High-Voltage Circuit Breaker Fault Diagnosis Using a Hybrid Feature Transformation Approach Based on Random Forest and Stacked Autoencoder

Suliang Ma, *Graduate Student Member, IEEE*, Mingxuan Chen, *Graduate Student Member, IEEE*, Jianwen Wu, *Senior Member, IEEE*, Yuhao Wang, Bowen Jia, and Yuan Jiang, *Member, IEEE*

***Abstract*—In recent years, machine learning techniques have been applied to test the fault type in high-voltage circuit breakers (HVCBs). Most related research involves in improving the classification method for higher precision. Nevertheless, as an important part of the diagnosis, the feature information description of the vibration signal of an HVCB has been neglected; in particular, its diversity and significance are rarely considered in many real-world fault-diagnosis applications. Therefore, in this paper, a hybrid feature transformation is proposed to optimize the diagnosis performance for HVCB faults. First, we introduce a nonlinear feature mapping in the wavelet package time–frequency energy rate feature space based on random forest binary coding to extend the feature width. Then, a stacked autoencoder neural network is used for compressing the feature depth. Finally, five typical classifiers are applied in the hybrid feature space based on the experimental dataset. The superiority of the proposed feature optimal approach is verified by comparing the results in the three abovementioned feature spaces.**

***Index Terms*—Fault diagnosis, feature transformation, high-voltage circuit breaker (HVCB), random forest (RF), stacked autoencoder (SAE), wavelet packet decomposition.**

I. Introduction

HIGH-VOLTAGE circuit breaker (HVCB) faults can lead to significant physical harm as well as considerable economic loss for the power grid [1], [2]. Therefore, several experts and scholars have focused their research efforts on HVCB fault-diagnosis technology. In addition, intelligent mechanical fault-diagnosis technologies based on machine learning have been successfully applied for various machines, including gearboxes and wind turbines, among others. Therefore, establishing a reliable, precise, and intelligent identification model for HVCB faults has become a hot topic.

A. Previous Work

1) HVCB Diagnosis: In recent times, the vibration characteristics during the closing or opening process of an HVCB has replaced the contact displacement [3], [4] and electromagnet coil current information [5], [6] as the distinguishing method for the operational condition of an HVCB; for example, an intelligent method based on density peaks clustering algorithm fused kernel fuzzy c-means and support vector machine (SVM) was proposed for HVCB mechanical fault diagnosis, which considerably improved the accuracy of fault diagnosis [7]. In another study, a genetic algorithm (GA) was utilized to optimize the radial basis function parameters of the SVM; the experimental results of this study indicate that the classification accuracy of the GA-SVM approach is higher than those of the artificial neural network and traditional SVM approaches [8]. Furthermore, a hybrid classifier based on support vector data description and fuzzy c-means was designed to improve the identification accuracy of HVCB faults; in this case, the experiments were performed on a real SF6 HVCB and validated [9]. However, in turn, these research works indicate that it is difficult to analyze the vibration characteristics of HVCBs owing to their nonstationary and nonlinear peculiarities.

2) Time–Frequency Analysis-Based Approaches and Feature Extraction: As an important part of the HVCB fault analysis methodology, feature extraction can considerably restrict HVCB diagnosis accuracy. Considering this, time–frequency analysis tools have been widely used to extract fault feature vectors of the HVCB vibration signals. In general, there are two methods based on the intrinsic model function of the HVCB original signal, including empirical mode decomposition (EMD) [10] and its improved approaches [11], [12] as well as those involving some basis functions, such as short-time Fourier transform (STFT) [13] and wavelet transform (WT) [14]. EMD involves decomposing complicated signals into several intrinsic mode functions containing local characteristic timescale

Manuscript received May 24, 2018; revised July 30, 2018, September 25, 2018, and October 12, 2018; accepted October 18, 2018. Date of publication November 7, 2018; date of current version July 31, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 51377007 and Grant 51677002, and in part by China Postdoctoral Science Foundation under Grant 2018M631307. *(Corresponding author: Jianwen Wu.)*

The authors are with the School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China (e-mail: masuliang@buaa.edu.cn; mingxuan_chen@buaa.edu.cn; wujianwen@buaa.edu.cn; 18810371198@163.com; jiabowen109@126.com; jiangy_luckystar@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2018.2879308

information of the HVCB original signal [10]; however, it suffers from disadvantages, such as the end effect, high computational complexity, and nonconformity of decomposition among different signals, spectrum overlapping, and chaotic boundary frequency among different intrinsic modes. Although some subsequent studies do propose improved methods [11], [12] to solve these abovementioned fractional problems, the other disadvantages still limit their application. In contrast to the EMD approach, STFT was proposed as an adjustable windowed Fourier transform to represent a signal's spectral variations during a particular interval of time, such as during the vibration signal [13]. However, the different window length parameters of STFT significantly affect the time–frequency spectrum result; in addition, the basis function for STFT is still a sin (cos) function with infinite energy for the entire time domain, which is not suitable for representing a local pulsing signal. To overcome this problem, the Haar wavelets family, proposed by Alfred Haar in 1909, was used to analyze a given signal in terms of functions that are more localized in time than the harmonic functions used in the Fourier analysis [14]. Later, wavelet and wavelet packet transform (WPT) theory rapidly developed and was applied in many fields [15]–[18]. For example, in [15], the authors proposed a different frequency band signal fusion method based on binary WPT to form the waveform feature space. In particular, the use of continuous WT, discrete WT, as well as WPT and second-generation WTs has been studied for fault diagnosis in rotary machines [16]. Kumar and Kumar [17] processed the vibration signal using WT, and then extracted prominent features as well as constructed the scale marginal integration graph, which shows multiscale energy distribution and can be used to identify the defective condition of a centrifugal pump. In another study, dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes are applied to deep residual networks [18].

3) Machine Learning-Based Approaches: Many researchers and scholars have also developed machine learning-based approaches for fault identification and diagnosis to improve classification technology. Rennie [19] proposed a logistic regression (LR) model for a classification task and explained that a multiclass generalization of LR, which is commonly referred to as Softmax. Decision tree (DT) classifier can be used for motor fault-diagnosis tasks or to detect breast cancer based on medical data [20], [21]. In contrast, Vernekar adopted a naive Bayes algorithm as a fault classifier to investigate the status of a monoblock centrifugal pump or engine [22]. Although a back propagation neural network (BPNN) was successfully adopted for fault detection based on extracted time–frequency features [23] and 15 calculated time-domain statistical parameters [24], a good BPNN for diagnosis relies heavily on a large-scale dataset. However, in the case of a small sample, SVM and its improved variants, one class support vector machine (OCSVM), have been applied for identifying mechanical faults in HVCBs; such methods have achieved satisfactory diagnostic performance [25], [26]. The above methods are effectively verified in their respective application scenarios. Furthermore, it has been noted that the training process of a single classifier is affected by the global error rate. Therefore, the models might be biased toward the majority class and ignore the minority class. Moreover, incorrectly collected samples or dispersion in the dataset might potentially cause overfitting. However, as proposed by Breiman [27] and studied by Biau *et al*. [28], a random forest (RF) is an improved machine learning algorithm based on a random subspace and ensemble learning, which involves the construction of a set of individual and weak classifiers and subsequently their combination is used to classify new data [29]. And an improved RF algorithm has successfully been applied to identifying the HVBC's mechanical faults [30]. In real industrial environments, it is difficult to acquire a large number of samples for classifier training and estimate the significance of feature candidates for fault diagnosis; considering this, RF-based models have a better performance rate than other classical classifiers [31], [32].

