IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 15, 2021, accepted August 22, 2021, date of publication August 27, 2021, date of current version September 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3108345

Running Status Diagnosis of Onboard Traction Transformers Based on Kernel Principal Component Analysis and Fuzzy Clustering

JUNMIN ZHU[1], SHUAIBING LI[2], (Member, IEEE), AND HAIYING DONG[2], (Member, IEEE)
[1]School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, China
[2]School of New Energy and Power Engineering, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, China

Corresponding authors: Shuaibing Li (lishuaibing1105@163.com) and Haiying Dong (hydong@mail.lzjtu.cn)

This work was supported in part by the Tianyou Youth Talent Lift Program of Lanzhou Jiaotong University, in part by the University Innovation Fund Project of Gansu Provincial Department of Education under Grant 2020A-036, and in part by the Youth Science Fund Project of Lanzhou Jiaotong University under Grant 2019029.

ABSTRACT The onboard traction transformer is a critical equipment of high-speed trains, its running state directly affects the safety and stability of a train's operation. Given the complexity of the running condition of the onboard traction transformer, this paper proposes a running state diagnosis algorithm based on kernel principal component analysis (KPCA) and fuzzy clustering. To fully extract the status information of the onboard traction transformer, the aging characteristics of insulating oil and main insulation are analyzed under different running mileage as the first step. Thereby, to eliminate the signal redundancy, the status feature set of the onboard traction transformer is analyzed by KPCA combined with the characteristic quantities of the traditional dissolved gas analysis (DGA), and the eigenvalues with the contribution rate of over 95% are used as new eigenvectors. Finally, a status diagnosis model is established by using fuzzy clustering analysis, considering the limitations of fault data of onboard traction transformer. The results from field collected data show that the proposed method is effective in diagnosing the running status of the onboard traction transformer.

INDEX TERMS Onboard traction transformer, running status diagnosis, insulation aging, kernel principal component analysis, fuzzy clustering.

I. INTRODUCTION

Onboard traction transformer is an important component of the power supply system of high-speed trains. According to statistics, among the railway power supply equipment failures, the faults caused by traction transformer account for more than 87.3% [1]. To ensure the safety and stable operation of trains, the railway department mainly adopts two strategies: daily maintenance and fault maintenance. However, some problems still exist, such as low maintenance efficiency, lack of human experience, and high safety risk [2], [3]. With the increase of the operation speed and operation density of high-speed railways, higher requirements are put forward for the operation and maintenance levels as well as detection means of onboard traction transformers. Recently, scholars have paid much attentions on oil-immersed power transformers [4]–[7]. However, considering the onboard traction transformer in the particularity of the internal structure and external environment [8], the fault diagnosis and running status analysis algorithm of power transformers is no longer suitable for onboard traction transformers. The intelligent fault diagnosis of onboard traction transformers mostly remains in the analysis and positioning after the problem occurs [9], [10], and there are few diagnostic algorithms in the running states of "Good", "Fair", "Poor", and "Faulty".

To realize the status diagnosis of an onboard traction transformer in its whole operation life, the extraction of status information is the primary premise. Jiang *et al.* [11], based on the partial discharge characteristics of the main insulating bushing, conducted UHF tests for different damaged bushings. Meanwhile, time-frequency domain signal characteristics were taken as state information, but only casing damage was considered as feature information, resulting in problems such as inadequate feature characterization and insensitivity

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

to weak features. Tao *et al.* [12] applied the wavelet decomposition and entropy theory to extract the detail component of transformer mechanical vibration signal as state feature information. However, it is not applicable due to the influence of the external environment and operation conditions. Liu *et al.* [13] analyzed the correlation coefficient of dissolved gas in transformer insulating oil to determine the fault type, but it is easily affected by individual differences.

In terms of the transformer's state diagnosis algorithm, Li *et al.* [14] proposed establishing the transformer fault diagnosis model using a neural network and decision tree. However, due to the limited real data of transformer fault samples and the need to train the diagnosis algorithm in advance, there were some problems, including "poor" practicability and difficult selection of training parameters. Lee *et al.* [15] established an expert system using the fault state mapping data set of oil-immersed transformers. However, considering the influence of individual differences of transformers, it is lack of promotion significance. Zhang *et al.* [16] proposed an improved multivariable support vector algorithm to diagnose the five fault states of transformers, but it is relatively difficult to optimize the parameters by using dissolved gas analysis (DGA) as its feature.

