal approximation function is used in the design of automatic encoders, and sparse constraints are added to the automatic encoder optimization [20]. The optimization model of the ELM sparse autoencoder can be expressed as Equation (18):

$$O_{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \|H\beta - X\|_2 + \|\beta\|_{\ell_1} \right\} \quad (18)$$

where X represents the input data, H represents the random map output, and β is the hidden layer weight to be obtained. In order to generate more sparse and compact features of the inputs, ℓ_1 optimization is performed for the establishment of ELM [20].

The problem in (17) is solved by a fast iteration shrinkage-thresholding algorithm. The implementation process is as follows:

- (1) Calculate the Lipschitz constant γ of the gradient of the smooth convex function ∇p .
- (2) Iterate through $y_1 = \beta_0 \in R^n$, $t_1 = 1$ as an initialization point. For $j(j \geq 1)$ the following holds.
 1. $\beta_j = S_{\gamma}(y_j)$, where S_{γ} is the Formula (19).

$$S_{\gamma} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{\gamma}{2} \|\beta - (\beta_{j-1} - \frac{1}{\gamma} \nabla p(\beta(j-1)))\|^2 + q(\beta) \right\} \quad (19)$$

2. $t_{j+1} = \frac{1 + \sqrt{1 + 4t_j^2}}{2}$
3. $y_{j+1} = \beta_j + (\frac{t_j}{t_{j+1}})(\beta_j - \beta_{j-1})$

3.3. H-ELM Framework

H-ELM is constructed in multiple layers. As shown in Figure 1, unlike the greedy layered training of the traditional deep learning framework, it can be seen that the H-ELM training framework is structurally divided into two separate phases: (1) unsupervised hierarchical feature representation and (2) supervised feature classification [20].

The autoencoder in the H-ELM framework is a self-encoder with sparse constraints. The implementation of the ELM sparse autoencoder is shown in Figure 1b above. It can be seen from the figure that unlike the automatic encoder in deep learning, the input weight of the ELM sparse autoencoder is established by searching the loop path from the random space. ELM theory demonstrates that ELM training using stochastic mapped input weights can approximate any input

data. That is to say, if the automatic encoder is trained according to the concept of ELM, once the automatic encoder is initialized, the parameters do not need to be fine-tuned.

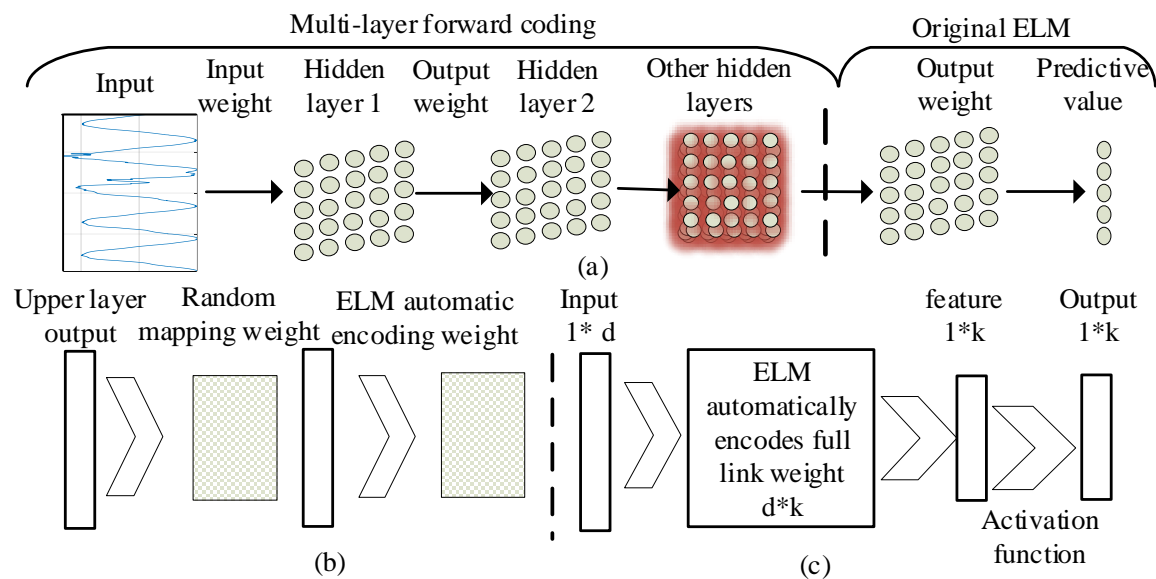


Figure 1. Hierarchical extreme learning machine (H-ELM) learning algorithm: (a) H-ELM overall framework; (b) ELM automatic coding implementation; (c) H-ELM single hidden layout.