B. Contribution

The analysis of most previous works indicates that they involve designing more powerful detectors for diagnosis. But when there is a problem of singular value and large noise in the sample data, a single strong classifier, such as SVM or DT, may show an overfitting problem. And a depth learning network generally requires a large amount of data during the training process; a small amount of data may cause the deep network to appear as an underfitting problem. In the application scenarios of this paper, the HVCB vibration data have large noise and strong dispersibility. It is very difficult to measure, because a large number of separation processes will affect the life and structure of the circuit breaker. In order to enhance fault feature variability, solve small sample problems, reduce dispersion and noise effects, a novel hybrid feature transformation method is proposed by considering the characteristics of the HVCB fault-diagnosis problem such that the transferable features could be learned from feature width and depth. Then, five standard machine learning classifiers, namely Softmax, DT, One-Class SVM, BPNN, and RF, are constructed to compare the diagnosis results in various feature spaces. To the best of our knowledge, this is the first attempt for solving feature significance issues in fault diagnosis by transforming and compressing features of the HVCB vibration based on RF coding and stacked autoencoder (SAE) neural network, respectively. The primary contributions of this paper can be summarized as follows.

1) The feature for the representation of various failures typically includes various vibration information with clear differences among the various faults; thus, to enhance characteristic description of the HVCB vibration information, a binary coding feature transformation based on RF is proposed. With ensemble learning and random subspace technology, the RF model is regarded as a nonlinear feature mapping model rather than a classifier. In summary, the original and continuous feature is transferred into a discrete binary coding feature string.
2) Because the characteristic diversity will lead to the inclusion of considerable superfluous information that might affect the diagnosis result, an SAE neural network is used to reduce the dimensionality of the feature space by deeply fusing some most prominent features from the bin coding feature space.

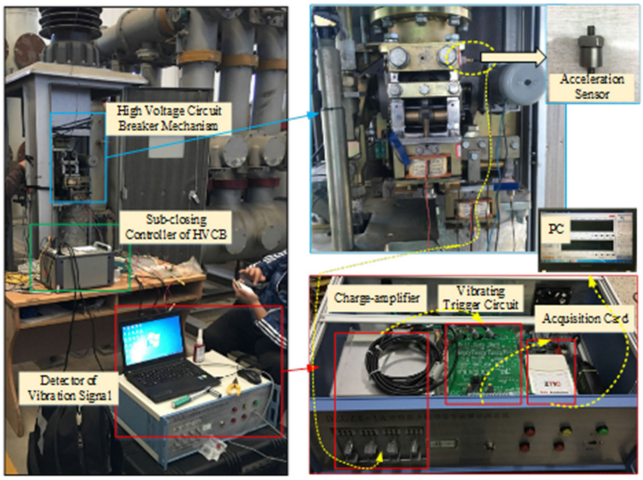

Fig. 1. Experimental acquisition system with HVCB.

3) The hybrid RF binary coding (RFBC)+SAE feature transformation can notably improve the accuracy of diagnosis of HVCB faults for a single classifier based on the experimental results in this paper. In particular, our proposed method presents a novel way to diagnose fault types in other applications as well, including planetary gearboxes, wind turbines, rolling element bearings, and centrifugal pumps.

C. Work Organization

The remainder of this paper is organized as follows. Section II introduces the acquisition system based on the vibration of HVCBs as well as the experimental datasets under variable operating conditions. Then, the procedure for vibration feature information extraction, transformation, and compression is briefly described in Section III. In Section IV, we present the improved results provided by our proposed hybrid feature transformation method by showing the feature significance in the feature evolution process and comparing the diagnosis accuracy with those of other classifiers in different feature spaces. Finally, Section V provides our conclusions.

II. Experimental Setup and Inspection

For the experiments in this paper, we used an LW30-252 type SF6 HVCB (Shandong TaiKai High Voltage Switchgear Co, Ltd., Shandong, China), control box for opening (closing) of the HVCB, and vibration signal acquisition system consisting of an acceleration sensor YD-111T, charge-amplifier TS5863, vibrating trigger circuit, and data acquisition card EM9118B; this experimental setup is shown in Fig. 1. The acquisition system has a measuring range, sensitivity, natural frequency, frequency response, sampling rate, and sampling period of $\pm 10\ 000$ g (where g $= 9.8\ \mathrm{m{\cdot}s^{-2}}$), $0.5\ \mathrm{mV{\cdot}g^{-1}}$, 45 kHz, 15 kHz, 300 kHz, and 120 (60) ms in the close (open) state, respectively.

In particular, the vibration information is collected using this acquisition system; when vibration occurs, the piezoelectric accelerometer YD-111T produces charge accumulation. The charge amplifier converts these accumulated charges into a voltage signal; if the voltage amplitude is greater than a certain threshold, a trigger circuit causes the acquisition card to obtain these voltage values for the time period representing the vibration acceleration. Then, this vibration information is stored in a PC. In Fig. 1, the yellow-dotted lines represent the acquisition flow of vibration signals. It is important to note that previous studies have reported that the accuracy of vibration feature-based fault-diagnosis method is affected by the installation site and method of the acceleration sensor. In this paper, the acceleration sensor was installed using the threaded fastening method; in addition, based on measuring the vibration signals at multiple sites, the final installation site shown in Fig. 1 was selected as the optimal installation site.

TABLE I
Summary of States for HVCB Considered in This Paper

Health Condition	Description of States	Category Label	Sample Number
Healthy	Normal case	Class 1	50
Faulty	Closing spring fatigue	Class 2	50
	Opening spring fatigue	Class 3	50
	Damping increase for transmission shaft	Class 4	50
	Oil damper leaking	Class 5	50
	Looseness of base screw	Class 6	50

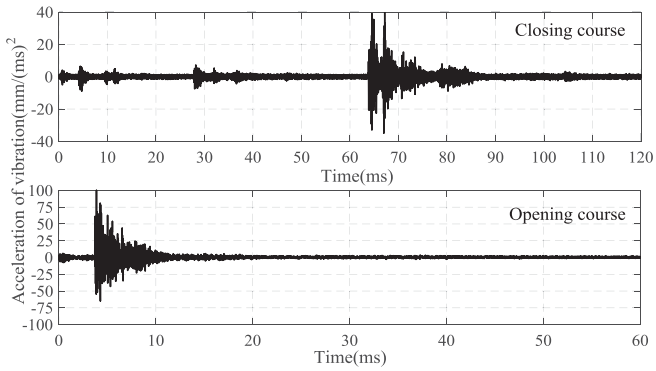

Fig. 2. Vibration signals under normal condition during the opening and closing processes.

To verify the efficacy of the proposed fault-diagnosis method, six states of the HVCB are considered for diagnosis; these are listed and described in Table I. In particular, we measured 50 vibration signals per state; then, we randomly selected 35 samples for training and the remainder were used for testing. The HVCB's vibration signal under normal condition is shown in Fig. 2.

As shown in Fig. 2, the closing process of an HVCB involves a multistage impact and attenuation process, with the maximum vibration acceleration of approximately 4000 g. In contrast, the opening process can only be represented by a trumpet-type envelope, with a maximum vibration acceleration of approximately 10 000 g.

Considering the operation mechanism of an HVCB, it is known that some components of HVCB, such as the closing spring and oil damper, only affect the closing process.

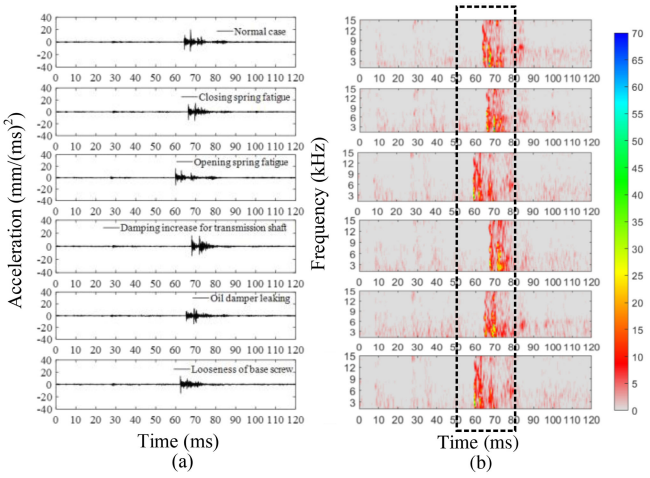

Fig. 3. Comparison of vibration signals under different cases. (a) Vibration signals in time domain. (b) Wavelet package spectrum of vibration signals.

Therefore, in this paper, the vibration signal during the closing process is adopted for HVCB fault diagnosis.