In terms of the latest transformer's fault diagnosis technology, Ghoneim *et al.* [17] introduced gas concentration percentage limit and gas ratio as characteristic quantity based on transformer dissolved gas analysis, and combined with teaching-learning based optimization model to diagnose transformer fault. To improve the diagnostic accuracy, the comprehensive index of DGA features is considered. Ward *et al.* [18] combined DGA features with partial discharge features to analyze the fault types of power transformers, which overcomes the shortcoming of insufficient individual fault character. Aarathy *et al.* [19] carried out a spectral analysis on the insulating oil of power transformers and established a connection between the spectral change results and the running state, which helps to establish the aging characteristic index of power transformers.

Unlike traditional power transformers, onboard traction transformer has a high coupling degree in its internal structure and is often affected by harsh environment, which finally results in high accident rates and significant security risks [9]. With above deficiencies, a single characteristic index is unable to fully characterize the running status characteristics, an comprehensive analysis of the status feature is needed. Considering fuzzy clustering analysis is based on the similarity between sample objects [20], and it has the characteristics of small sample analysis and does not require advanced training, which has also been widely applied to the status diagnosis of aero-engine and rolling bearing equipment [21], [22].

Based on the above analysis, this paper proposes a diagnosis algorithm of an onboard traction transformer's running status by combining the kernel principal component analysis (KPCA) theory and fuzzy clustering analysis. Firstly, to extract the effective state characteristics of onboard traction transformers, the relationship between the conductivity of insulating oil, furfural of insulating oil, electrical strength of insulating oil, moisture in insulating oil, the temperature of insulating oil, and train operating mileage is analyzed. Once a linear relationship is found, it will be selected as the status feature. Secondly, the time-domain characteristics of the main dielectric spectrum are taken as the status feature, and the initial status feature set is constructed with DGA as feature. After processing by KPCA, the eigenvalue with a contribution rate of more than 95% is selected as the running state feature vector without the loss of signal characteristics. Finally, the fuzzy clustering analysis algorithm is used to input the feature vectors under different running status into the onboard traction transformer's state diagnosis model, and the dynamic clustering diagram is formed to realize real-time state diagnosis. The verification and analysis results show that the proposed method can accurately diagnose the running status of the onboard traction transformer.

The main contributions of this paper are as follows: 1) the diagnostic algorithm for classifying the onboard traction transformer in the state of "Good", "Fair", "Poor" and "Fault" is established, which can not only judge the "Fault", but also has the ability to distinguish the state between "Good" and "Fault". 2) By combining the aging characteristics of insulating oil and main insulation with DGA characteristics, some weak features of onboard traction transformers in the different running statuses can be fully characterized. 3) The advantage of fuzzy clustering analysis in small sample analysis and intuitive expression is applied to effectively and accurately diagnose the running status of onboard traction transformer without training.

II. STATE DIAGNOSIS ALGORITHM

A. KERNEL PRINCIPAL COMPONENT ANALYSIS

To solve the problem of feature sets composed of nonlinear data, such as low correlation and large dimension, the KPCA is used to map the data set into linear space [23]. After linear principal component analysis (PCA), the data features are filtered and sorted. The specific steps of the KPCA algorithm are as follows:

1) Suppose the nonlinear mapping matrix is φ, and the column vector of the data feature set $\boldsymbol{Q}$ is $\boldsymbol{x}_i (i = 1, 2, \ldots N)$, then the covariance matrix $\boldsymbol{S}$ corresponding to feature set $\boldsymbol{Q}$ can be expressed as:

$$\boldsymbol{S} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\varphi}(\boldsymbol{x}_i)\boldsymbol{\varphi}(\boldsymbol{x}_i)^T \tag{1}$$

2) In a high-dimensional linear space, the feature set $\boldsymbol{R}$ is processed by PCA, and the mapping matrix $\boldsymbol{\varphi}$ is multiplied by both ends of the equation:

$$\lambda \boldsymbol{V} = \boldsymbol{S}\boldsymbol{V} \tag{2}$$

$$\lambda[\boldsymbol{\varphi}(\boldsymbol{x}_i)\boldsymbol{V}] = \boldsymbol{\varphi}(\boldsymbol{x}_i)\boldsymbol{S}\boldsymbol{V}, i = 1, 2, \cdots N \tag{3}$$

where $\boldsymbol{V}$ is the eigenvector of the high-dimensional space, λ is the corresponding eigenvalue.