3.4. Classification Process

The classification process of this method is shown in Figure 2. The classification prediction model used in this paper is an algorithm based on hierarchical learning. The main trend of hierarchical learning is to conduct research based on deep learning. Deep learning training is challenging, requires a lot of data, and requires pre-processing data, therefore, it is difficult to apply to the classification of power quality disturbances. The H-ELM framework used in this paper is a hierarchical structure, mainly consisting of two parts: “feature extraction and supervised feature classification”. The H-ELM algorithm has a more compact and more meaningful feature representation than the original ELM. Utilizing the advantages of ELM random feature mapping, the hierarchical coding output is randomly projected before the final decision, so that better classification results can be achieved, and the learning speed is faster. The hidden layer of the H-ELM framework is trained in the forward training mode. Once the previous layer is established, the weight of the current layer can be fixed without fine-tuning. Therefore, the proposed algorithm has high accuracy and a fast classification performance.

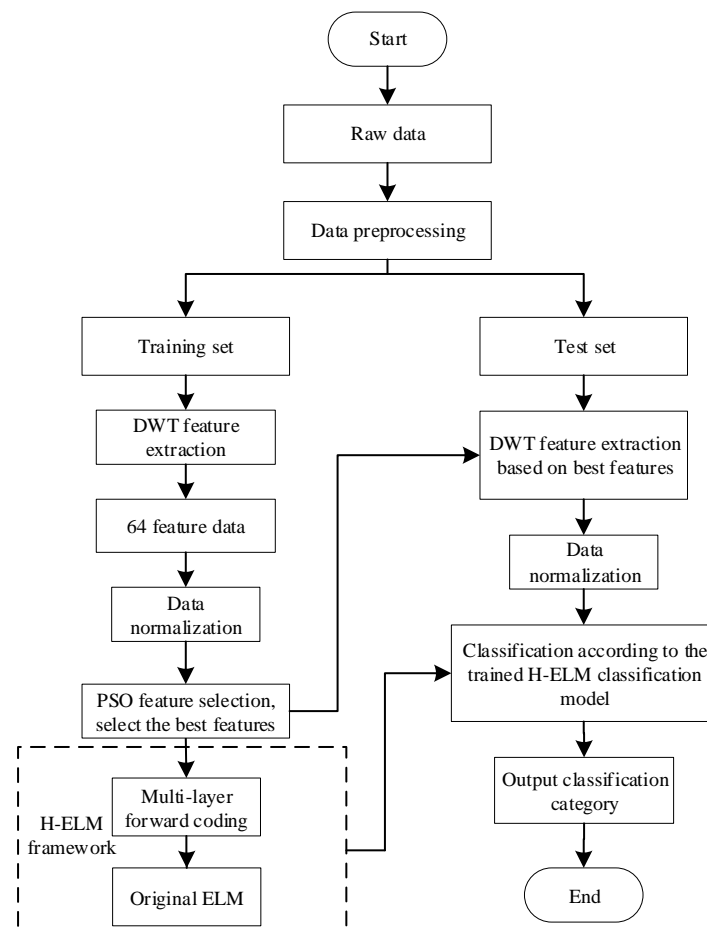


Figure 2. Classification flowchart based on the H-ELM algorithm.

4. Simulation Analysis and Result Verification

The proposed method uses the parametric equations of 15 PQD signals, including pure sine waves, to evaluate the classification performance of the proposed algorithm. The PQD simulation data set consists of nine single types, namely pure sinusoidal waveforms, sag, swell, interrupt, harmonics, Oscillatory transient, flicker, notch and spikes. The six complex PQD signals include sag with harmonics, swell with harmonics, interruption with harmonics, harmonic with flicker, flicker with sag and flicker with swell. The parameter variation of the power quality disturbance equation conforms to the parametric equation of the Institute of Electrical and Electronics Engineers 1159(IEEE-1159) standard [22].