III. Proposed Method

In HVCB diagnosis, feature extraction is a key procedure that can strongly affect diagnosis precision. First, based on the time–frequency energy ratio with the WPT, an original feature space can be described, as discussed in Section III-A. Then, our novel hybrid feature transform method based on RFBC and SAE networks is implemented that can effectively improve the diagnose accuracy of various faults; this is discussed in Section III-B and C. Lastly, the general procedure of our proposed method is described in Section III-D.

A. Feature Extraction Based on WPT

As already stated in the previous section, the spectral variations in the vibration signal constantly change during this interval of time. As a well-developed time–frequency analysis technology, the WT has been applied to extract the HVCB vibration information in various fields; WT allows the use of long time intervals to obtain precise low-frequency information and shorter regions where high-frequency information is required; however, WT cannot effectively split the high-frequency bands containing rich fault modulation information. Therefore, a better representation of the vibration signal can be obtained by extending WT to WPT, which can simultaneously split the detailed and approximation versions of the vibration signal. Analysis of vibration signal difference of HVCB under different working conditions is based on wavelet packet analysis theory. The vibration signal in time domain and the wavelet package spectrum are shown in Fig. 3.

It can be seen from the figure that the energy distribution of different faults in time–frequency spectrum has significant difference compared to the normal situation. Taking closing spring fatigue as an example, during the closing process, the closing spring will cause the closing ability to decrease due

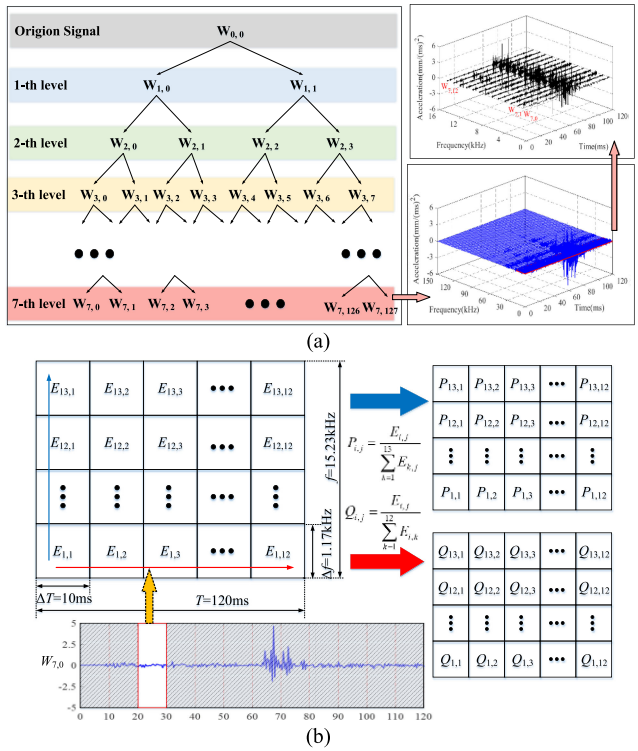

Fig. 4. WPT and feature extraction. (a) WPT time–frequency analysis. (b) Feature extraction.

to fatigue, the closing speed will be slower, and the closing buckle impact (maximum vibration peak) will lag behind, and its vibration capacity will decrease. The energy in each frequency band is more dispersed. Therefore, it is feasible to describe the performance of vibration information under different faults by the change of energy in the time–frequency spectrum.

The binary tree framework of the WPT algorithm divided into seven resolution levels for an origin vibration signal is shown in Fig. 4(a). The jth level and nth sub-band of the reconstructed wavelet coefficients W_j^n represent $1/2^n$ of the frequency information. In this paper, each vibration signal was subjected to a seven-level binary wavelet packet decomposition and wavelet coefficient reconstruction using the db3 wavelet basis. The system sampling frequency was set to 300 kHz; in addition, the width of each of the generated 2^7 frequency bands was 1.17 kHz. Considering that the acquisition system had a frequency response range from 0 to 15 kHz, the No. 0–12 wavelet coefficients at the seventh level wavelet packet were employed to characterize the time–frequency characteristics of the vibration signal. Then, the vibration signal during the closing process was split into 13 $\times$ 12 signal segments; in addition, the energy of each signal segment was calculated using the equation $E = \sqrt{\sum_{t=t_0}^{t_N} x^2(t)}$. This is shown in Fig. 4(b).

Moreover, considering the dispersity of the vibration signal energy of the HVCB and corresponding normalization processing, the proportion of each signal segment in the time–frequency direction was calculated using (1), where $P_{i,j}$ and $Q_{i,j}$ represent the proportion of the jth signal in the frequency direction within the ith time period as well as the proportion of the ith signal in

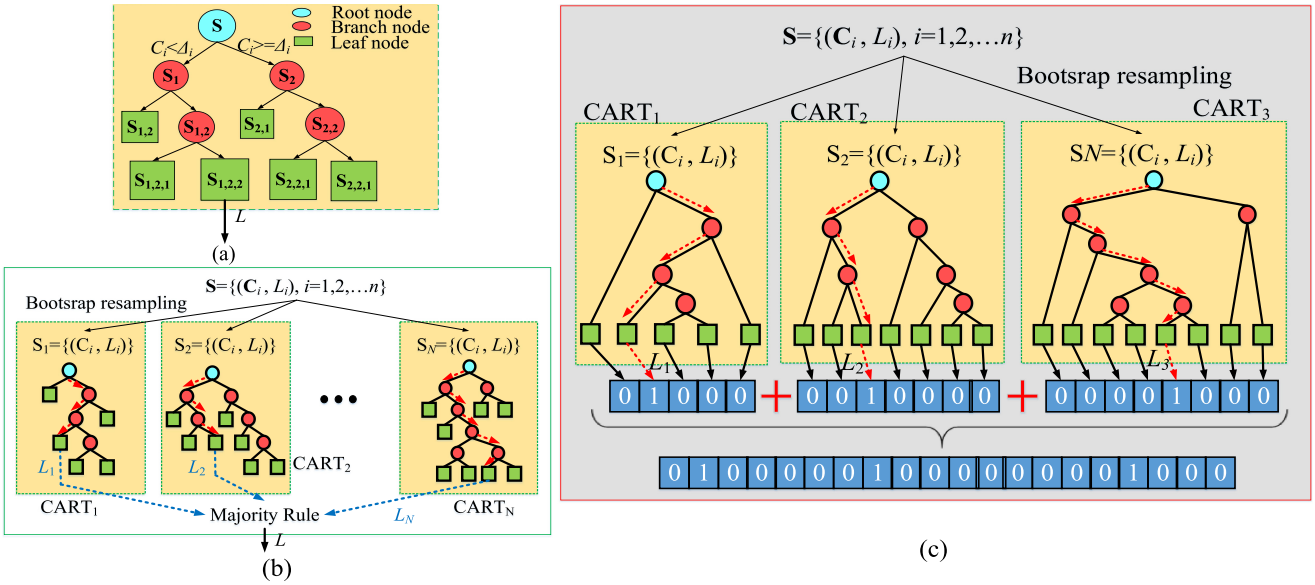

Fig. 5. Feature space transformation based on RF. (a) CART, (b) Random Forest Classification, (c) Feature space transform based on Random Forest.

the time direction within the *j*th frequency band, respectively [30]

$$\begin{cases} P_{i,j} = E_{i,j} / \sum_{k=1}^{13} E_{k,j} \\ Q_{i,j} = E_{i,j} / \sum_{k=1}^{12} E_{i,k} \end{cases}. \tag{1}$$

For the feature extraction process, the difference in the distribution order of sequence was considered and wavelet packet analysis-based time–frequency energy was regarded as a HVCB vibration signal feature to form the original feature space wavelet package time-frequency energy rate (WTFER), as shown in

$$\begin{aligned} \mathrm{WTFER} &= [\mathrm{WTFER}_t, \mathrm{WTFER}_f] \\ &= [P_{1,1}, P_{1,2}, \ldots P_{1,12}, P_{2,1}, \ldots, \\ &\quad P_{13,12}, Q_{1,1}, \ldots, Q_{13,12}]. \end{aligned} \tag{2}$$

B. Feature Space Transform With RF

In general, two methods are available to transform the feature space to improve diagnosis accuracy. The first such transformation is based on matrix theory. In particular, we can consider that the value of a categorical feature is equivalent to that of the coordinate and categorical feature and can be construed as the basis. Then, a transformation matrix that maps the original coordinate to a new coordinate is designed for more clear characterization. Another simple but effective method to learn nonlinear transformations is to bin the feature and treat the bin index as a categorical feature; for classification, this means that each categorical feature has only two solutions, 0 or 1.

