Received April 22, 2019, accepted May 4, 2019, date of publication May 7, 2019, date of current version May 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2915252

Mechanical Faults Diagnosis of High-Voltage Circuit Breaker via Hybrid Features and Integrated Extreme Learning Machine

WEI GAO[1,2], RONG-JONG WAI[2], (Senior Member, IEEE), SU-PENG QIAO[1], AND MOU-FA GUO[1]

[1]College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China
[2]Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taipei 106, Taiwan

Corresponding author: Rong-Jong Wai (rjwai@mail.ntust.edu.tw)

This work was financially supported in part by the Ministry of Science and Technology of Taiwan under Grant MOST 106-2221-E-011-028-MY2, and in part by the National Natural Science Foundation of China under Grant 51677030.

ABSTRACT As key electrical equipment in the power system, the normal operation of a high-voltage circuit breaker is related to the reliability and economy of the power supply. In this paper, a mechanical fault diagnostic method for a high-voltage circuit breaker via the hybrid feature extraction and the integrated extreme learning machine (IELM) is investigated. First, the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) is used to decompose the vibration signal to obtain intrinsic mode functions (IMF). Then, the sub-band reconstruction of each order IMF component is performed by combining the Hilbert transform and the band-pass filter in order to obtain the time-frequency matrix. Moreover, mechanical fault feature vectors can be formed by the time-frequency entropy and the singular entropy, which are extracted by transforming the time-frequency matrix into the energy matrix and normalizing the frequency bands via the normal cumulative distribution function (NCDF). In addition, an IELM is built for the fault classification. The advantages of the proposed CEEMDAN scheme in combination with band-pass filtering can eliminate the modal aliasing, reduce the number of auxiliary noise additions, and improve the decomposition efficiency. Besides, the performance of the singular entropy normalized by the NCDF is more stable, and the IELM composed of multiple weak classifiers can solve the shortcomings of the traditional extreme learning machine. The experimental results based on measured data show the proposed method can effectively diagnose the mechanical failure via small samples of high-voltage circuit breakers.

INDEX TERMS High-voltage circuit breaker, vibration signal, complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), time-frequency entropy, singular entropy, integrated extreme learning machine (IELM).

I. INTRODUCTION

With the development of economy, the requirements for the reliability and stability of a power system are constantly improving. As important electrical equipment in the power system, the high-voltage circuit breaker has dual functions of optimizing the network structure and protecting the safety of the power grid. Once the high-voltage circuit breaker fails, it may cause significant economic losses or even compromise system safety. Therefore, it is extremely important to accurately identify defects and faults of high-voltage circuit breakers.

The statistical analyses indicate that 80% of faults in high-voltage circuit breakers are caused by poor mechanical properties, and most of them are problems of operating mechanisms [1], [2]. The diagnostic and eliminative methods for traditional mechanical faults mainly rely on scheduled maintenance [3]. Unfortunately, scheduled maintenance often has the disadvantages of over-repair and waste of human resources. In addition, repeated disassembly and repair may cause mechanical structural changes and even bring new faults [3]. With the development of artificial intelligence (AI) technologies, the diagnostic methods for

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo.

mechanical faults have gradually combined with AI technologies. Among many emerging methods, vibration signals [3]–[14], contact stroke displacements [15], [16], and electromagnetic coil currents [17] are typical factors for diagnostic methods. The non-intrusive fault diagnostic method based on vibration signals has gradually become the mainstream research due to the convenient data acquisition.

The vibration signal of high-voltage circuit breakers usually exhibits non-stationary and non-linear behavior owing to the interaction between mechanical structures [5]. Therefore, the signal decomposition is always required to obtain the characteristics in the time-frequency domain before the feature extraction. General methods via time-frequency analyses mainly include the wavelet transform (WT) [18], the wavelet packet transform (WPT) [5]–[7], the empirical mode decomposition (EMD) [8], [19], the ensemble empirical mode decomposition (EEMD) [9], [10], [20], the variational mode decomposition (VMD) [11], [12], [21], etc. These algorithms have been widely used in the processing of vibration signals. The WT is a decomposition method based on a basis function, which is similar to the Fourier transform of a tunable window. As for the decomposition form, the selective range of basis functions in the WT is wider than that of the Fourier transform. However, the WT has a limitation in decomposing the high-frequency information. In order to improve this shortcoming, the WPT has been developed on the basis of the WT. The WPT can effectively decompose high-frequency and low-frequency information, and take the advantages of wavelet orthogonality, completeness and locality into account at the same time. However, both the WT and the WPT belong to non-adaptive decomposition methods. Before the decomposition, the basis function needs to be determined, and the decomposition result is greatly affected by the basis function. Besides, there are some unavoidable defects between them, such as energy leakage, interference problems, and so on. The EMD is a modal component-based decomposition method to adaptively decompose complex signals into a series of intrinsic mode functions (IMF) containing local feature information. Liu *et al.* [8] adopted the EMD to decompose the vibration signal of a high-voltage circuit breaker, obtained the envelope of each order IMF components by the Hilbert transform, and extracted the envelope energy entropy as the features. The extracted features have obvious recognition in identifying normal and faulty signals. But, it also lacks sufficient data for verification, and the performance of the method in [8] is not further analyzed. Although the EMD is an adaptive decomposition method, it also has certain limitations. For example, the recursive decomposition method causes the envelope estimation error to be transmitted all the time, and results in the modal aliasing, which is hard to accurately separate components with similar frequencies. For the phenomenon of the modal aliasing, the EEMD is derived from the EMD. The EEMD increases continuity to different scales by adding auxiliary noise to the original signal, which suppresses the modal aliasing to some extent. Unfortunately, there are often residual noise components in the reconstructed signal, which directly affects the accurate analysis of the signal. Unlike recursive modal decomposition methods (such as EMD and EEMD), the VMD can eliminate the modal aliasing phenomenon of the EMD. In addition, the components of the VMD have limited bandwidth, and its center frequency can achieve online estimation. However, there are certain deficiencies in the VMD [21]. It is difficult to select the modal number and modal frequency bandwidth control parameters. The number of modalities usually needs to be set manually, and is often selected by the central frequency observation method to be greatly influenced by subjective factors. The vibration waveform of the high-voltage circuit breaker is too complicated, and it is difficult to obtain a suitable modal number by the center frequency observation method. In addition, the selection of the modal frequency bandwidth control parameters will affect the ability to consistently interfere with noise.

The complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [22]–[25] is proposed to solve these problems. Compared to the EEMD [26], the CEEMDAN obtains the IMF component by adding adaptive auxiliary noise and calculating the unique margin. As a result, it can maximize the zero reconstruction error of the decomposed signal. Because the CEEMDAN eliminates the modal aliasing with fewer times of additive noise, it greatly reduces the cost of computing resources. Lv *et al.* [23] proposed a rolling bearing fault diagnostic method based on the CEEMDAN and improved the multivariate multi-scale sample entropy (MMSE). From the results in [23], the CEEMDAN performs better than the EEMD in mitigating modal aliasing, which can extract the characteristic frequency of the faults more accurately. In addition, Ren *et al.* [24] studied the performance of the EMD, the EEMD, the complementary EEMD (CEEMD) and the CEEMDAN in wind speed prediction, and combined these algorithms with artificial neural networks and support vector regression (SVR) to predict the wind speed in time series. From the comparisons in [24], the CEEMDAN-SVR performs best than all other methods. In view of the above analyses, this study uses the CEEMDAN as the time-frequency analysis method for the vibration signal of a high-voltage circuit breaker.

Among the feature extraction methods for vibration signals of high-voltage circuit breakers, the entropy has been widely used as the characteristic attribute [27]. Huang *et al.* [4] selected the WT to decompose the signal according to the vibration signal energy distribution reflected by the S-transform, constructed the time-frequency matrix through the module operation, and converted it into an energy matrix. Moreover, the energy entropy from the time domain and the frequency domain was extracted to construct the time-frequency entropy, and the one-class support vector machine (OCSVM) was used to determine whether there are mechanical failures or not. He *et al.* [13] proposed a new method for the fault identification of high-voltage circuit breakers based on the density peak clustering, the kernel fuzzy C-means clustering and the support vector machine (SVM).