The power quality disturbance signal specifications are: amplitude 1pu, duration $t = 0.2$ s, total period $T = 10$, total sampling point 1280 and sampling frequency 6.4 kHz. Each power quality disturbance type simulation generates 300 signals, to give a total of 4500 signals. These signals are stored in a matrix of size 4500×1280 . A similar matrix increases the Gaussian white noise ratio by 50, 40, 30 and 20 dB at the signal-to-noise ratio (SNR). Part of the simulation signal of power quality disturbance is shown in Figure 3.

An important feature of the multilayer perceptron extreme learning machine is that the classification speed is fast and the algorithm runs for a short time. Compared with the machine learning algorithm, the algorithm has the advantages of high classification accuracy and fast classification speed. In order to verify that the multilayer perceptron also has this characteristic in the classification of power quality disturbance, this paper compares the speed of classification of the same dataset by three algorithms. Table 3 shows the total running time of the three algorithms. All the simulations were accomplished in MATLAB 2016a software on a laptop with Intel Core i5-M-520 processor at 2.40 GHz clock speed

and 8 GB of RAM [18]. As is shown in Table 3, the operating speed of H-ELM is superior to the PNN algorithm.

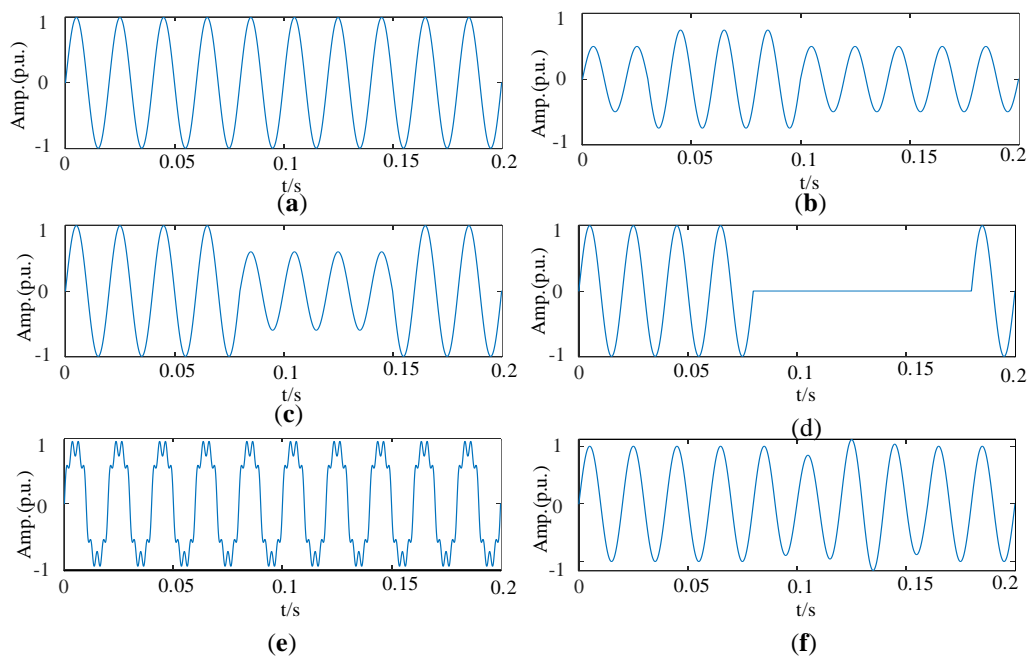


Figure 3. Power quality measured signal: (a) normal signal; (b) swell; (c) sag; (d) interruption; (e) harmonics; (f) flicker.

Table 3. Algorithm running time comparison. SNR, signal-to-noise ratio; PNN, probabilistic neural networks.

SNR	PNN(s)	H-ELM(s)	PSO-H-ELM(s)
20 dB	3.106	0.864	0.325
30 dB	3.134	0.863	0.330
40 dB	3.109	0.738	0.338
50 dB	3.107	0.953	0.283

In order to verify the classification effect of the multilayer perceptron extreme learning machine, we compared it with five other existing methods. The comparison results are shown in Table 4 and Figure 4. It can be seen intuitively from Figure 4 that our method has a good classification effect compared to other methods. Using the original data set classification, the H-ELM algorithm has a higher classification performance than the other three machine learning algorithms. When using the best feature set for classification, under the 20 dB signal-to-noise ratio, the classification effect of the algorithm is significantly improved. The principal component analysis and support vector machines (PCASVM) algorithm also has good PQD recognition accuracy, but its training speed is slow, for the same amount of data, and its operating speed is 30 times that of the H-ELM algorithm. Therefore, the high performance of the H-ELM can be seen.