As previously mentioned, RF is a novel, ensemble machine-learning algorithm, which combines the bagging integrated learning technology with the random subspace theory. The structure of RF for classification is shown in Fig. 5(b); the steps of the general process for RF generation are as follows.

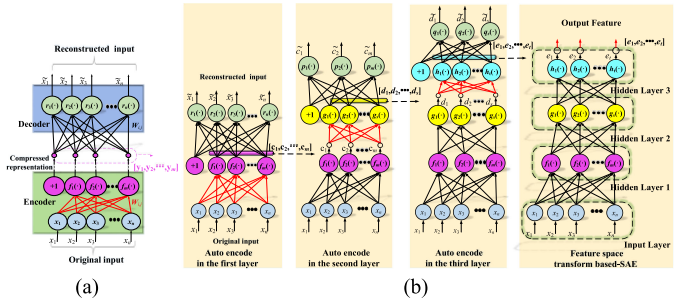

Fig. 6. Feature compression based on AE and SAE. (a) AE. (b) SAE.

Step 1: The training set $\mathbf{S}$ is generated from the original training set using the bootstrap resampling method; $\mathbf{S} = \{(C_i, L_i),\ i = 1, 2, \ldots n\}$, $(C_i, L_i) \in \mathbf{R}_d \times \mathrm{R}$, (C_i, L_i) represent feature set and label of the *i*th sampling.

Step 2: A DT is generated using the classification and regression tree (CART) algorithm based on the Gini index; the structure of the CART algorithm is shown in Fig. 5(a). It should be noted that the DT in the RF model is different from that in the CART algorithm where *m* characteristic attribute values are randomly selected from *d* characteristic attribute values for each node.

Step 3: Repeat Step 1 until the number of DTs reaches the set threshold value.

As already stated in the previous literature, the output of each tree is a certain label or class in the traditional RF model and the majority voting rule is defined by [29]–[32]

$$L = \arg_{\boldsymbol{L}}^{\max} \sum_{j=1}^{N_{\mathrm{tree}}} I(\mathrm{CART}_j(\boldsymbol{C}_i) = L_k) \tag{3}$$

where $I(\mathbf{C}_i)$ is an indication function, and arg() is a value function representing the number of trees that classify the test sample as the classification vector L_k.

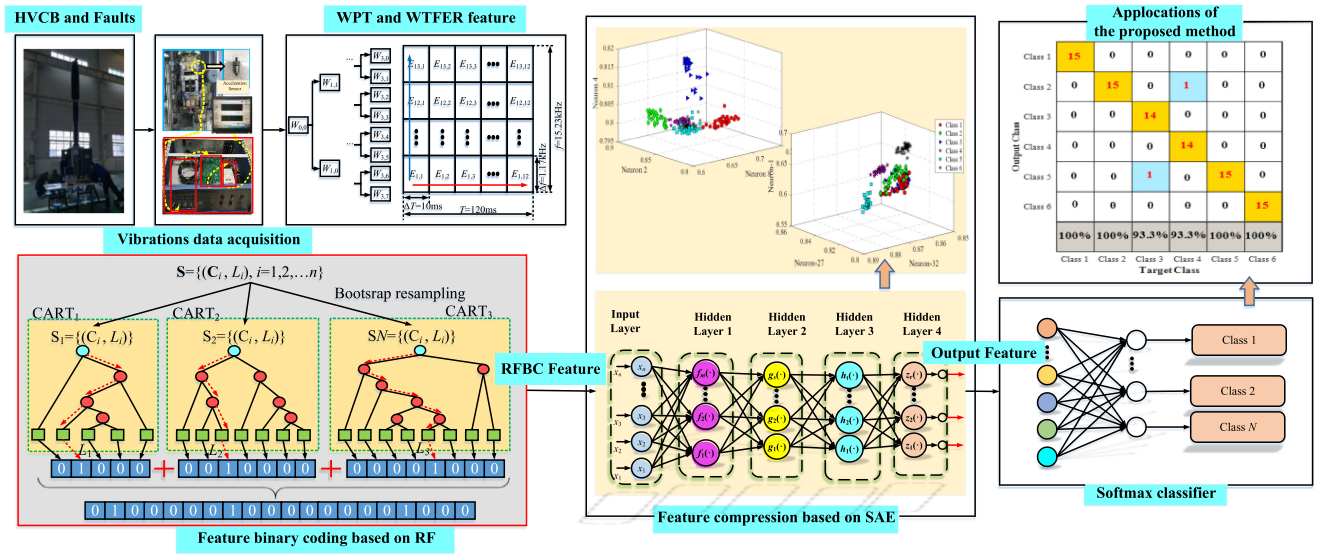

Fig. 7. Framework of the proposed method.

In this paper, we show that RF is a powerful and convenient method to implement nonlinear transformations to map the original feature to a binary coding space, rather than perform a classification. Each individual tree in RF is treated as a categorical feature that takes the index of the leaf at which the instance is present. In particular, we use 1-of-K coding for these types of features. The last binary coding feature space is expanded by combining the coding of every tree in the RF; for example, consider the RF model in Fig. 5(c) with three DTs, where the first, second, and third trees have five, seven, and eight leaves, respectively. If an instance ends up in leaf 2 in the first tree, leaf 3 in second tree, and leaf 5 in the last tree, the last binary coding feature space will be the following binary vector: [0; 1; 0; 0; 0; 0; 0; 1; 0; 0; 0; 0; 0; 0; 0; 0; 1; 0; 0; 0;].

C. Feature Compression With SAE

It is evident that the number of bits in a binary coding is large and sparse based on the RF considering that there are many CARTs in the RF and nodes on each CART. This indicates that it is important to use high-dimensional binary coding space compression and fusion. Therefore, the autoencoder (AE), which is a typical unsupervised feature learning neural network, might be an effective solution for fault diagnosis. If a single AE is unduly used for compression, the result might be worse than without using it because there might be a considerable loss of productive information. Therefore, in this paper, we design multiple AEs, stack them sequentially, and optimize them layer by layer to achieve a better compression feature; this is depicted in Fig. 6.

In particular, Fig. 6(a) presents a basic AE neural network that consists of an encoder and decoder. If the number of neurons in a hidden layer is less than the dimensionality of the input data and the output data of decoder are approximately equal to the input, we can assume that the input data are compressed by the encoder. In other words, the output of the hidden layer represents input information. Consider a training sample $\boldsymbol{x} = [x_1, x_2, \ldots x_n]^{\mathrm{T}} \in \Re^n$, the encoder maps the input vector x to a hidden representation $\boldsymbol{y} = [y_1, y_2, \ldots y_m]^{\mathrm{T}} \in \Re^m$ using the sigmoid function [33]–[35] as follows:

$$\boldsymbol{y} = f_\theta(\boldsymbol{x}) = s_f(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}) \tag{4}$$

where $s_f(\alpha) = 1/(1+e^{-\alpha})$ is the activation function of the encoder, and $\theta = \{\boldsymbol{W}, \boldsymbol{b}\}$ is the parameter of the encoder. In addition, $\boldsymbol{W} \in \Re^{m\times n}$ is the weight matrix, and $\boldsymbol{b} \in \Re^m$ is a bias vector.

Then, the hidden vector y is transformed back into a reconstruction vector $\tilde{\boldsymbol{x}} = [\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n]^{\mathrm{T}} \in \Re^n$ using the decoder, which decodes the vector as follows:

$$\tilde{\boldsymbol{x}} = r_\theta(\boldsymbol{y}) = s_g(\tilde{\boldsymbol{W}}\boldsymbol{y} + \tilde{\boldsymbol{b}}) \tag{5}$$

where $s_g(\alpha) = c\alpha$ is the activation function of the decoder, and $\tilde{\theta} = \{\tilde{\boldsymbol{W}}, \tilde{\boldsymbol{b}}\}$ is the parameter of decoder. As in the previous case, $\tilde{\boldsymbol{W}} \in \Re^{m\times n}$ is the weight matrix, and $\tilde{\boldsymbol{b}} \in \Re^m$ is a bias vector.