3) Linear correlation is used to express the feature vector $\boldsymbol{V}$ as (4), where $\boldsymbol{a}$ is the correlation coefficient matrix.

$$\boldsymbol{V} = \sum_{k=1}^{N} \alpha_k \boldsymbol{\varphi}(\boldsymbol{x}_k) = \boldsymbol{\alpha\varphi} \tag{4}$$

The kernel matrix $\boldsymbol{K}$ is defined:

$$\boldsymbol{K}_{ij} = \left[\boldsymbol{\varphi}(x_i)\boldsymbol{\varphi}(x_j)\right] \tag{5}$$

Combining Equations (1), (4), and (5) and substituting Equation (3), one can obtain:

$$N\lambda \boldsymbol{K\alpha} = \boldsymbol{KK\alpha} \Rightarrow N\lambda\boldsymbol{\alpha} = \boldsymbol{K\alpha} \tag{6}$$

Thus, the eigenvalue of the kernel matrix $\boldsymbol{K}$ is the eigenvalue of the covariance matrix $\boldsymbol{S}$, and the direction of the principal component of the linear space can be obtained.

4) In this paper, the radial basis kernel function is selected to construct the kernel matrix, and the cumulative contribution rate of eigenvalues n_p is defined:

$$k(x_i, y_j) = \exp[-\left\|x_i - y_j\right\|^2/(2\sigma^2)] \tag{7}$$

$$n_p = \sum_{j=1}^{p} \lambda_j / \sum_{i=1}^{11} \lambda_i \tag{8}$$

The eigenvalues of the kernel matrix $\boldsymbol{K}$ are sorted in descending order, and the cumulative contribution rate of more than 95% is taken as the new eigenvector. Thus, the signal redundancy and characteristic dimensions are reduced without the loss of signal features.

B. FUZZY CLUSTERING ANALYSIS

Fuzzy clustering analysis is based on fuzzy mathematics, which uses the membership similarity between samples to complete their clustering analysis. It has the advantages of intuitive expression, small sample analysis, and no need for training in advance. The specific steps are as follows:

1) FUZZY MATRIX X

In the sample set $\boldsymbol{U} = \{\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots \boldsymbol{x}_n\}$, the fuzzy matrix $\boldsymbol{X}$ is established with m feature indexes in each sample:

$$\boldsymbol{x_i} = (x_{i1}, x_{i2}, \cdots, x_{im}) \quad (i = 1, 2, \cdots n) \tag{9}$$

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{10}$$

C. FUZZY STANDARD MATRIX X″

Under the influence of different dimensions and properties of characteristic indexes, the fuzzy matrix $\boldsymbol{X}$ needs to be standardized in the following two ways:

Translation • Standard deviation

$$x'_{ik} = \frac{x_{ik} - \bar{x}_k}{s_k} \quad (i = 1, 2, \cdots, n; k = 1, 2, \cdots, m) \tag{11}$$

where $\bar{x}_k = \frac{1}{n}\sum\limits_{i=1}^{n} x_{ik}$, $s_k = \sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}(x_{ik} - \bar{x}_k)^2}$.

Translation · Range

$$x''_{ik} = \frac{x'_{ik} - \min\limits_{1\le i\le n}\{x'_{ik}\}}{\max\limits_{1\le i\le n}\{x'_{ik}\} - \min\limits_{1\le i\le n}\{x'_{ik}\}} \quad (k = 1, 2, \cdots, m) \tag{12}$$

Through the above transformation, the fuzzy standard matrix $\boldsymbol{X}''$ is obtained, which satisfies all the characteristic indexes $x''_{ij} \in [0, 1]$.