Table 4. Comparison of classification accuracy of five methods.

Classifier	50 dB(%)	40 dB(%)	30 dB(%)	20 dB(%)
PCASVM	98.30	97.40	96.30	95.00
ELM	98.40	97.23	94.90	91.32
PNN	94.67	94.32	93.64	91.00
H-ELM	98.10	97.60	96.27	93.20
PSO-H-ELM	98.60	98.10	97.67	95.20

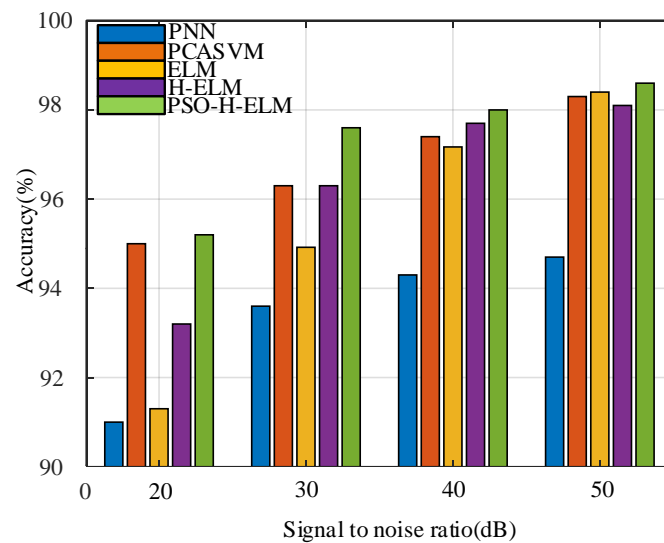


Figure 4. Five methods of classification accuracy rate histogram. PCA SVM—principal component analysis and support vector machines.

Table 4 shows that the classification performance of the H-ELM algorithm is best when using the best feature set selected by PSO for classification. The performance of the pure H-ELM algorithm is better than that of other machine learning algorithms for the classification of power quality disturbances. It can be clearly seen from the results of the first four sets of experiments that the classification performance of the H-ELM classifier is better under the same data volume and the same feature set. Compared with the existing methods, the classification effect of the model proposed in this paper is obviously improved under various signal-to-noise ratios. It is proved that H-ELM has better classification performance for power quality disturbance data.

In the simulation analysis, in order to verify that the features selected by the PSO feature selection promoted the classification results, the experiments using the original 63 features for classification and the classification using the best feature set were analyzed. The experimental results are shown in Table 5. From the four aspects of training time, training accuracy, test time and test accuracy, it can be clearly seen that the features determined by the PSO feature selection are used for classification, which has better classification accuracy and faster classification speed. Feature selection is only the determination phase of the initial feature, and is not embedded in the program. Just like the expert determines the feature, it only implements this process through the optimization method, so as to get a better classification effect.

Table 5. Comparison between the PSO-H-ELM and H-ELM algorithms.

SNR	Classifier	Train Time (s)	Train Accuracy (%)	Test Time (s)	Test Accuracy (%)
20 dB	PSO-H-ELM	0.155	95.55	0.150	95.20
	H-ELM	0.204	95.54	0.164	93.20
30 dB	PSO-H-ELM	0.101	97.57	0.140	97.67
	H-ELM	0.194	97.25	0.144	96.27
40 dB	PSO-H-ELM	0.120	98.61	0.163	98.10
	H-ELM	0.202	98.53	0.170	97.60
50 dB	PSO-H-ELM	0.169	99.24	0.120	98.60
	H-ELM	0.200	98.93	0.115	98.10

Figure 5 shows a comparison of training and test times for categorizing the best features selected using the PSO feature selection and classifying them with the original feature set. It can be seen from the figure that the classification is performed using the best feature set, and the training speed and test speed are significantly better than the original features.