The parameter set $\{\theta, \tilde{\theta}\} = \{\boldsymbol{W}, \boldsymbol{b}, \tilde{\boldsymbol{W}}, \tilde{\boldsymbol{b}}\}$ of the basic AE is optimized using the gradient descent method to minimize the average reconstruction error [33]–[35] as follows:

$$J_{AE}(\theta, \tilde{\theta}) = -\sum_{i=1}^{n}[x_i \log(\tilde{x}_i) + (1-x_i)\log(1-\tilde{x}_i)] \tag{6}$$

TABLE II
PARAMETERS USED IN OUR EXPERIMENTS

Description	Value
Number of DTs in the RF	100
Maximum feature space dimension at each node of the DT	18
Number of hidden layers in the SAE	4
Numbers of neurons in the first hidden layer of the SAE	1024
Numbers of neurons in the second hidden layer of SAE	512
Numbers of neurons in the third hidden layer of SAE	256
Numbers of neurons in the fourth hidden layer of SAE	128
Learning rate of the SAE	0.1
Maximum epochs	1000

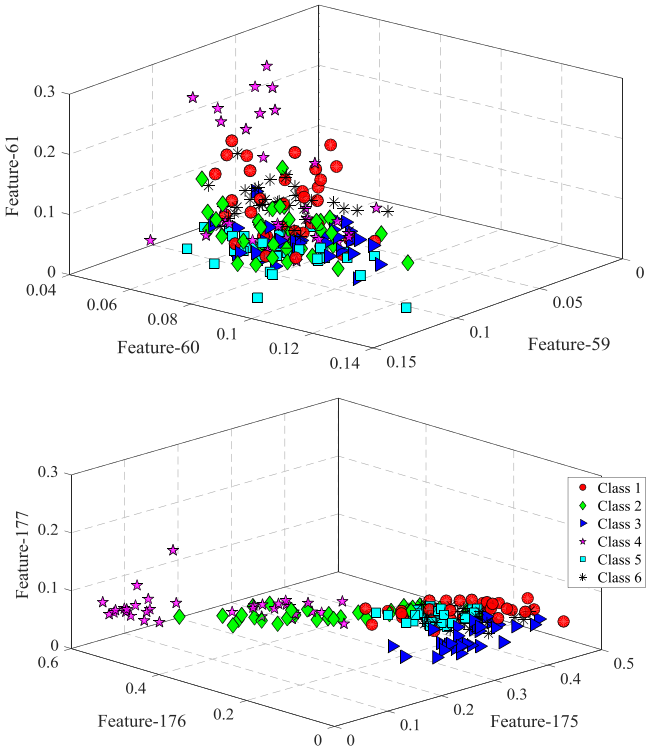

Fig. 8. Features distribution in the WTFER based on WPT.

where $J_{AE}(\boldsymbol{\theta})$ is the cross entropy regarded as the loss function, which measures the difference between $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$.

The topological graph of the SAE is shown in Fig. 6(b). First, a basic encoder is trained as the first layer of the SAE, and the input data are the binary coding features, wherein each x_i is the bit of the binary coding. Then, the hidden output $[c_1, c_2, \ldots, c_m]$ of the first layer is regarded as the input data of the second layer for training the second AE. It is important to note that the weight of the first layer does not change owing to the training process in the second AE. Thus, a deep SAE neural network is generated. The output of the SAE represents the binary coding feature based on the RF and serves as the input for the classifier for the fault diagnosis of HVCBs.

D. General Procedure of the Proposed Method

In this paper, a novel feature transform and compression method for HVCB fault diagnosis is developed. The framework of our proposed method is shown in Fig. 7 and the general steps are summarized as follows.

Step 1: Under various fault scenarios, the vibration data of the HVCB are collected using the acquisition system.
Step 2: WPT is used to obtain the information of the vibration signals and create a feature space using WTFER.
Step 3: Based on the RF algorithm, a binary coding feature space is used to abstract the original feature description.
Step 4: The SAE is adopted to reduce the dimensionality of the coding feature space; then, the compressed feature space is obtained.

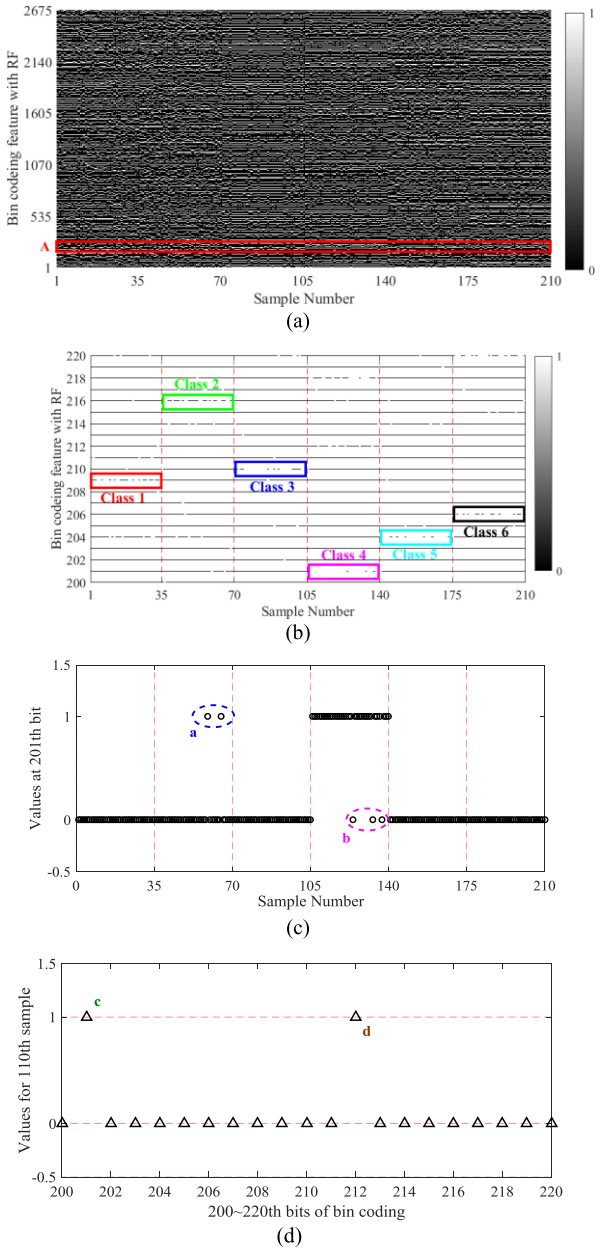

Fig. 9. Feature description based on RF transformation. (a) Goal graph. (b) Local graph. (c) Curve at 201st bit for all training datasets. (d) Bits 200–220th of the bin coding for the 110th sample of the training dataset.

Step 5: The compressed, clustering features are inputted into a Softmax algorithm for fault classification.
Step 6: A network combining SAE with the Softmax algorithm is retrained for fine tuning the weights of the network for fault classification.
Step 7: The performance of the proposed method is validated using the testing samples.