First, the local mean decomposition was used to decompose the sound and vibration signals, and the maximum original signal correlation law was used to filter the waveform. Then, the product functions were quantized by the approximate entropy, and the entropy value was taken as the characteristic of the fault signal. The result in [13] showed that the feature recognition accuracy extracted by double clustering is significantly higher than that of a single clustering method. Although the features extracted by the entropy as a characteristic attribute have certain characterization ability, the influence of disturbance and noise cannot be avoided, especially the statistic that directly quantizes the time series, such as the sample entropy, the permutation entropy, and the approximate entropy. The time-frequency entropy [4], [14] was composed of the time-energy entropy and the frequency-energy entropy, which takes the block energy as the analytic object and mainly characterizes the energy distribution of each frequency-band or period. Compared with the direct quantization of the time-series complexity, the influence of disturbance and noise is weak. In view of good characterization ability of the time-frequency entropy, the time-frequency entropy can be chosen for extracting the mechanical vibration signal of a high-voltage circuit breaker. Since the existence of disturbance and noise, the characterization ability of the time-frequency entropy also has some certain limitations. On this basis, another feature extraction method, namely the singular entropy, is adopted to solve the problem of insufficient representation ability of single type features. Although the singular value decomposition [28], [29] has strong robustness against noise and interference in the background, excessive influence value will affect the stability of singular values. In recent years, the normal cumulative distribution function (NCDF) is always used to process the signal before the feature extraction, and weaken the influence of the impact maximum values on the singular value decomposition [30]. The singular entropy takes the singular value as the analytic object, which can achieve both dimensionality reduction and the energy distribution characteristics of the time-frequency matrix.

After the feature extraction, choosing a classifier with good performance is a key issue. Traditional classification methods include the back-propagation neural network (BPNN) [6], [7], the support vector machine (SVM) [4], [11], the K-nearest neighbor method (KNN) [31], [32], the decision tree (DT) [33], [34], etc. Among them, the BPNN has good anti-noise and generalization ability [35]. But, its classification performance depends on a large number of training samples, and it is prone to local convergence. The SVM is suitable for the classification problem of small samples with nonlinear and high-dimensional features. Due to the structure of binary tree classification, the training time of the SVM is longer [10]. Besides, the classification results are greatly affected by the parameters. As for the KNN method, the parameter estimation and training process are not required so that it is easy to implement. However, the selection of the number of the nearest neighbors and the choice of distance will affect the final recognition result. Although the DT is easy to understand and implement, it suffers from the overfitting and insufficient generalization problems.

Intelligent diagnostic strategies including fuzzy theory and neural network have been commonly used for fault diagnoses [36], [37]. Song *et al.* [36] used statistical tests in the frequency domain to sequentially extract features for diagnosing a number of anomalous states from the measured signals, and combined the trivalent logic inference diagnosis theory with the possibility and fuzzy theory to establish a stepwise fuzzy diagnosis method for diagnosing faults in rotating mechanical structures. However, there are difficulties in obtaining fuzzy diagnostic knowledge and relying too much on fuzzy knowledge bases. Wang *et al.* [37] proposed a novel fault recognition method for rotating machinery on the basis of multi-sensor data fusion and bottleneck layer optimized convolutional neural network (MB-CNN). Unfortunately, the computation burden for the MB-CNN is a latent problem.

The extreme learning machine (ELM) [38], [39] is a new fast learning algorithm that avoids the problem of falling into a local optimum. Because the weights and offsets between the input layer and the hidden layer of the ELM are randomly generated, the generated models are inconsistent each time, and the classification effect is also different. In view of the above analyses, the integrated extreme learning machine (IELM) is investigated in this study. The IELM consists of a group of ELM weak classifiers generated by small samples, which not only includes the advantages of rapidity and global in the ELM, but also applies the ELM with large sample classification to the classification problem of small samples.

In this study, a novel mechanical fault diagnostic method for high-voltage circuit breakers is investigated. First, the CEEMDAN is applied to the time-frequency analysis of the vibration signal. Moreover, the sub-band reconstruction of the decomposed components is performed by combining the Hilbert transform and the band-pass filter in order to analyze the signal at the same scale. In addition, the time-frequency entropy from the energy matrix and the singular entropy normalized by the NCDF are extracted, and the characteristics of this diversity are sufficient to enhance the characterization ability of the original signal. Furthermore, the IELM is trained by the hybrid features, and is used to classify small fault samples of high-voltage circuit breakers.

This study is organized as follows. Following the Introduction, Section II introduces the principle of the CEEMDAN algorithm. Section III investigates the methods of the time-frequency entropy and the singular entropy extraction. Section IV explains the structure of the IELM. Section V analyzes and discusses the experimental process and results. Section VI summarizes the methods presented in this study.

II. WAVEFORM DECOMPOSITION BASED ON CEEMDAN

In the CEEMDAN algorithm, the operator $E_j(\cdot)$ is defined to obtain the j^{th}-order IMF component through the EMD. $\omega_i(t)$ is the Gaussian white noise obeying the N (1,0) distribution,

and ε_i is the coefficient for controlling the signal-to-noise ratio of the auxiliary noise and the original signal. Given the signal $x(t)$ to be processed, the CEEMDAN algorithm [22] can be described as follow:

(i) The first IMF component can be obtained by the EMD decomposition of $x(t)$ plus $\varepsilon_0\omega_i(t)$.

$$c_1(t) = \frac{1}{I}\sum_{i=1}^{I} E_1(x(t) + \varepsilon_0\omega_i(t)) \tag{1}$$

where $i = 1, 2, \ldots, I$, in which I is the adding times of the auxiliary noise.

(ii) The first margin can be calculated by

$$r_1(t) = x(t) - c_1(t) \tag{2}$$

(iii) The second IMF component can be obtained by the EMD decomposition of $r_1(t)$ plus $\varepsilon_1 E_1(\omega_i(t))$, that is,

$$c_2(t) = \frac{1}{I}\sum_{i=1}^{I} E_1(r_1(t) + \varepsilon_1 E_1(\omega_i(t))) \tag{3}$$

(iv) Similar to the steps in (ii)-(iii), the $(k+1)^{th}$ IMF component can be expressed as

$$c_{k+1}(t) = \frac{1}{I}\sum_{i=1}^{I} E_1(r_k(t) + \varepsilon_k E_k(\omega_i(t))) \tag{4}$$

(v) Until the margin satisfies the termination condition of the residual component, the original signal $x(t)$ can be decomposed into

$$x(t) = \sum_{k=1}^{K} c_k(t) + r(t) \tag{5}$$

where K is the total order of the IMF components, and $r(t)$ is the residual component.

III. FAULT FEATURE EXTRACTION

A. WAVEFORM RECONSTRUCTION BASED ON HILBERT TRANSFORM AND BAND-PASS FILTERING

When the CEEMDAN is performed on the same type of vibration signals, the number of acquired components may be different, or the bandwidths of the components with respect to the same order may also be different. If features are directly extracted from these components, there is no uniform specification. In other words, the features extracted from the same type of vibration signals in the same order are different, and it is difficult to extract the general feature information for affecting the later diagnostic design. In this study, a waveform reconstruction method based on the Hilbert transform and the band-pass filter are used to reconstruct the decomposed waveform into each frequency band, extract the general features from the signals in the same category, and investigate the difference between different-type signals. The process of waveform reconstruction and band-pass filtering is expressed as follows:

(i) First, a Hilbert transform [19] is performed on each order IMF component $c_k(t)$:

$$H[c_k(t)] = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{c_k(\tau)}{t-\tau}\mathrm{d}\tau \tag{6}$$

Then, a parsing signal can be constructed by

$$z_k(t) = c_k(t) + \mathrm{j}H[c_k(t)] = \alpha_k(t)e^{\mathrm{j}\phi_k(t)} \tag{7}$$

Thus, the corresponding phase function can be expressed as

$$\phi_k(t) = \arctan\frac{H[c_k(t)]}{c_k(t)} \tag{8}$$

In addition, the instantaneous frequency of the phase function can be calculated by

$$f_k(t) = \frac{1}{2\pi} \times \frac{\mathrm{d}\phi_k(t)}{\mathrm{d}t} \tag{9}$$

(ii) It is assumed that the number of pre-divided bands is M. A reconstructed signal of the $(F_{m-1}, F_m)|_{m=1,2,\ldots,M}$ band can be obtained by traversing each IMF component $c_k(t)$. When the instantaneous frequency with respect to each data point in $c_k(t)$ is within this frequency band, the data point is maintained; otherwise, it will be cleared. After that, a new sequence can be formed. By accumulating all new sequences $(D(t) = \sum_{k=1}^{K} \widehat{c}_k(t))$ to obtain the waveforms in the $(F_{m-1}, F_m)|_{m=1,2,\ldots,M}$ band, the original waveform can be decomposed into each of the specified frequency bands.

The division of IMF components in different frequency bands is convenient by analyzing the changes of vibration signals at the same scale. Based on the normative time-frequency matrix, it becomes meaningful to seek commonality or characteristics. Since different types of vibration signals often differ in the same frequency bands, the method of waveform reconstruction described herein is very beneficial for signal analysis.