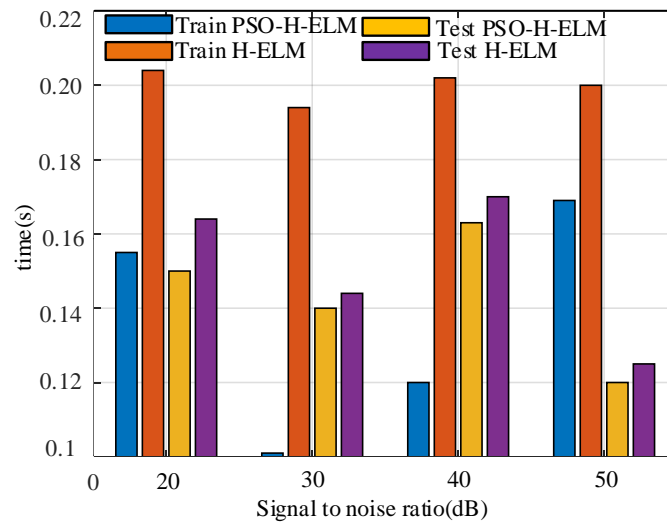


Figure 5. Test and training time histogram.

Figure 6 shows a comparison of the classification accuracy using the best feature set and the original feature set. As can be seen from the figure, the blue histogram is significantly higher than the orange histogram. It shows that the classifier's classification accuracy is significantly improved after using the best feature set.

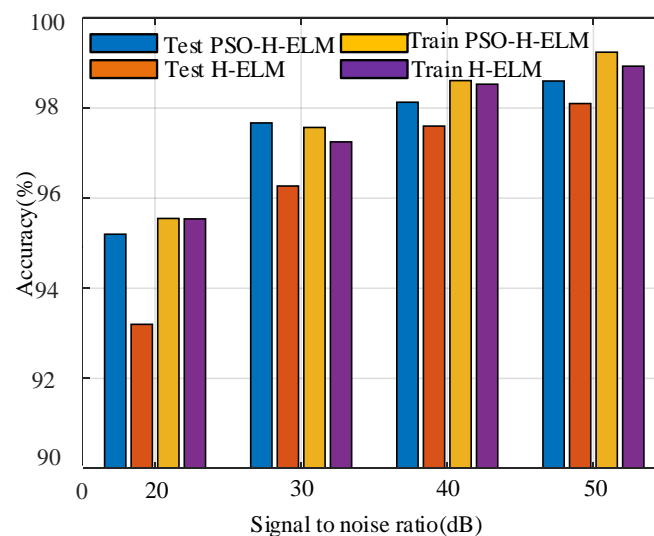


Figure 6. Test and training accuracy histogram.

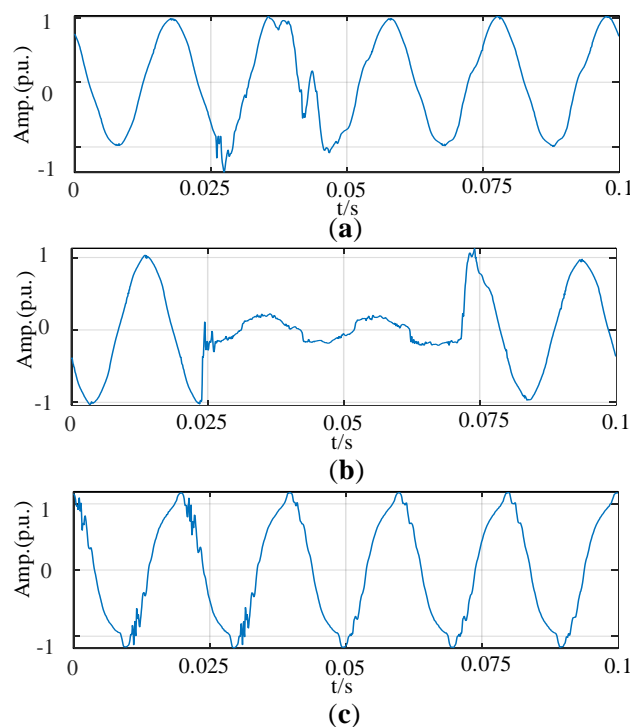
Table 6 shows the classification accuracy rate of each disturbance, and that the overall disturbance classification accuracy rate is above 95%, which satisfies the actual classification needs. The misclassified samples are mainly concentrated in the 20 dB signal, with and without harmonic disturbance, indicating that high noise has greater interference with the identification of signal harmonics.

Table 6. Classification effect of each disturbance type of PSO-H-ELM.