IV. Experimental Verification

Our proposed method was implemented using MATLAB and applied for fault diagnosis of HVCBs. As mentioned in Section II, 35 samples that are randomly selected for each class

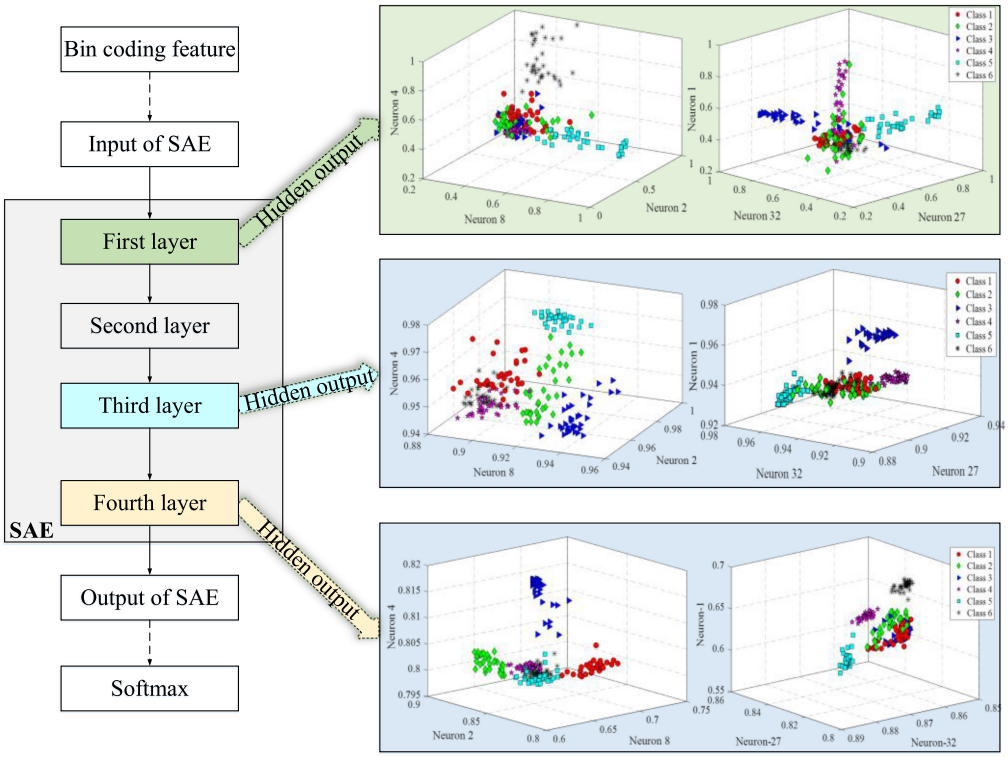

Fig. 10. Feature compression process based on SAE.

are applied to train the proposed model. Then, the model was tested using the remaining samples. The parameters used in our experiments are listed in Table II. The diagnosis performance of the proposed method is discussed in Section IV-A. Furthermore, in Section IV-B, the use of various approaches (Softmax, DT, OCSVM, BPNN, and RF) under different feature spaces (WTFER, RFBC, and RFBC+SAE) are discussed and compared to highlight the improvement results provided by our proposed method.

A. Performance of the Proposed Method

In order to evaluate the performance of the proposed approach, the evolution process of the feature space in the case of the WTFER space, bin coding space based on the RF, and compression space-based SAE are evaluated; these are shown in Figs. 8–10, respectively. In addition, the diagnosis results with the training and test datasets are shown in Fig. 11.

In particular, Fig. 8 shows the training data distribution for six classes in the case of the six-dimensional WTFER feature space. It can be distinctly seen that most samples are clustered together except in the case of Classes 2 and 4. However, it is difficult to build a better classifier to realize satisfactory identification precision in such feature spaces, especially using some linear classifiers.

In Fig. 9(a), the bin coding feature space and feature transformation based on the RF are shown. The horizontal axis represents the sample number, i.e., 1–35th, 36–70th, 71–105th, 106–140th, 141–175th, and 176–210th belonging to Classes 1–6, respectively, whereas the vertical axis describes the bits of the bin coding or node ranking in the RF. In this paper, we assume that the scale of the DT is 100, which indicates that 100 bits of the bin coding are 1, whereas the others are 0. In Fig. 9(b)–(d), the coding result using RF derived from the local region of the bin coding is displayed. For example, the 201st bit of Class 4 samples (106–140th) is 1, whereas others are 0 [see Fig. 9(c)]. Similarly, the 209th, 216th, 210th, 204th, and 206th bits can primarily be used for distinguishing other classes, as depicted in Fig. 9(b). Fig. 9(d) lists the output of the 200–220th bits for the 110th sample (which belongs to Class 4) and shows that the 201st and 212th bits of the coding feature being 1 indicate that they belong to Class 4 with high probability. In particular, if the 201st or 210th bits of a sample are 1, the sample may be defined as Class 4. Considering Fig. 9(b) and (c), we found the 201st bit is a better feature for differentiating Class 4 samples compared with the 212th bit. This indicates that although bin coding-based RF successfully establishes some useful features for HVCB diagnosis, superfluous dimensionality also exists in feature space. Thus, it is crucial to eliminate redundant information to further improve diagnostic accuracy.

The bin coding features are imported into the SAE. The output of SAE will span the proposed hybrid feature space and be applied to a classifier, as shown in Fig. 10. In particular, Fig. 10 shows the three-dimensional representations of high-dimensional feature maps at different layers in the SAE. Although a comparison between the feature maps involves unavoidable errors due to the loss of information during dimensionality reduction, it is obvious that diverse classes are heavily overlapped at the input layer, whereas they become more separable at deeper layers. Furthermore, specifically, the last layer

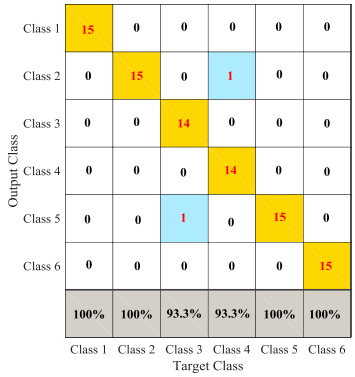

Fig. 11. Diagnosis results of using the proposed method.

feature can completely identify the various classes from at least one dimensionality of the SAE feature space, i.e., the diagnosis precision will be higher compared to without the use of SAE.

Fig. 11 shows the multiclass confusion matrix of the proposed method with a Softmax classifier. The multiclass confusion matrix records the classification results of all the conditions in a detailed manner, including both classification information and misclassification information. The horizontal axis of the confusion matrix represents actual label of classification, whereas the vertical axis represents the predicted label; the element value $N_{i,j}$ of the confusion matrix represents the matching or mismatching sample number between ith target class and predicted label. It should be noted that the elements on the main diagonal are the same, whereas others are not. In addition, the element in the last row of the confusion matrix represents the classification accuracy of each condition. In the case shown in Fig. 11, we can observe that the lowest precision is 93.3% which occurs for Conditions 3 and 4. Furthermore, for all test samples, the accuracy of the proposed method is 97.8%, which indicates that the proposed method has a better performance rate for HVCB fault diagnosis.

B. Comparison With Traditional Methods

Five classification methods (Softmax, DT, OCSVM, BPNN, and RF) for three feature spaces (WTFER, RFBC, and RFBC+SAE) are applied to diagnose the test dataset. In the evaluation system of classifiers, classification accuracy rate is taken as the key performance metric in this work [8]. For completeness, and to ensure that the accuracy values are a reliable indicator of overall performance, the precision and recall of all classifiers are also investigated. Furthermore, F-measure is another widely used criterion, which contains both the precision rate and recall rate [36]. Therefore, the accuracy rate and F-measure are applied to evaluate the performance of different feature spaces with various classifiers in this paper, as shown in Fig. 12.

Based on these results, the following inferences can be drawn.

1) After the RFBC, RFBC+SAE diagnostic process, the F value of the SoftMax method is continuously increased substantially under the six classes, and the F values of most other methods are also increased, which indicates that the feature transformation method described in this paper is conducive to diagnosis.
2) The highest accuracy of 97.8% is obtained using the proposed method (Softmax+RFBC+SAE), which is significantly higher than the accuracy rates of 87.8%, 91.1%, 94.4%, and 93.3%, using the DT, OCSVM, BPNN, and RF+RFBC+SAE methods, respectively.
3) After feature transformation with RF, the accuracy of the single classification is considerably improved; however, the ensemble learning RF is only slightly enhanced. This can be attributed to the fact that the ensemble learning was performed using the RF feature transformation. Based on the RFBC feature space, an RF model is an iterative method of classification.
4) Based on the SAE, feature compression can deeply fuse the effective information for diagnosis, reasonably reducing the unfavorable effect owing to ensemble feature transformation (i.e., RFBC) and further improving the classification performance.