B. HYBRID FEATURE EXTRACTION BASED ON ENTROPY

The vibration signal of a high-voltage circuit breaker is the result of the composite vibration of multiple mechanical components. From the perspective of the frequency domain, it is the superposition of multi-frequency characteristic information. But, it is an attenuation process of multiple vibration energy from the perspective of the time domain. When the mechanical structure changes, the path of the vibration energy transfer may change to accordingly cause the amplitude and frequency variation of each mechanical component. As a result, the amplitude and frequency of the composite vibration signal also change. In order to analyze the vibration signal accurately, it is necessary to extract the features from the time domain and the frequency domain. From the concept of entropy, the degree of irregularity of a system or time series can be quantified by the probability distribution of its state. Therefore, two entropy features, which can be used to characterize the mechanical state of a high-voltage circuit breaker, are adopted in this study.

The time-frequency entropy [4], [14] takes the energy matrix as the analytic object and statistically calculates the energy distribution for each time period and each frequency band, respectively. Whether it is the distribution of the energy probability of each frequency band or each time period, it is directly related to the path of the vibration energy transmission. Different energy transfer paths will cause changes in the amplitude and frequency of each mechanical component. In other words, the amplitude variation will cause a change in the energy probability distribution for each time period, and the frequency variation will cause a change in the energy probability distribution of each frequency band. Therefore, the time-frequency entropy is used to find the energy distribution characteristics of the vibration signals at different mechanical states in the time-frequency domain.

The singular entropy uses the time-frequency matrix at each time period as the analytic object to describe the distribution statistics of the singular values of the time-frequency matrix. The singular value decomposition can extract the feature of signal stability under the occurrence of disturbance and noise, and the distribution of amplitude from large to small in order. Because the degree of decline of singular value distribution in different time periods is different, the probability distribution of the singular value is also different. Thus, the singular entropy is extracted as the other feature of a high-voltage circuit breaker in this study.

1) TIME-FREQUENCY ENTROPY

In the frequency domain, the reconstructed waveform is divided into M sub-waveforms at different frequency bands to form a time-frequency matrix of $M \times N$ size. In the time domain, the time-frequency matrix is equally divided into L time periods to form a $M \times L$ block time-frequency matrix, wherein each block matrix has a length N/L. Let $D_l^m(t)(l = 1, 2, \ldots, L; m = 1, 2, \ldots, M)$ be the time series corresponding to the l^{th} period in the m^{th} frequency band, and the calculation formula of its energy value $E_{l,m}$ can be represented as

$$E_{l,m} = \sqrt{\sum_{t=1}^{N/L} D_l^m(t)^2} \tag{10}$$

From (10), the block time-frequency matrix can be converted into the block energy matrix. Then, the energy matrix is normalized according to the following conversion formula:

$$P_{l,m} = E_{l,m}/E \tag{11}$$

where E is the total energy value of the matrix; $P_{l,m}$ is the proportion of $E_{l,m}$ in E.

As an indicator to evaluate the complexity of time series, the Shannon entropy is often used to characterize the uncertainty of the signal. The greater the uncertainty, the larger the entropy, and the greater the amount of information needs to clarify it. The Shannon entropy Θ can be calculated as

$$\Theta = -\sum_{i=1}^{R} \varsigma_i \log \varsigma_i \tag{12}$$

In this study, the Shannon entropy is used as a characteristic attribute. The time-frequency entropy features are extracted in the energy matrix from the time-frequency domain direction. The specific calculation formulas can be expressed as follows:

$$\begin{aligned} W_l^t &= -\sum_{k=1}^{M} (P_{l,k} \times \log(P_{l,k})) \\ W_m^f &= -\sum_{k=1}^{L} (P_{k,m} \times \log(P_{k,m})) \end{aligned} \tag{13}$$

In (13), W_l^t is a characteristic value with respect to the l^{th} time period in the time domain, and reflects the distribution of energy in different frequency bands during this period, which is simply referred to as the time-domain entropy (TE). W_m^f is a characteristic value with respect to the m^{th} frequency band in the frequency domain, and reflects the distribution of energy in different time periods during this band, which is simply referred to as the frequency-domain entropy (FE). $W = [W_1^t, W_2^t, \ldots, W_L^t, W_1^f, W_2^f, \ldots, W_M^f]$ is the total feature set, which is the time-frequency entropy (TFE) described herein. It can be seen from the form of the feature vector that the TFE are composed of the TE and the FE.

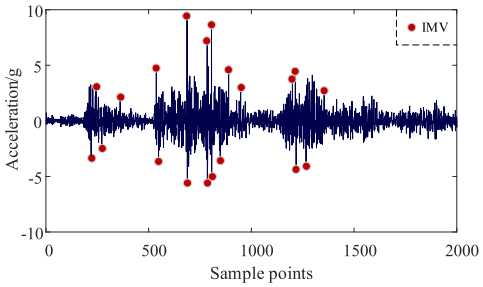


FIGURE 1. Impact maximum value of vibration signal.

2) NCDF-SINGULAR ENTROPY

The vibration signal generated from the operation of the high-voltage circuit breaker has a certain amount of non-periodic impact maximum values (IMV) as shown in Fig. 1. These values are caused by the strong collision between mechanical components and the influence of interference factors. Thus, the amplitudes of vibration signals do not decrease synchronously with the attenuation of vibration energy. These values are sensitive to the singular value decomposition, and it is difficult to obtain stable singular value characteristics by the singular value decomposition of the time-frequency matrix directly. Therefore, the NCDF is introduced in this study to conduct the standardized processing of vibration signals at all frequency bands for weakening the impact of

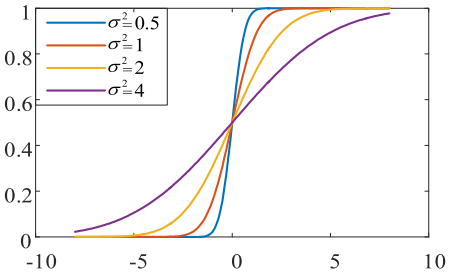


FIGURE 2. NCDF distribution with respect to different variances.

impact maximum value on the singular value decomposition and improving the stability of singular value characteristics.

The NCDF is an integral of a probability density function obeying a normal distribution and can fully describe the probability distribution of the random variable X. It can be expressed as

$$F(x;\mu;\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{(t-\mu)^2}{2\sigma^2})dt \qquad (14)$$

where μ, σ and σ^2 are the mean, the standard deviation, and the variance of the random variable X, respectively. Figure 2 shows the NCDF distribution for different variances when the mean is zero. It is obvious that the NCDF curves with respect to different magnitudes of σ^2 are steep in the middle and smooth at both ends, and the steepness is related to the magnitude of σ^2. The reasonable parameter selection can effectively distribute the impact maximum values to the two poles of the curve, which essentially highlights the overall distribution of the vibration signal, thereby reducing the impact of the impact maximum values. After completing the normalization processing by the NCDF for each frequency band, the steps of the singular entropy extraction are explained as follows.

Suppose $\boldsymbol{D}$ is a matrix of $a \times b$ size according to the singular value decomposition in [28], for any matrix with $a \times b$ dimension, there is always a matrix $\boldsymbol{U}$ with $a \times c$ dimension, a matrix $\boldsymbol{V}$ with $c \times b$ dimension, a diagonal matrix Λ with $c \times c$ dimension, so that the matrix $\boldsymbol{D}$ can be expressed as

$$\boldsymbol{D} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T \qquad (15)$$

Among them, the diagonal element $\lambda_k (k = 1, 2, \ldots, c)$ in the matrix Λ is called the singular value of the matrix $\boldsymbol{D}$. Singular values are non-negative and are arranged in a decreasing manner. Then, the singular values can be normalized by

$$p_k = \lambda_k / \sum_{k=1}^{c} \lambda_k \qquad (16)$$

where p_k is the proportion of λ_k in the total singular value. In addition, the singular entropy (SE) can be defined as

$$S_E = -\sum_{k=1}^{c} p_k \log p_k \qquad (17)$$

In order to obtain finer features, the time-frequency matrix consisting of reconstructed waveforms needs to be divided. For a time-frequency matrix of $M \times N$ dimension, it can be equally divided into G block in the time direction, wherein each block matrix has the dimension of $M \times (N/G)$. The SE feature vector can be extracted from each block time-frequency matrix.

The NCDF-singular entropy of the high-voltage circuit breaker vibration signal is implemented by the following steps: 1) Perform the CEEMDAN decomposition on the vibration signals, and then reconstruct the waveform; 2) Normalize the reconstructed waveform with the NCDF; 3) Perform the singular value decomposition and calculate the singular entropy.