Disturbance Type	Classification Accuracy (%)			
	50 dB	40 dB	30 dB	20 dB
Normal	100	100	100	98
Sag	96	96	94	94
Swell	88	86	88	90
Interruption	100	100	100	100
Harmonics	100	100	100	100
Oscillatory transient	100	100	100	100
Spike	100	100	100	100
Flicker	96	94	96	84
Periodic notch	100	100	100	100
Sag with harmonics	98	98	92	82
Swell with harmonics	100	100	100	98
Interruption with harmonics	100	100	100	100
Flicker with Harmonic	100	100	100	100
Flicker with sag	98	98	96	84
Flicker with swell	100	100	100	100

5. Real Signal Classification Verification

To further verify the feasibility of the proposed method in the actual signal, in this section, a set of actual signals is used to test the effectiveness of the H-ELM. The data set is provided by the IEEE Power Engineering Society database [23,24] for PQD classification. This data set has been tested in reference [25] for power quality classification effects to meet the needs of the experiment. The sampling rate of the supplied signal is 256 points per cycle. Each signal has a length of 1536. The obtained waveforms are determined label by label, and the data set is processed according to the data enhancement method. The actual disturbance signal is shown in Figure 7.

**Figure 7.** Power quality measured signal (a) transient oscillation; (b) sag; (c) harmonics.

The actual power quality disturbance data has an unbalanced feature. For example, the type of disturbance of voltage sag accounts for more than 80% of all disturbances. There is no relevant research

on this problem. This paper proposes a data enhancement method to preprocess data. In computer vision, data enhancement is often used to increase the number of training samples to enhance the generalization performance of the classifier. In this paper, for the problem of data imbalance, the data enhancement method is adopted, the amount of data is equalized and the data enhancement operation is performed for disturbances such as flicker with less data volume. The data enhancement operation mainly adopts random cropping, moderately increases random noise, reverses the signal, etc., and performs random extraction verification on all the operation signals to ensure that the data after data enhancement belongs to the disturbance type data. The data enhancement operation is shown in Figure 8.

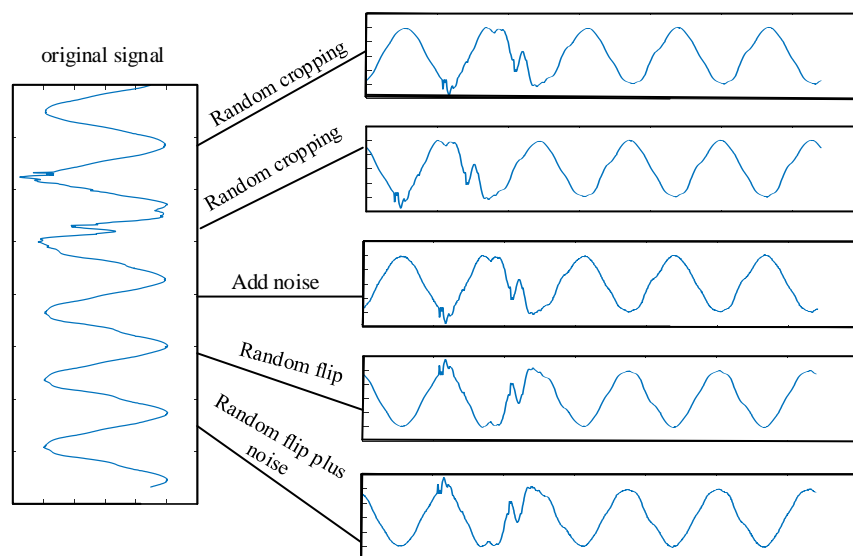


Figure 8. Data enhancement operation diagram.

The measured data set is classified and verified according to the method in this paper. The optimal feature combination and classifier parameters are shown in Table 7.

Table 7. The H-ELM algorithm parameter setting.

Parameter	Numerical Value
Best feature set	$\mu_3, \mu_7, \sigma_2, \sigma_7, \text{RMS}_4, \text{RMS}_6, \text{KT}_6, \text{KT}_7, \text{KT}_9, \text{Ent}_2, \text{Ent}_6, \text{Ent}_7, \text{Ent}_8, \text{RG}_1$
H-ELM hidden layer node	$N1=N2=10, N3=290$
L2 penalty P on the last layer of ELM	2^{-30}
Scale factor S	0.8

We verified the real signals by five methods. The classification accuracy rate and algorithm running time of the five methods are shown in Table 8. It can be seen from the table that the proposed method achieves better classification performance on the measured data. Since the amount of real signal data is small, the samples of various types of signals are small, and even after the data enhancement operation, the amount of data is only 1000 sets. Therefore, the improvement of classification performance by the best features is affected. Since the real signal is more complex than the simulated signal, the real signal accuracy is reduced compared to the simulated signal.