1) FUZZY SIMILARITY MATRIX R

The similarity coefficient between characteristic indexes is defined as $r_{ij} = R(x_i, x_j)$. Generally, the fuzzy equivalent matrix is established by the method of quantity product, angle cosine, and correlation coefficient, as shown in (13). In this paper, the fuzzy similarity matrix $\boldsymbol{R}$ is established by the exponential similarity coefficient method with more obvious features:

$$r_{ij} = \frac{1}{m}\sum_{k=1}^{m}\exp[-\frac{3}{4}\cdot\frac{(x_{ik} - x_{jk})^2}{s_k^2}] \tag{13}$$

where $s_k = \frac{1}{n}\sum\limits_{i=1}^{n}(x_{ik} - \bar{x}_{ik})^2, \bar{x}_k = \frac{1}{n}\sum\limits_{i-1}^{n} x_{ik}(k = 1, 2, \cdots m)$.

2) FUZZY EQUIVALENCE MATRIX R*

To make the fuzzy similarity matrix $\boldsymbol{R}$ transitive, when the transfer packet $t(\boldsymbol{R}) = \boldsymbol{R}^{2K}$ has a natural positive integer K such that $\boldsymbol{R}^{2K} = \boldsymbol{R}^{2(K+1)}$, $t(\boldsymbol{R})$ is the corresponding fuzzy equivalent matrix $\boldsymbol{R}^*$. The transfer packet algorithm is shown in (14).

$$\boldsymbol{R}^2 = \boldsymbol{R} \circ \boldsymbol{R}, \boldsymbol{R}^4 = \boldsymbol{R}^2 \circ \boldsymbol{R}^2, \cdots, \tag{14}$$

3) FUZZY BOOLEAN MATRIX

As shown in (15), the confidence factor $\lambda \in [0, 1]$ is used to determine the 0 and 1 distributions of each characteristic index in the fuzzy equivalence matrix $\boldsymbol{R}^*$, and then the corresponding fuzzy equivalence matrix $\boldsymbol{R}_\lambda$ is formed.

$$r_{ij} = \begin{cases} 1, & r_{ij} \ge \lambda \\ 0, & r_{ij} < \lambda \end{cases} \tag{15}$$

The sample set mentioned above can be summarized as X→ $\boldsymbol{X}'' \to \boldsymbol{R} \to \boldsymbol{R}^* \to \boldsymbol{R}_\lambda$. In the fuzzy Boolean matrix $\boldsymbol{R}_\lambda$, when λ is given a specific value, the same column vector of the sample set object is classified into one class. As λ changes from 1 to 0, the number of classifications gradually decreases. Finally, the sample objects are grouped into one category, thus forming a dynamic clustering diagram to intuitively express the classification situation.

D. DIAGNOSTIC EVALUATION CRITERIA

To realize the evaluation of the status diagnosis results, the status recognition rate is used as the evaluation standard, and the algorithm is shown in (16).

$$Accuracy = \frac{p}{t} \times 100\% \tag{16}$$

where t is the number of samples in the test set, and p is the number of samples with a correct diagnosis.

III. RUNNING STATE DIAGNOSIS OF ONBOARD TRACTION OF STATE CATEGORIES

A. SELECTION OF STATE CHARACTERISTIC QUANTITIES AND CLASSIFICATION OF STATE CATEGORIES

As the power source of the high-speed train, the onboard traction transformer is mainly used to convert the 25 kV catenary voltage to a low voltage. Its running state is the basis to ensure the efficient, safe, and stable running of the train. Restricted by the code for design of railway lines (GB 50090-2006) [24], the onboard traction transformer installed at the bottom of the train is affected not only by the external adverse environment, such as mechanical vibration, climate conditions, and running time, but also by internal operating conditions, such as voltage fluctuation, excitation current, and poor heat dissipation, which make it different from the general power transformer. The onboard traction transformer is composed of a high-voltage winding and four windings, and the coupling degree between the components is high. To realize the extraction of its running state characteristics, a single DGA is used to diagnose the state, which does not apply to the onboard traction transformer.

Given the structural particularity of onboard traction transformers, this paper analyzes the aging characteristics of insulating oil and main insulation, analyzes the correlation between the characteristics and the running status of a test transformer, and uses DGA to characterize the state characteristic quantity of an onboard traction transformer. In the aging characteristics of the main insulation, the characteristic quantity is obtained by the frequency domain analysis of its dielectric spectrum. In the aging characteristics test and DGA analysis of insulating oil, the state characteristics included are shown in Table 1, in which $f_1 \sim f_5$ are the aging characteristics of insulating oil, while $d_1 \sim d_8$ stands for dissolved gas in the oil.