IV. INTEGRATED EXTREME LEARNING MACHINE

The ELM is a feedforward neural network with single hidden layer and random learning [38]. It has the characteristics of randomly generating input weights and determining output weights by analyses. Since the input weights are randomly assigned, the ELM does not need to adjust the network parameters in a learning process, and requires less computational memory than traditional learning algorithms.

The ELM consists of an input layer, a hidden layer, and an output layer. Given the training set Z_N as

$$Z_N = \left\{ (x_j, y_j) \,|x_j \in \mathrm{R}^n, y_j \in \mathrm{R}^w \right\}_{j=1}^{J} \qquad (18)$$

where x_j is the input vector of $n \times 1$ size; y_j is the target vector of $w \times 1$ size, and J is the total number of samples. The output function of an ELM with q nodes in the hidden layer can be expressed as

$$f(x_j) = \sum_{i=1}^{q} \beta_i \cdot \varphi(\alpha_i \cdot x_j + g_j) = y_j, \quad j = 1, 2, \ldots, J \qquad (19)$$

where φ is the activation function; α_i is the weight of the connection between the i^{th} hidden layer node and the input node; β_i is the connection weight between the i^{th} hidden layer node and the output node; g_i is the bias of the i^{th} hidden layer node. By the principle of the ELM, the output of the ELM can be represented as

$$H\beta = Y \qquad (20)$$

where H is the output matrix of the hidden layer. The solution of β can be obtained by calculating the least squares solution of (20) as

$$\min_{\beta} \|H\beta - Y\| \qquad (21)$$

From (21), one can obtain the solution as

$$\hat{\beta} = H^{\dagger}Y \qquad (22)$$

where $H^{\dagger}$ is the Moore-Penrose generalized inverse matrix of H.

The use of the ELM as a classifier for the mechanical fault diagnostic model of high-voltage circuit breakers has two disadvantages. Because the weights and the biases are

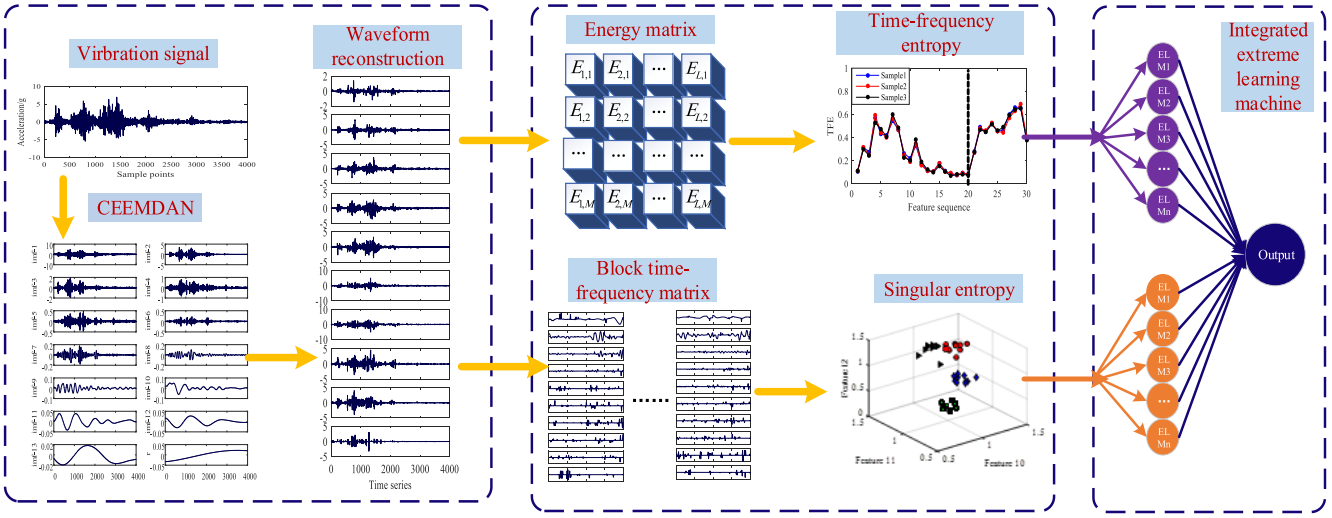


FIGURE 3. Framework of proposed mechanical fault diagnostic method.

randomly generated, the network generated by each training process is inconsistent, and its performance is also significantly different. Moreover, for each generated model, the evaluation optimal criteria are insufficient. In addition, the training of the ELM network requires a large number of samples. Unfortunately, the mechanical failure of the high-voltage circuit breaker belongs to a small sample event, so the result generated by the ELM classifier will be weak. In view of the shortcomings of a single ELM, an integrated extreme learning machine (IELM) is proposed in this study. The IELM is composed of many weak classifiers generated by a group of ELMs. The process of implementing decision classification can be expressed as follows.

For an ELM with multiple output nodes, it usually makes the category with the highest output value as the final output label. For example, given a test data x, $f(x) = [f_1(x), f_2(x), \ldots, f_w(x)]$ is the value of each output node of the ELM, where w is the number of output nodes and $f_i(.)$ is the functional model of the ELM. For the convenient description, the value of $f(x)$ is an integer within the range from 1 to w. The decision function of the ELM with multiple output nodes can be expressed as

$$\text{label}(x) = \arg\max_i f_i(x), \quad i \in [1, 2, \ldots, w] \tag{23}$$

where arg(.) is the value function for obtaining the independent variable, and the right side of (23) indicates the category label with respect to the maximum situation.

The IELM is composed of a set of ELMs, whose decision functions can be represented as

$$\text{Label}(x) = \arg\max_i \text{sum}(i), \quad i \in [1, 2, \ldots, w] \tag{24}$$

where sum(i) represents the number of sub-classifiers whose prediction result is i.

In this study, a mechanical fault diagnostic method for high-voltage circuit breakers via the hybrid feature extraction and the IELM is proposed. The framework of the proposed method is depicted in Fig. 3.

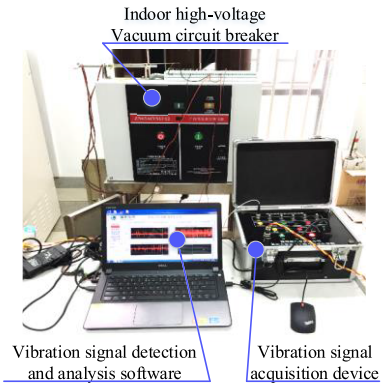


FIGURE 4. Photograph of experimental setup for data acquisition.

V. EXPERIMENTAL ANALYSIS AND RESULTS

A. EXPERIMENTAL PLATFORM INSTRUCTION

An indoor vacuum circuit breaker (ZN63A-12) is used in this experiment, and the photograph of the experimental setup for data collection is depicted in Fig. 4. The sensor adopts the piezoelectric acceleration sensor (LC0102T model), whose sensitivity, measurable range, and the frequency response range is 10.06 mV/g, ±500 g, from 2 Hz to 13,000 Hz, respectively. The multi-function equipment (USB-6211) manufactured by the NI corporation is used as the data acquisition unit, and its update rate is 250kS/s.

Four actual mechanical states are simulated on this platform, and the corresponding data acquisition is performed. In this data set, the data at the normal state, the base-loosen state, the C-phase leading-closing state, and the C-phase hysteresis-closing state are recorded as from Class 1 to

Class 4, respectively. Classes 3 and 4 belong to the asynchronous closing failure. In addition, in order to verify whether the proposed method is valid in practical field, the experimental data can be obtained under the background interference, such as the circuit breaker activity, passing people or passing the vehicle.

B. FEATURE EXTRACTION FROM VIBRATION SIGNALS

In this study, the vibration signal of each mechanical state is collected at a sampling frequency of 20 kHz, and the total sampling time is 0.5 s. In order to alleviate the influence of the zero drift of the sensor, the vibration signal needs to be intercepted before the feature extraction. The waveform 10ms before startup and 190ms after startup (4000 points in total) are taken as valuable data information.

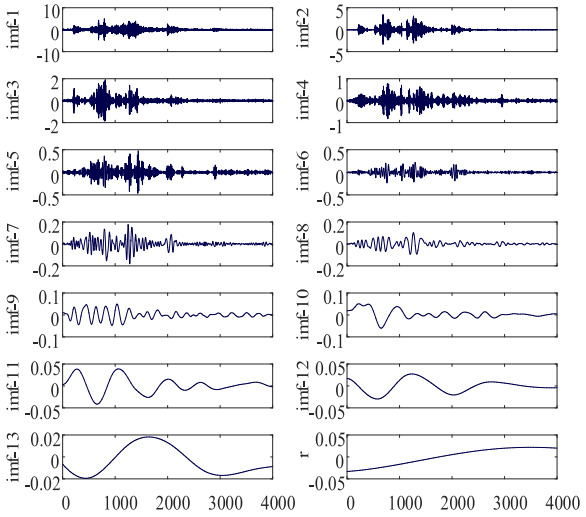


FIGURE 5. CEEMDAN decomposition with respect to each order component at base-loosen state.