Table 8. Comparison of classification accuracy and algorithm running time of five methods.

Classifier	Classification Accuracy (%)	Algorithm Runtime (s)
PCASVM	83.94	1.28
ELM	89.73	0.34
PNN	84.93	0.98
H-ELM	92.56	0.12
PSO-H-ELM	93.01	0.08

The best feature set is selected by the feature selection algorithm. The disturbance classification results obtained by the method of the present invention are shown in Table 9. There is no data imbalance treatment, the classification accuracy is 92% after the data enhancement process, the average recognition accuracy of the disturbance is 93% after eliminating the influence of data imbalance. The results of the real data classification are lower than the simulation results. The main reason is that the training data is less, the data contains multiple disturbances, and the labeling is inaccurate. In general, the method has a good classification effect and can be adapted to the disturbance classification in the actual power grid.

Table 9. True signal classification result.

Signal Type	True Signal Accuracy (%)
Sag	91.6
Swell	93.1
Harmonics	92.0
Oscillatory transient	98.4
Waveform distortion	88.9
Overall accuracy	93.0

6. Conclusions

Aiming at the problem of identifying complex power quality disturbances, a method for fast and accurate identification of power quality complex disturbance based on DWT and H-ELM is proposed. The simulation results of the example are as follows.

(1) The feature extraction is performed by DWT, the feature selection is performed by the PSO feature selection algorithm, and the feature combination with the best classification performance is selected. Based on the selected classification features, a network with good generalization performance can be trained.

(2) The data enhancement method is used to deal with the problems of data imbalance in the power quality disturbance data and the small amount of data, and to enhance the generalization ability of the network. H-ELM is faster than traditional machine learning algorithms, it can extract more important information from features, and its generalization ability is stronger.

(3) Compared with the deep learning algorithm, the advantage of this method is that it adopts a lightweight structure that combines the ability to process big data and maintain rapidity. Compared with most machine learning methods, the training speed of this method and the ability to process a large data volume is outstanding. It can be used for offline large data volume processing as well as for online analysis.

In the area of disturbance identification of power quality, how to build the local knowledge map of power systems, analyze the underlying data, and realize streaming online processing is the direction needed to improve the real-time classification of power quality disturbance, which needs further research.

Author Contributions: J.W. contributed to establishing the model and writing the paper. Many ideas were suggested by Z.X. to support the work, and Xu performed the simulations. Y.C. analyzed the data, reviewed the work and modified the paper. In general, all authors cooperated during all stages of the research.