In order to avoid a perceived bias based on the experiment, ten trials were conducted to compare the diagnosis performance in

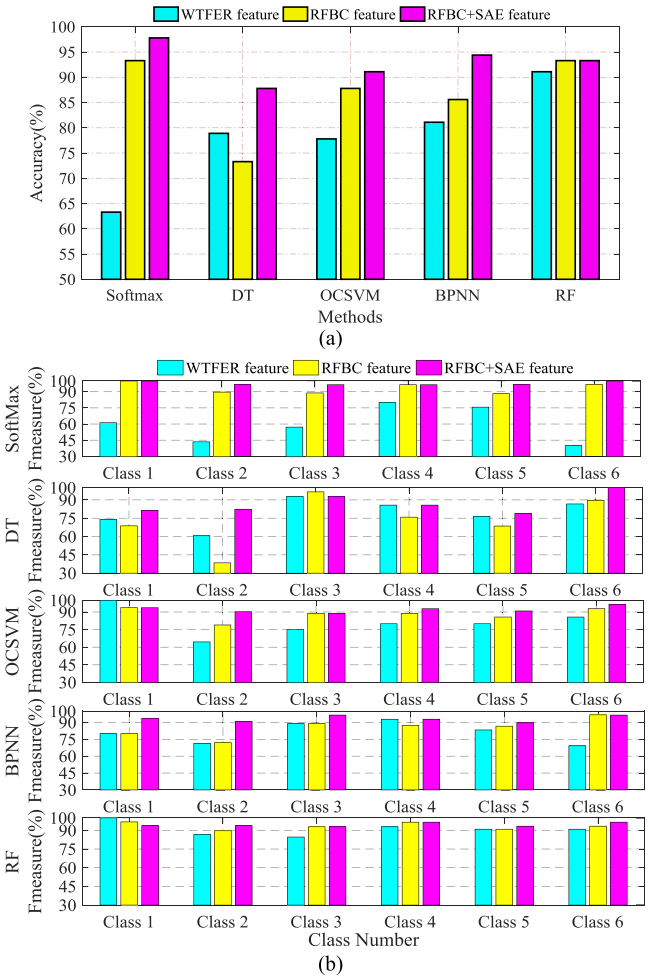

Fig. 12. Comparison results. (a) Accuracy for different feature spaces with various classifiers. (b) F-measure for different feature spaces with various classifiers.

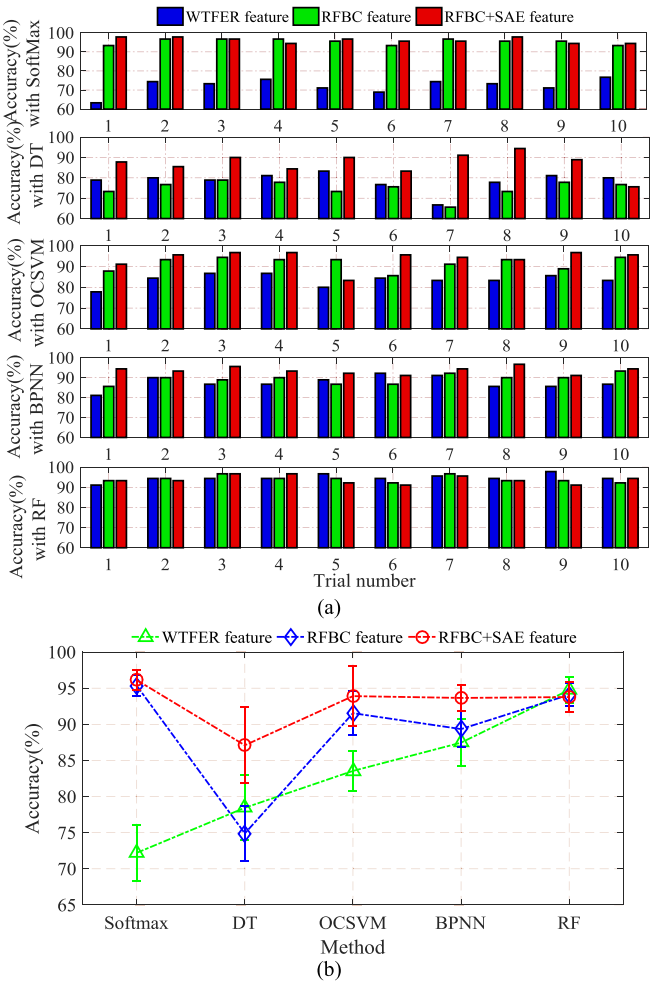

Fig. 13. Comparison results of ten trials using different methods. (a) Diagnosis accuracy of ten trials using different methods. (b) Average accuracy and standard deviation curves of ten trials using different methods.

TABLE III
AVERAGE ACCURACY AND STANDARD DEVIATION VALUES OF VARIOUS CLASSIFICATIONS WITH DIFFERENT FEATURE SPACES FOR TEN TRIALS

Parameter	Feature space	Value				
		Softmax	DT	OCSVM	BPNN	RF
Average accuracy	WTFER	72.21	78.45	83.55	87.46	94.76
	RFBC	95.35	74.90	91.54	89.34	94.09
	RFBC +SAE	96.12	87.10	93.90	93.65	93.77
Standard deviation	WTFER	3.89	4.53	2.82	3.22	1.76
	RFBC	1.49	3.86	3.08	2.44	1.60
	RFBC +SAE	1.43	5.23	4.12	1.83	2.06

various feature spaces using different classification approaches. In each trial, we randomly select training and test data from the raw dataset in a ratio of 7:3. Fig. 13(a) shows the detailed diagnosis results for each trial; in addition, the average accuracy and standard deviation of the five classification methods with three feature spaces in all the trials are listed in Table III and shown graphically in Fig. 13(b). It is clear that the accuracy of the Softmax+RFBC+SAE method is the highest in the case of most trials; in particular, its average accuracy is the highest, standard deviation is the least, and performance is best than the other approaches.

V. CONCLUSION

It is important to develop a good set of features for fault diagnosis of HVCBs. Thus, in this paper, a feature transformation and compression method was proposed. First, using the traditional WPT time–frequency analysis, we developed the basic characteristic description. Then, based on ensemble learning and random subspace technology, a RF was applied to coding a binary feature vector to extend the feature width. In addition, the SAE neural network was used to compress the binary feature using feature depth. Finally, an optimal feature space represented the information of HVCB under various fault types.

The usefulness of feature learning was verified by a comparison among five machine learning methods in a series of experiments. Experimental results indicated that the average accuracy in ten trials for the hybrid feature space (Softmax+RFBC+SAE) performed better than the other classification methods in WTFER, RFBC, and other hybrid feature spaces. In addition, the optimal feature could more effectively reduce a classifier's burden and improve the diagnosis result of HVCB.

Although the method proposed in this paper involved a reliable feature transformation approach and was verified against an entire set of diagnosis technology approaches for HVCB fault diagnosis, the proposed method's parameters will further be discussed. Furthermore, the proposed method was established on manual feature extraction. In a future work, typical fault feature automatic extraction will also be considered.

REFERENCES

[1] A. Janssen, D. Makareinis, and C. E. Solver, "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 SF6 circuit breakers by statistical analysis on the basis of the field data," *Elect. Power Syst. Res.*, vol. 103, pp. 105–113, 2013.
[3] W. Niu, G. Liang, H. Yuan, and L. Baoshu, "A fault diagnosis method of high voltage circuit breaker based on moving contact motion trajectory and ELM," *Math. Probl. Eng.*, vol. 2016, pp. 1–10, 2016.
[4] A. Forootani, A. A. Afzalian, and A. N. Ghomsheh, "Model-based fault analysis of a high-voltage circuit breaker operating mechanism," *Turkish J. Elect. Eng. Comput. Sci.*, vol. 25, pp. 2349–2362, 2016.
[5] M. Fei, P. Yi, K. Zhu, and J. Zheng, "On-line hybrid fault diagnosis method for high voltage circuit breaker," *J. Intell. Fuzzy Syst.*, vol. 33, pp. 1–12, 2017.
[6] K. Zhu, F. Mei, and J. Zheng, "Adaptive fault diagnosis of HVCBs based on P-SVDD and P-KFCM," *Neurocomputing*, vol. 240, pp. 127–136, 2017.
[7] H. E. Mengyuan *et al.*, "Research of circuit breaker intelligent fault diagnosis method based on double clustering," *IEICE Electron. Express*, vol. 14, no. 17, pp. 1–10, 2017.
[8] J. Huang, X. Hu, and F. Yang, "Support vector machine with genetic algorithm for machinery fault diagnosis of high voltage circuit breaker," *Measurement*, vol. 44, pp. 1018–1027, 2011.
[9] 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, 2016, Art. no. 322.