The signal is decomposed by the CEEMDAN, and the added Gaussian white noise is 0.2 times the amplitude of the original signal standard deviation, and the adding time is 20. For example, Figure 5 is the IMF components and the residual component extracted by the vibration signal decomposition at the base-loosen state, where the horizontal axis is a time series with 4000 data points, and the vertical axis is the amplitude of the vibration signal. It is obvious that the amplitudes of the IMF components after the sixth order are small.

TABLE 1. Correlation between different-order reconstructed signals and original signals.

Order	1st	2nd	3rd	4th	5th	6th	7th
Class 1	0.819	0.921	0.976	0.990	0.994	**0.997**	0.998
Class 2	0.769	0.901	0.960	0.989	0.995	**0.997**	0.999
Class 3	0.797	0.911	0.975	0.997	0.999	**0.999**	0.999
Class 4	0.794	0.921	0.979	0.991	0.996	**0.998**	0.999

Progressive reconstruction is carried out for each order component signal successively, and the correlation with the original signal is calculated to obtain the results in Table 1. As can be seen from Fig. 5, the amplitude of the 7th IMF waveform is significantly reduced. From Table 1, the correlation between the reconstructed signal of the first 6-order IMF component and the original signal reaches 0.997, and the signal can be almost completely described. Therefore, this study selects the first 6-order of IMF components for subsequent analyses to reduce the effects of low-frequency spurious waveforms.

As for dividing the frequency band, a total of 10 frequency bands are allocated, and the bandwidth is 1000 Hz, i.e., the frequency band is from 0 Hz to 1000 Hz, from 1,000 Hz to 2,000 Hz, …, from 9,000 Hz to 10,000 Hz, respectively. In order to extract the time-frequency entropy, each frequency band of data is equally divided into twenty sets to form 10×20 block time-frequency matrices, where each block matrix has a size of 1×200. It should be pointed out that if the number of the block is too small, it is easy to cover up important information, and the extracted features are unrepresentative. However, if the number of the block is too large, the important information is dispersed and the corresponding features are difficult to extract. The number of the block selected in this study is the best result through multiple experimental trials.

Three samples are randomly selected for each mechanical state, and the TFE is extracted. As a result, a two-dimensional feature map is drawn in Fig. 6. The first 20 features are the time domain entropy and the last 10 features are the frequency domain entropy. It is obviously shown that the states of C-phase leading-closing and hysteresis-closing are significantly different from the other two states in the time domain characteristics. In particular, the difference between the C-phase hysteresis-closing state and the other three states is mainly reflected in the features from 2^{nd} to 11^{th} features. However, the difference between the C-phase leading-closing state and the normal state or the base-loosen state is reflected in the features at 7^{th} and 11^{th} feature sequences. Since the normal state and the base-loosen state are only slightly different, the features at the time domain are not sufficient to distinguish them. Similarly, although there is a slight difference between four mechanical states from 21^{st} to 25^{th} features, it is not obvious in the frequency domain. It can be seen that the discrimination of the time-frequency entropy features of each mechanical state depends only on a few features, and the vibration energy may change randomly under the background of noise and disturbance, resulting in the instability of these features based on the energy matrix extraction. Therefore, it is very significant to use a variety of hybrid features to enhance the ability to characterize signals.

Before the singular entropy feature is extracted, the NCDF function is used to normalize each frequency band in order to weaken the influence of the impact maxima values on the singular value decomposition. The value of the variance (σ^2) determines the degree of standardization. If the value of σ^2 is too small, the NCDF curve is too steep and most of the information will be mapped at both ends of the curve. The details of the signal will be submerged, even small noise signals will be amplified. If the value of σ^2 is too large, the NCDF

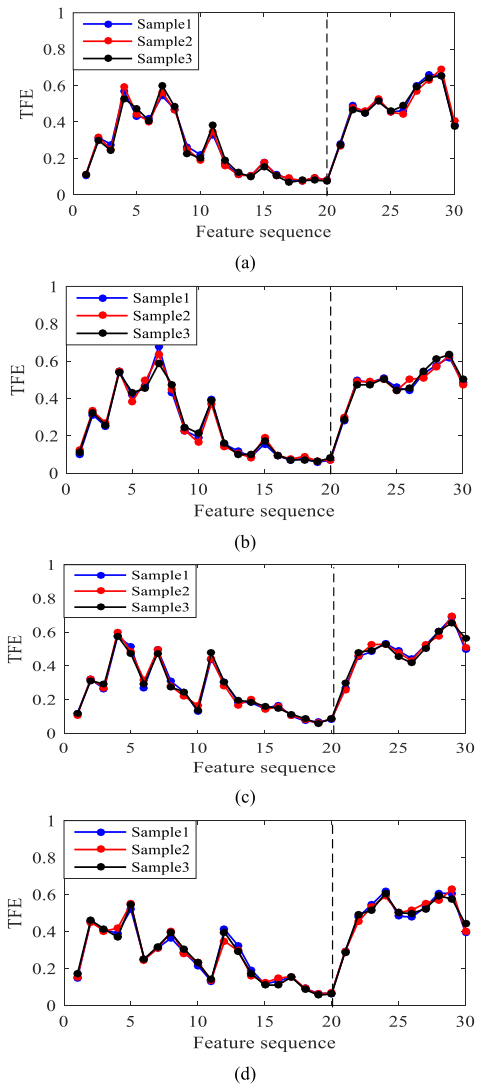


FIGURE 6. Time-frequency entropy with respect to four different mechanical states: (a) Normal state; (b) Base-loosen state; (c) C-phase leading-closing state; (d) C-phase hysteresis-closing state.

curve is too flat. Most of the information will distribute in the middle of the gradual area to be approximately linearity, and cannot achieve the effect of weakening the impact signal. For example, there is a comparison of signals at the Class 1 state before and after the normalization by the NCDF in Fig. 7(a)∼(d). In Fig. 7(a), the first 2000 data points of the signal are subjected to the standardization, and in the different range of σ^2 they are respectively mapped into the range from 0 to 1 as shown in Fig. 7(b), 7(c), and 7(d). In Fig. 7(b), most of the data are mapped to 0 and 1, and the details of the signal are almost completely submerged. In Fig. 7(c) and 7(d), as the value of σ^2 increases, the standardized signal tends to the linear mapping of the original signal. As for $\sigma^2 = 2$, the normalized signals of the other three states (Class 2, Class 3 and Class 4) are depicted in Fig. 7(e), 7(f) and 7(g), respectively. Obviously, the impact maximum values for the signal standardized by the NCDF can be significantly reduced, which is beneficial for the singular value decomposition. In order to achieve the effect of weakening the impact value, this study compares the results of many experiments, and finally one sets the value of σ^2 to be 2.

In the time domain, 100 data points are divided into 40 time slots to form 40 time-frequency matrixes, wherein each block matrix has a size of 10×100, then the SE of each block matrix can be obtained. The singular entropy characteristics of three groups of vibration signal samples at the C-phase leading-closing state with or without the NCDF standardization is depicted in Fig. 8. For the case without normalization, all the singular entropy of the same type of vibration signal tends to the same value, and there is no significant difference in the singular entropy at different periods. After the NCDF standardization, not only the singular entropy of the same period approaches the same value, but also the characteristic differences in different periods are particularly prominent.

In order to visualize the singular entropy characteristics, the partial singular entropy extracted from vibration signals of four different mechanical states in three-dimensional coordinates is depicted in Fig. 9, where Fig. 9(a), 9(b), 9(c) and 9(d) describe the characteristic distribution of the singular entropy interval of 10^{th} - 12^{th}, 23^{th} - 25^{th}, 26^{th} - 28^{th} and 30^{th} - 32^{th} for different cases, respectively; 10 samples are randomly selected for each case. For the interval shown in Fig. 9(b), although there is a certain difference between Class 1 and the other three classes, this difference is too small to be difficult for accurately classifying. If it is combined with the interval represented by Fig. 9(a) and 9(c) to form complete information, the feature differences between Class 1, Class 3 and Class 4 are significantly increased. The set of these multi-interval features is very effective for the expression of fault features in vibration signals.

C. ANALYSES OF RESULTS

1) ELM PARAMETER SELECTION

In order to enhance the classification performance of a single ELM, the number of hidden layer neurons and the type of activation functions need to be determined before the classifier is trained. Moreover, as for the selection of the number of neurons in the hidden layer and the type of activation functions, the classifiers of two feature types should be discussed separately. Since the mechanical failure of the circuit breaker is a small probability event, this study collects 50 samples for each mechanical state, i.e. the total number of samples is 200. Among them, 76% of the samples are used to train the classifier model, and 24% of the samples are used to test the accuracy of the classification.