Funding: This research was funded by SGCC program: Research on Extensive Application and Benefit Evaluation of Typical Power Substitution Technology Considering Power Quality Influence, grant number 52182018000H.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, J.; Pang, W. Application of fisher discriminant analysis in steady-state power quality evaluation of grid-connected photovoltaic system. *Electr. Power Autom. Equip.* **2017**, *37*, 50–54.
2. Wang, Y.; Xiao, X.; Liu, Y.; Liu, B. Classifying features selection and classification based on mahalanobis distance for complex short time power quality disturbances. *Power Syst. Technol.* **2014**, *38*, 1064–1069.
3. Singh, U.; Singh, S.N. Optimal feature selection via NSGA-II for power quality disturbances classification. *IEEE Trans. Ind. Inform.* **2018**, *14*, 2994–3002. [[CrossRef](#)]
4. Liu, X.; Liu, B.; Xu, D. Fuzzy classification of power quality signals based on pattern linguistic values. *Trans. China Electrotech. Soc.* **2015**, *30*, 392–399.
5. Huang, J.; Qu, H.; Li, X. Classification for hybrid power quality disturbance based on STFT and its spectral kurtosis. *Power Syst. Technol.* **2016**, *40*, 3184–3191.
6. Li, J.; Teng, Z.; Tang, Q. Detection and classification of power quality disturbances using double resolution S-Transform and DAG-SVMs. *IEEE Trans. Instrum. Meas.* **2016**, *66*, 2302–2312. [[CrossRef](#)]
7. Qu, H.; Liu, H.; Li, X.; Huang, J. Feature combination optimization for multi-disturbance classification of power quality. *Electr. Power Autom. Equip.* **2017**, *37*, 146–152.
8. Chen, X.; Li, K.; Xiao, J.; Meng, Q.; Cai, D. A method of real-Time power quality disturbance classification. *Trans. China Electrotech. Soc.* **2017**, *32*, 45–55.
9. Yin, B.; He, Y.; Zhu, Y. Detection and classification of power quality multi-disturbances based on generalized S-transform and fuzzy SOM neural net-work. *Proc. CSEE* **2015**, *35*, 866–872.
10. Huang, N.; Zhang, W.; Cai, G.; Xu, D. Power quality disturbances classification with improved multi-resolution fast S-transform. *Power Syst. Technol.* **2015**, *39*, 1412–1418.
11. Ahila, R.; Sadasivam, V.; Manimala, K. An integrated PSO for parameter determination and feature selection of ELM and its application in classification of power system disturbances. *Appl. Soft Comput.* **2015**, *32*, 23–37. [[CrossRef](#)]
12. Chen, H.; Zhang, G. Power quality disturbance identification using decision tree and support vector machine. *Power Syst. Technol.* **2013**, *37*, 1272–1278.
13. Khokhar, S.; Zin, A.A.M.; Memon, A.P.; Mokhtar, A.S. A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network. *Measurement* **2017**, *95*, 246–259. [[CrossRef](#)]
14. Liu, Z.; Cui, Y.; Li, W. A classification method for complex power quality disturbances using EEMD and rank wavelet SVM. *IEEE Trans. Smart Grid* **2015**, *6*, 1678–1685. [[CrossRef](#)]
15. Zhang, Q.; Liu, Z.; Zhu, L.; Zhang, Y. Recognition of multiple power quality disturbances using multi-label wavelet support vector machine. *Proc. CSEE* **2013**, *33*, 114–120.
16. Borges, F.A.S.; Ferañades, R.A.S.; Silva, I.N. Feature Extraction and Power Quality Disturbances Classification Using Smart Meters Signals. *IEEE Trans. Ind. Inform.* **2016**, *12*, 824–833. [[CrossRef](#)]
17. Wang, H.; Wang, P.; Liu, T.; Zhang, W. Power quality disturbance classification based on growing and pruning optimal RBF neural network. *Power Syst. Technol.* **2018**, *42*, 2408–2415.
18. Mohammadi, F.; Zheng, C.; Su, R. Fault Diagnosis in Smart Grid Based on Data-Driven Computational Methods. In Proceedings of the 5th International Conference on Applied Research in Electrical, Mechanical, and Mechatronics Engineering, Tehran, Iran, 24 January 2019.
19. Murat, U.; Selcuk, Y.; Muhsin, T.G. An effective wavelet-based feature extraction method for classification of power quality disturbance signals. *Electr. Power Syst. Res.* **2008**, *78*, 1747–1755.
20. Tang, J.; Deng, C.; Huang, G. Extreme learning machine for multilayer perceptron. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 809–821. [[CrossRef](#)]
21. Huang, G.; Zhou, H.; Ding, X.; Zhang, R. Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Trans. Syst. Man Cybern.* **2012**, *42*, 513–529. [[CrossRef](#)]
22. IEEE P 1159 Working Group. *IEEE Standard 1159-2009. Recommended Practice for Monitoring Electric Power Quality*; IEEE: Piscataway, NJ, USA, 2009.

23. IEEE Power Engineering Society. IEEE PES Working Group P1433 Power Quality Definitions. Available online: <http://grouper.ieee.org/groups/1159/2/testwave.html> (accessed on 2 February 2011).
24. IEEE Power Engineering Society. IEEE PES Working Group Power Quality Definitions. Available online: <http://grouper.ieee.org/groups/1159/3/docs.html> (accessed on 12 October 2012).
25. Luo, Y.; Li, K.; Li, Y.; Cai, D.; Zhao, C.; Meng, Q. Three-Layer Bayesian Network for Classification of Complex Power Quality Disturbances. *IEEE Trans. Ind. Inform.* **2018**, *14*, 3997–4006. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).