[10] M. Liu, K. Wang, L. Sun, and J. Zhen, "Applying empirical mode decomposition (EMD) and entropy to diagnose circuit breaker faults," *Optik*, vol. 126, pp. 2338–2342, 2015.
[11] J. Zhang, M. Liu, K. Wang, and L. Sun, "Mechanical fault diagnosis for HV circuit breakers based on ensemble empirical mode decomposition energy entropy and support vector machine," *Math. Probl. Eng.*, 2015, Art. no. 101757.
[12] 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, 2016, Art. no. 1887.
[13] J. B. Allen and L. Rabiner, "A unified approach to short-time Fourier analysis and synthesis," *Proc. IEEE*, vol. 65, no. 11, pp. 1558–1564, Nov. 1977.
[14] C. Gargour, M. Gabrea, V. Ramachandran, and J.-M. Lina, "A short introduction to wavelets and their applications," *IEEE Circuits Syst. Mag.*, vol. 9, no. 2, pp. 57–68, Second quarter 2009.
[15] Y. Wang, G. Xu, L. Lin, and K. Jiang, "Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 54/55, pp. 259–276, 2015.
[16] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Process.*, vol. 96, pp. 1–15, 2014.
[17] A. Kumar and R. Kumar, "Time-frequency analysis and support vector machine in automatic detection of defect from vibration signal of centrifugal pump," *Measurement*, vol. 108, pp. 119–133, 2017.
[18] M. Zhao, M. Kang, B. Tang, and M. Pecht, "Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4290–4300, May 2018.
[19] J. D. Rennie, "Regularized logistic regression is strictly convex," 2005, unpublished. [Online]. Available: people.csail.mit.edu/jrennie/writing/convexLR.pdf.
[20] M. Seera, C. P. Lim, and K. L. Chu, "Motor fault detection and diagnosis using a hybrid FMM-CART model with online learning," *J. Intell. Manuf.*, vol. 27, pp. 1–13, 2016.
[21] A. T. Azar and S. M. El-Metwally, "Decision tree classifiers for automated medical diagnosis," *Neural Comput. Appl.*, vol. 23, pp. 2387–2403, 2013.
[22] K. Vernekar, H. Kumar, and K. V. Gangadharan, "Engine gearbox fault diagnosis using empirical mode decomposition method and naïve Bayes algorithm," *Sādhanā*, vol. 42, pp. 1–11, 2017.
[23] T. Boukra, A. Lebaroud, and G. Clerc, "Statistical and neural-network approaches for the classification of induction machine faults using the ambiguity plane representation," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 4034–4042, Sep. 2013.
[24] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3398–3407, Aug. 2013.
[25] N. 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, 2015, Art. no. 7.
[26] J. Ni, C. Zhang, and S. X. Yang, "An adaptive approach based on KPCA and SVM for real-time fault diagnosis of HVCBs," *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 1960–1971, Jul. 2011.
[27] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
[28] G. Biau, L. Devroye, and G. Lugosi, "Consistency of random forests and other averaging classifiers," *J. Mach. Learn. Res.*, vol. 9, pp. 2015–2033, 2009.
[29] A. Verikas, A. Gelzinis, and M. Bacauskiene, "Mining data with random forests: A survey and results of new tests," *Pattern Recognit.*, vol. 44, pp. 330–349, 2011.
[30] S. Ma *et al.*, "Intelligent fault diagnosis of HVCB with feature space optimization-based random forest," *Sensors*, vol. 18, no. 4, 2018, Art. no. 1221.
[31] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, and C. Li, "Fault diagnosis in spur gears based on genetic algorithm and random forest," *Mech. Syst. Signal Process.*, vol. 70/71, pp. 87–103, 2016.
[32] M. Seera, M. L. D. Wong, and A. K. Nandi, "Classification of ball bearing faults using a hybrid intelligent model," *Appl. Soft Comput.*, vol. 57, pp. 427–435, 2017.
[33] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727–2736, Mar. 2018.
[34] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi, "Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features," *Mech. Syst. Signal Process.*, vol. 99, pp. 459–477, 2018.
[35] K. Li *et al.*, "Multi-label spacecraft electrical signal classification method based on DBN and random forest," *PLoS One*, vol. 12, no. 5, 2017, Art. no. e0176614.
[36] H. Shao *et al.*, "Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing," *Mech. Syst. Signal Process.*, vol. 100, pp. 743–765, 2018.

Suliang Ma (GS'17) was born in Anshan, China, in 1988. He received the B.S. degree in electrical engineering and its automation and the M.S. degree in control engineering from the China University of Mining and Technology, Beijing, China, in 2011 and 2014, respectively. He is currently working toward the Ph.D. degree in electric machines and electric apparatus at Beijing University of Aeronautics and Astronautics, Beijing.

From 2012 to 2013, he was a Graduate Student Researcher with the China Electric Power Research Institute, focusing on research, optimization, and design of the smart grid with wind, photovoltaic and energy storage. His research interests include data-driven prognostics and health management for high-voltage switchgear, automeasurement technology, control theory and application, power electronics technology.

Mingxuan Chen (GS'17) was born in China in 1989. He received the B.S. degree from the China University of Mining and Technology, Beijing, China, in 2011, and the M.S. degree from the Inner Mongolia University of Technology, Huhhot, China, in 2014, both in electrical engineering and power electronics. He is currently working toward the Ph.D. degree in electrical engineering at Behihang University, Beijing.

From 2012 to 2013, he was a Graduate Student Researcher with China Electric Power Research Institute, focusing on research of photovoltaic system and energy storage. His research interests include power electronics technology, and power electronics for the renewable energy system.

Jianwen Wu (M'15–SM'17) received the B.S. and M.S. degrees from the Shenyang University of Technology, Shen Yang, China, in 1984 and 1987, respectively, and the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 1995, all in electrical engineering.

From 1995 to 1998, he was a Postdoctoral Fellow with the Huazhong University of Science and Technology, Wuhan, China. Since 2001, he has been a Member of faculty of the School of Automation Science and Electrical Engineering, Beijing, China, where he is currently a Professor. His research interests include vacuum arc theory, intelligent electrical apparatus, and power electronics technology.

Dr. Wu is a Senior Member of China Electrotechnical Society.

Yuhao Wang was born in Yuncheng, China, in 1993. He received B.S. degree in automation from the Beijing University of Civil Engineering and Architecture, Beijing, China, in 2015. He is currently working toward the M.S. degree in electric machines and electric apparatus at Beijing University of Aeronautics and Astronautics, Beijing.

His research interests include intelligent electrical apparatus and automeasurement technology.

Bowen Jia was born in Beijing, China, in 1992. He received the M.S. degree in electrical engineering in 2017 from Beihang University, Beijing, where he is currently working toward the Ph.D. degree in electric machines and electric apparatus.

His research interests include arc fault diagnosis, plasma characteristics, and the mechanism of dc arc interruption in hydrogen and mixed gases, the design of high-powered dc relays, and dc breaker technology.

Yuan Jiang (M'17) was born in Dandong, China, in 1985. He received the B.S. and M.S. degrees in materials science and engineering and the Ph.D. degree in electrical engineering from Beihang University, Beijing, China, in 2008, 2011, and 2016, respectively.

From July 2016 to May 2017, he was with the China North Industries Group Corporation. He is currently a Postdoctoral Fellow with Beihang University. His research interests include the theory and application of vacuum arcs and the control of intelligent electrical apparatuses.