The K-fold cross-validation method is applied herein to obtain a combination of parameters that makes the ELM classification ability more robust. Note that, the trained and validated samples are from the same 152 samples. In this study, the K value is set to 10, which is a typical 10-fold cross-validation method. With the changes of the number of neurons, the functions of sigmoidal, sinusoidal hardlim, rectified linear units (Relu), Tanh and improved sigmoidal (Isigmoid) [40] are respectively used as activation functions.

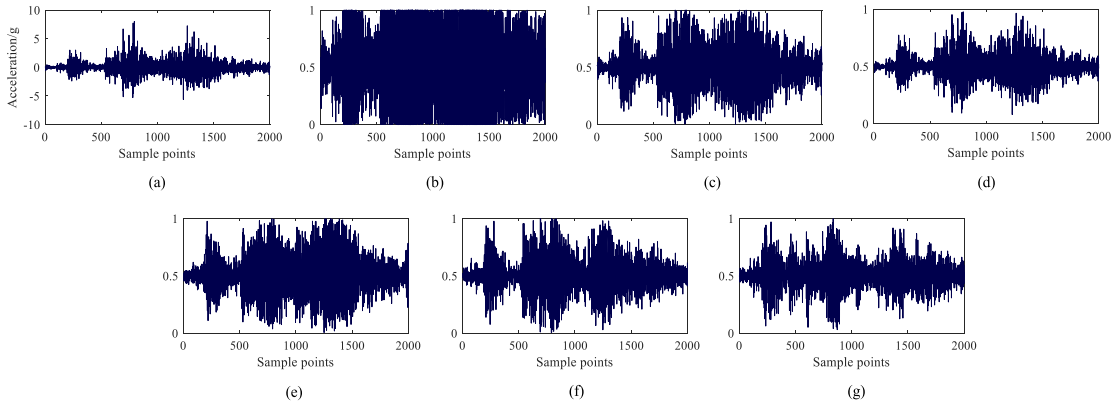


FIGURE 7. NCDF normalized signal for different state vibration signals: (a) Original signal of Class 1; (b) Normalized signal as $\sigma^2 = 0.5$ of Class 1; (c) Normalized signal as $\sigma^2 = 2$ of Class 1; (d) Normalized signal as $\sigma^2 = 4$ of Class 1; (e) Normalized signal as $\sigma^2 = 2$ of Class 2; (f) Normalized signal as $\sigma^2 = 2$ of Class 3; (g) Normalized signal as $\sigma^2 = 2$ of Class 4.

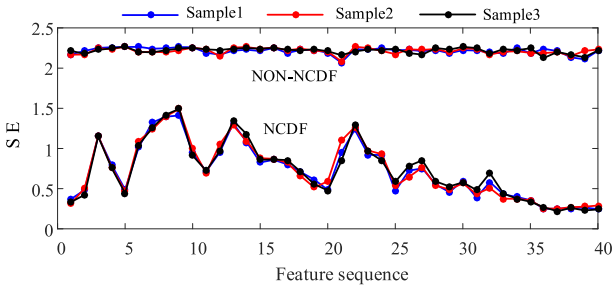


FIGURE 8. Comparison of singular entropy with or without NCDF normalization.

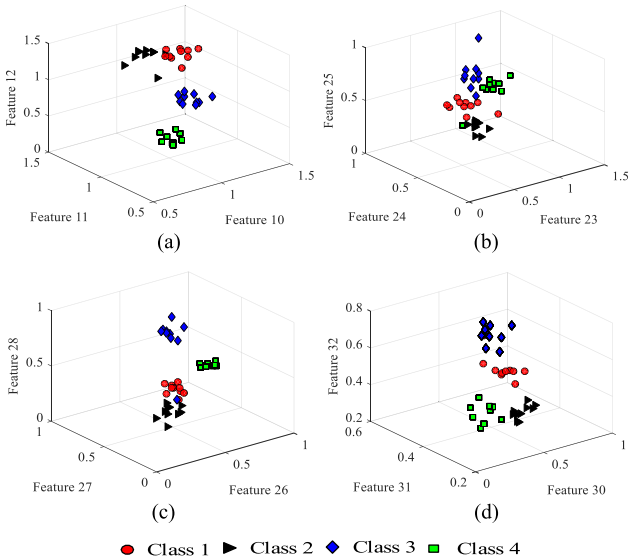


FIGURE 9. Partial singular entropy with respect to different mechanical states (a) features of 10^{th}-12^{th}; (b) features of 23^{th}-25^{th}; (c) features of 26^{th}-28^{th}; (d) features of 30^{th}-32^{th}.

The average accuracy calculated by the cross-validation method is used as an indicator for evaluating each combination of parameters.

The average accuracies of TFE and SE characteristics in different neuron numbers and activation functions are summarized in Tables 2 and 3, respectively. When 50 neurons are used in combination with the sinusoidal function, the ELM trained by TFE or SE can achieve slightly higher classification accuracies than other parameter combinations. Thus, they are selected as the parameters of each sub-classifier in subsequent experiments.

TABLE 2. Accuracies of TFE features under different neurons and activation functions.

Number	Sigmoid	Sin	Hardlim	Relu	Tanh	Isigmoid
50	90.7%	**94.7%**	76.5%	92.7%	93.4%	91.4%
100	89.5%	90.7%	86.8%	86.7%	88.0%	86.0%
150	65. 7%	62.5%	91.4%	82.8%	56.4%	68.7%
200	82.1%	84.8%	88.1%	79.5%	79.4%	67.5%
250	84.2%	82.7%	81.5%	59.8%	87.4%	71.1%
300	85.5%	87.4%	76.9%	60.4%	88.8%	70.4%
350	86.8%	86.1%	64.5%	57.3%	90.0%	65.4%

TABLE 3. Accuracies of SE features under different neurons and activation functions.

Number	Sigmoid	Sin	Hardlim	Relu	Tanh	Isigmoid
50	94.0%	**95.4%**	81.6%	93.4%	90.1%	93.4%
100	90.2%	91.4%	82.3%	90.0%	88.8%	88.1%
150	73.6%	76.9%	82.9%	88.1%	62.4%	85.5%
200	82.2%	90.1%	84.8%	88.1%	74.9%	73.8%
250	89.5%	89.5%	81.7%	70.3%	82.9%	59.1%
300	93.4%	92.7%	82.2%	67.0%	88.1%	64.6%
350	92.1%	90.1%	78.8%	63.6%	88.8%	67.6%

2) EXPERIMENTAL RESULTS OF PROPOSED METHOD

According to the above analytic results, the number of hidden layer neurons is set to 50 and the sinusoidal function is selected as the activation function. In addition, 50 sets of sub-classifiers are trained with the time-frequency entropy and the singular entropy as input vectors, respectively, i.e., a total of 100 sets of results are generated in each testing sample during the diagnostic process, and these 100 sets of results are used for comprehensive decision-making. The total number of correct classifications of 48 testing samples in 100 sub-classifiers is depicted in Fig. 10 with the form of histogram. The horizontal axis represents the testing sample. Samples 1^{st}-12^{th}, 13^{th}-24^{th}, 25^{th}-36^{th}, and 37^{th}-48^{th} belong to the

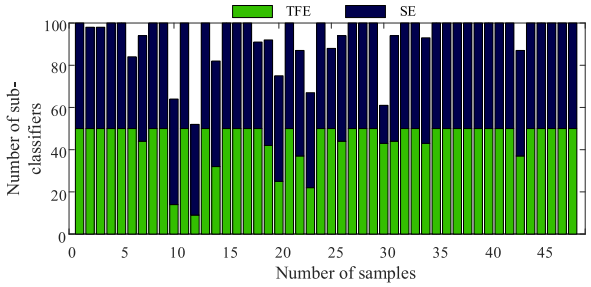


FIGURE 10. Number of correct classifications of test samples in 100 sub-classifiers.

normal state, the base-loosen state, the C-phase leading-closing state, and the C-phase hysteresis-closing state, respectively. The vertical axis represents the correct number of classifications. Blue color labeling is the correct classification number of SE classifiers, and the green color labeling is the correct classification number of TFE classifiers. The decision criterion for the IELM is: the more the numbers of classifiers with respect to a certain type, the greater the probability of determining it as this type.

For the first testing sample, the number of the normal state identified by the TFE classifier and the SE classifier is 50. That is, 100 classification results point to the normal state and are finally determined to be normal. It is clearly shown that the state characteristics of this sample are relatively obvious, so that each sub-classifier can correctly classify it. For the 12^{th} testing sample, the number of the TFE classifier recognized as the normal state is only nine. However, there are 43 SE classifiers to be identified as the normal state. Therefore, 52 of the 100 sub-classifiers trained by the two features identify it as the normal state, and the 12^{th} sample is accurately identified as the normal state according to the decision rule. Obviously, if only relying on the TFE classifier for decision, eventually there will be a miscalculation. For some samples with relatively weak state features, in the case where some of the features are weak, the other features are enhanced, and the complementarity of the two makes the overall features distinguishing significantly. Therefore, based on the hybrid features of the TFE and the SE, the intrinsic information of the signal can be mined deeper. As can be seen from Fig. 10, the total number of sub-classifiers correctly classified for each testing sample is greater than 50. According to the decision rule defined in this study, the recognition accuracy of the testing samples can reach 100%.

The TE, FE, TFE, and SE are compared as the features in this study to further verify the effect of the feature extraction. The source of experimental data set and the ratio of training data and testing data are the same as the previous experiments. Experimental results in Table 4 show that the TFE has higher identification accuracy in comparison with the TE and the FE. Although both the TFE and the SE can achieve a higher accuracy, they cannot be completely classified correctly, and the classification results of them do not show a mutual subset relationship. For example, the TFE is classified wrong in Class 1 and Class 2 sets, while the SE is classified wrong

TABLE 4. Classification results with respect to multiple features.

Feature	Class 1	Class 2	Class 3	Class 4	Accuracy
TE	9	10	11	11	85.42%
FE	10	11	10	11	87.50%
TFE	10	11	12	12	93.75%
SE	12	12	11	12	97.92%
TFE+SE	**12**	**12**	**12**	**12**	**100%**

in Class 3. The hybrid features of the TFE and the SE can solve the problem of insufficient feature representation for a single type.

On the other hand, one tries to input both TFE and SE features into a classifier for training. The effect is slightly better than the one of the TFE or the SE, separately. In order to improve the accuracy, multiple sub-classifiers are also used for integrating decision making. Although experimental results show that the identification accuracy can be further improved with the number of sub-categories increases, more sub-classifiers are required to stabilize the results in comparison with the method proposed in this study. Because the separate training method makes the classifier more diverse, the integrated classifier has a stronger cognitive ability.

TABLE 5. Classification results with respect to multiple sample sets.

Data set	Class 1	Class 2	Class 3	Class 4	Accuracy
Sample set 2	12	12	12	12	100%
Sample set 3	12	12	10	12	95.83%
Sample set 4	12	12	12	12	100%

Three new sample sets are respectively collected and observed to verify the robustness of the IELM. In order to distinguish from the first sample set, three new sample sets are respectively recorded as sample set 2, sample set 3, and sample set 4. Moreover, the number of each mechanical state samples, and the ratio of the training set and testing set in each sample set are consistent with the above experiments. Experimental results in Table 5 show that best classification accuracy also can be achieved for three newly collected sample sets. For the sample set 3 in Class 3 that was misjudged, it does not rule out that abnormal data or interference from the surrounding environment during the experimental collection process has changed the way the mechanical components interact. From a holistic perspective, the integrated learning approach makes the output more stable. The examined results of the above four sample sets strongly verify the robustness of the method described in this study.

3) COMPARISON WITH OTHER METHODS

In order to analyze the difference between the classification accuracy of the proposed method and other methods in previous researches, it is compared with the methods in [4], [6], [11]. In [4], the S-transformation is used to decompose the vibration signal, and the time-frequency entropy is extracted and classified by the OCSVM. In [6], the WPT is used to decompose the vibration signal, and the BPNN is

TABLE 6. Performance comparisons.

Method Name	Class 1	Class 2	Class 3	Class 4	Accuracy
Proposed method	**12**	**12**	**12**	**12**	**100%**
Ref. [4]	10	9	12	12	89.58%
Ref. [6]	9	10	8	11	79.17%
Ref. [11]	11	9	12	12	91.67%

used as the fault classifier after obtaining the time-frequency entropy. In [11], the VMD is used to decompose the vibration signal, and the singular value is obtained by block and then classified by the OCSVM. Experiment data comes from the sample set 2. Among them, 76% of each mechanical state is used to train the classifier model, and 24% of the samples are used to test the accuracy of the classification. The examined results of four methods are summarized in Table 6.

Although the time-frequency entropy is used as the characteristic in both [4] and [6], the results are quite different, mainly due to the difference between the SVM and the BPNN. Obviously, the BPNN has insufficient ability to classify small sample events. Huang *et al.* [4], [11] used the SVM as the classifier, which has a higher recognition rate than the one in [6]. But, there is also a certain amount of misjudgment. There are three main reasons for misjudgment: 1) Whether the components obtained by the decomposition can accurately describe the time-frequency quantity of the signal; 2) Whether the extracted features have the characterization ability, and 3) Whether the classifier has a strong generalization. The classification accuracy of the proposed method is higher than those of the other three methods, mainly reflected in the double feature to make up for the problem of insufficient representation ability of single-class features. In addition, the IELM is used as a classifier, which makes up for the lack of generalization ability of a single classifier.

VI. CONCLUSION

In this study, vibration signals are taken as analytic objects, and a method for mechanical fault diagnoses of high-voltage circuit breakers based on hybrid feature extraction and IELM is proposed. From theoretical analyses and experimental results, the major contributions of this study are summarized as follows:

(i) The vibration signal is decomposed by the CEEMDAN, and two different characteristics including the time-frequency entropy and the singular entropy are constructed. The waveform is normalized by the NCDF, which improves the stability of the singular entropy characterization signal. Moreover, the hybrid features avoid the problem of insufficient representation ability for single-type features. The corresponding experiments show that the classification accuracy with respect to the hybrid features can reach 100%, which is significantly higher than the one with respect to the single-type features.

(ii) Based on the idea of integrated decision-making, an IELM is constructed to compensate for the inconsistency of the single ELM model. On the other hand, the ELM for large training samples has also been successfully applied to solve the problem of small sample classification.

Under the same experimental samples, the performance of the proposed method is compared with other methods in [4], [6], [11]. It can conclude that the proposed method has certain improvements in the time-frequency signal analysis, the feature extraction, and the pattern recognition. The high diagnostic accuracy verifies the excellent performance of the proposed method.

REFERENCES

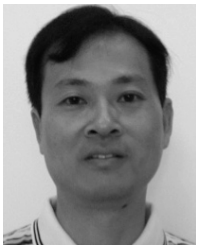
[1] A. Janssen, D. Makareinis, and C.-E. Sölver, "International surveys on circuit-breaker reliability data for substation and system studies," *IEEE Trans. Power Del.*, vol. 29, no. 2, pp. 808–814, Apr. 2014.
[2] X. Zhang, E. Gockenbach, Z. Liu, H. Chen, and L. Yang, "Reliability estimation of high voltage SF_6 circuit breakers by statistical analysis on the basis of the field data," *Electr. Power Syst. Res.*, vol. 103, pp. 105–113, Oct. 2013.
[3] N. Huang, L. Fang, G. Cai, D. Xu, H. Chen, and Y. Nie, "Mechanical fault diagnosis of high voltage circuit breakers with unknown fault type using hybrid classifier based on LMD and time segmentation energy entropy," *Entropy*, vol. 18, no. 9, p. 322, Sep. 2016.
[4] N. T. Huang *et al.*, "Mechanical fault diagnosis of high voltage circuit breakers based on wavelet time-frequency entropy and one-class support vector machine," *Entropy*, vol. 18, no. 1, p. 7, Dec. 2015.
[5] S. Ma, M. Chen, J. Wu, Y. Wang, B. Jia, and Y. Jiang, "Intelligent fault diagnosis of HVCB with feature space optimization-based random forest," *Sensors*, vol. 18, no. 4, p. 1221, Apr. 2018.
[6] S. Liu *et al.*, "Study of PSO-BP neural networks application in high-voltage circuit breakers mechanical fault diagnosis," in *Proc. China Int. Conf. Electr. Distrib.*, Aug. 2016, pp. 1–5.
[7] Y. W. Xing, M. L. Liu, P. Yang, Q. Peng, and B. Li, "Fault diagnosis method of HV circuit breaker based on wavelet packet time-frequency entropy and BP neural network," in *Proc. 29th Chin. Control Decis. Conf.*, May 2017, pp. 4143–4148.
[8] M. Liu, K. Q. Wang, L. Sun, and J. Zhen, "Applying empirical mode decomposition (EMD) and entropy to diagnose circuit breaker faults," *Optik*, vol. 126, no. 20, pp. 2338–2342, Oct. 2015.
[9] Z. Jianfeng, L. Mingliang, W. Keqi, X. Jingyan, and S. Shuli, "Feature extraction of HV circuit breaker based on ensemble empirical mode decomposition and correlation dimension," in *Proc. IEEE Int. Conf. Electron. Meas. Instrum.*, Jul. 2015, pp. 546–551.
[10] J. Zhang, M. Liu, K. Wang, and S. Sun, "Mechanical fault diagnosis for HV circuit breakers based on ensemble empirical mode decomposition energy entropy and support vector machine," *Math. Problems Eng.*, vol. 2015, Jun. 2015, Art. no. 101757.
[11] N. Huang, H. Chen, G. Cai, L. Fang, and Y. Wang, "Mechanical fault diagnosis of high voltage circuit breakers based on variational mode decomposition and multi-layer classifier," *Sensors*, vol. 16, no. 11, p. 1887, Nov. 2016.
[12] L. Dou, S. Wan, and C. Zhan, "Application of multiscale entropy in mechanical fault diagnosis of high voltage circuit breaker," *Entropy*, vol. 20, no. 5, p. 325, Apr. 2018.
[13] H. Mengyuan, D. Qiaolin, Z. Shutao, and W. Yao, "Research of circuit breaker intelligent fault diagnosis method based on double clustering," *IEICE Electron. Express*, vol. 14, no. 17, Aug. 2017, Art. no. 20170463.
[14] B. Li, M. Liu, Z. Guo, and Y. Ji, "Mechanical fault diagnosis of high voltage circuit breakers utilizing EWT-improved time frequency entropy and optimal GRNN classifier," *Entropy*, vol. 20, no. 6, pp. 448–470, Jun. 2018.
[15] W. Niu, G. Liang, H. Yuan, and B. Li, "A fault diagnosis method of high voltage circuit breaker based on moving contact motion trajectory and ELM," *Math. Problems Eng.*, vol. 2016, no. 8, Jan. 2016, Art. no. 3271042.
[16] A. Forootani, A. A. Afzalian, and A. N. Ghohmshe, "Model-based fault analysis of a high-voltage circuit breaker operating mechanism," *Turkish J. Elect. Eng. Comput. Sci.*, vol. 25, no. 3, pp. 2349–2362, May 2016.
[17] M. Fei, J. Mei, J. Zheng, and Y. Wang, "Development and application of distributed multilayer on-line monitoring system for high voltage vacuum circuit breaker," *J. Elect. Eng. Technol.*, vol. 8, no. 4, pp. 813–823, Jul. 2013.

[18] Y. Qin, "a new family of model-based impulsive wavelets and their sparse representation for rolling bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2716–2726, Mar. 2018.
[19] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London Ser. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
[20] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.
[21] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
[22] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2011, pp. 4144–4147.
[23] Y. Lv, R. Yuan, T. Wang, H. Li, and G. Song, "Health degradation monitoring and early fault diagnosis of a rolling bearing based on CEEMDAN and improved MMSE," *Materials*, vol. 11, no. 6, p. 1009, Jun. 2018.
[24] Y. Ren, P. N. Suganthan, and N. Srikanth, "A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 236–244, Dec. 2017.
[25] W. Zhang, Z. Qu, K. Zhang, W. Mao, Y. Ma, and X. Fan, "A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting," *Energy Convers. Manage.*, vol. 136, pp. 439–451, Mar. 2017.
[26] T. Liu, Z. Luo, J. Huang, and S. Yan, "A comparative study of four kinds of adaptive decomposition algorithms and their applications," *Sensors*, vol. 18, no. 7, pp. 2120–2170, Jul. 2018.
[27] Z. Wang *et al.*, "A novel fault diagnosis method of gearbox based on maximum kurtosis spectral entropy deconvolution," *IEEE Access*, vol. 7, pp. 29520–29532, Feb. 2019.
[28] Z. Qiao and Z. Pan, "SVD principle analysis and fault diagnosis for bearings based on the correlation coefficient," *Meas. Sci Technol.*, vol. 26, no. 8, Jul. 2015, Art. no. 085014.
[29] H. Jiang, J. Chen, G. Dong, T. Liu, and G. Chen, "Study on Hankel matrix-based SVD and its application in rolling element bearing fault diagnosis," *Mech. Syst. Signal Process.*, vols. 52–53, pp. 338–359, Feb. 2015.
[30] M. Rostaghi, M. R. Ashory, and H. Azami, "Application of dispersion entropy to status characterization of rotary machines," *J. Sound Vib.*, vol. 438, no. 6, pp. 291–308, Jan. 2018.
[31] X. Gan, H. Lu, G. Yang, and J. Liu, "Rolling bearing diagnosis based on composite multiscale weighted permutation entropy," *Entropy*, vol. 20, no. 11, p. 821, Oct. 2018.
[32] S. Dong, X. Xu, and R. Chen, "Application of fuzzy C-means method and classification model of optimized K-nearest neighbor for fault diagnosis of bearing," *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 38, no. 8, pp. 2255–2263, Dec. 2016.
[33] A. T. Azar and S. M. El-Metwally, "Decision tree classifiers for automated medical diagnosis," *Neural Comput. Appl.*, vol. 23, nos. 7–8, pp. 2387–2403, Dec. 2013.
[34] L. Song, H. Wang, and P. Chen, "Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 8, pp. 1887–1899, Aug. 2018.
[35] Z. T. Duan, Y. S. Wang, and Y. F. Xing, "Sound quality prediction of vehicle interior noise under multiple working conditions using back-propagation neural network model," *J. Transp. Technol.*, vol. 5, pp. 134–139, Apr. 2015.
[36] L. Song, H. Wang, and P. Chen, "Step-by-step fuzzy diagnosis method for equipment based on symptom extraction and trivalent logic fuzzy diagnosis theory," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3467–3478, Dec. 2018.
[37] H. Wang, S. Li, L. Song, and L. Cui, "A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals," *Comput. Ind.*, vol. 105, pp. 182–190, Feb. 2019.
[38] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2011.
[39] Z. Zhao, Z. Chen, Y. Chen, S. Wang, and H. Wang, "A class incremental extreme learning machine for activity recognition," *Cogn. Comput.*, vol. 6, no. 3, pp. 423–431, Sep. 2014.
[40] Y. Qin, X. Wang, and J. Zou, "The optimized deep belief networks with improved logistic Sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3814–3824, May 2019.

WEI GAO was born in Pingtan, China, in 1983. He received the B.S. degree in electrical engineering and automation and the M.S. degree in power system and automation from Fuzhou University, China, in 2005 and 2008, respectively. He is currently pursuing the Ph.D. degree with the National Taiwan University of Science and Technology, Taiwan. Since 2008, he has been a Lecturer with Fuzhou University. He has authored or coauthored about six journal papers (EI) and one book. He holds about three patents. His research interests include generation technology of photovoltaic and faults diagnosis of power equipment.

RONG-JONG WAI (M'99–SM'05) was born in Tainan, Taiwan, in 1974. He received the B.S. degree in electrical engineering and the Ph.D. degree in electronic engineering from Chung Yuan Christian University, Chung Li, Taiwan, in 1996 and 1999, respectively. From 1998 to 2015, he was with Yuan Ze University, Chung Li, where he was the Dean of the General Affairs, from 2008 to 2013, and the Chairman of the Department of Electrical Engineering, from 2014 to 2015. Since 2015, he has been with the National Taiwan University of Science and Technology, Taipei, Taiwan, where he is currently a Distinguished Professor, the Dean of the General Affairs, and the Director of the Energy Technology and Mechatronics Laboratory. He has authored more than 170 conference papers, over 180 international journal papers, four book chapters, and 57 inventive patents. His research interests include power electronics, motor servo drives, mechatronics, energy technology, and control theory applications. The outstanding achievement of his research is for contributions to real-time intelligent control in practical applications and high-efficiency power converters in energy technology. He is a fellow of the Institution of Engineering and Technology (U.K.) and a Senior Member of the Institute of Electrical and Electronics Engineers (USA).

SU-PENG QIAO received the B.S. degree from Fuzhou University Zhicheng College, China, in 2017. He is currently pursuing the M.S. degree with Fuzhou University. His research interest includes the diagnosis of power equipment failures in power distribution systems.



MOU-FA GUO was born in Fuzhou, China, in 1973. He received the B.S. and M.S. degrees from Fuzhou University, Fuzhou, China, in 1996 and 1999, respectively, and the Ph.D. degree from Yuan Ze University, Taiwan, in 2018, all in electrical engineering. Since 2000, he has been with Fuzhou University, where he is currently a Professor, and the Chairman of the Department of Electric Power Engineering. His research interests include power distribution systems and its automation, and the application of artificial intelligence in power distribution systems.