According to the above state characteristic indexes of an onboard traction transformer and the state-level classification standard in IEC60422-2013 [25], the state level of an onboard traction transformer is divided into four statuses: "Good", "Fair", "Poor", and "Fault".

B. STATUS DIAGNOSIS PROCESS

Given the different aging characteristics of insulating oil and the main insulation of the onboard traction transformer in different running states, this paper proposes combining DGA to extract state features. After processing by KPCA, fuzzy clustering analysis is used to complete the running status diagnosis of the onboard traction transformer. The specific diagnosis flow chart is shown in Figure 1.

TABLE 1. The characteristic quantity of the aging of insulating oil of an onboard traction transformer.

Feature code	State feature	Feature code	State feature
f_1(Ω.m)	conductivity	d_3 (μL/L)	CO_2
f_2 (mg/L)	furfural	d_4 (μL/L)	CH_4
f_3 (kV)	electrical strength	d_5 (μL/L)	C_2H_6
f_4 (ppm)	moisture	d_6 (μL/L)	C_2H_4
f_5 (°C)	temperature	d_7 (μL/L)	C_2H_2
d_1 (μL/L)	H_2	d_8 (μL/L)	total hydrocarbon
d_2 (μL/L)	CO		

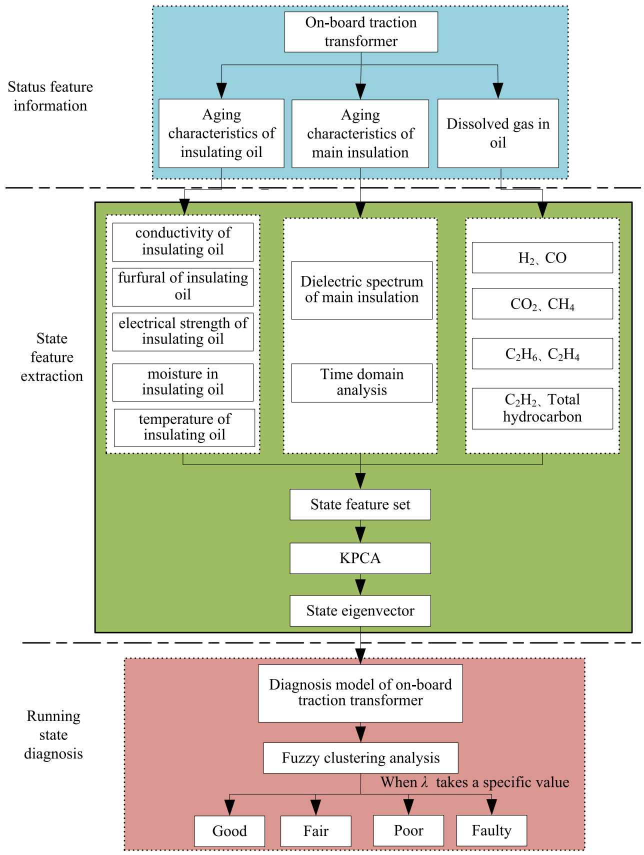


FIGURE 1. Flow chart of an onboard traction transformer's running state diagnosis.

1) STATUS FEATURE INFORMATION

Insulating oil and windings are important components of onboard traction transformers. The characteristics of the insulating oil and windings of the traction transformers of high-speed trains under different maintenance levels are tested, consistency between their state characteristics and maintenance levels is also analyzed. The information that can reflect the status of onboard traction transformers is obtained by combining with DGA.

2) STATE FEATURE EXTRACTION

During studying on the aging characteristics of insulating oil, the conductivity, furfural, electrical strength, moisture, and temperature of insulating oil are mainly tested. Through curve fitting of insulating oil characteristics of onboard traction transformers under different maintenance levels, the state characteristics are analyzed and extracted. Based on the dielectric spectrum of onboard traction transformer winding, the time-domain characteristics are extracted to represent the transformer's status. Combined with the dissolved gas in insulating oil, the state feature sets of the onboard traction transformer are established. To solve the problem of signal redundancy and correlation analysis, KPCA is used to process the state feature set. Then the feature vectors of the onboard traction transformer in different states are established.

3) RUNNING STATE DIAGNOSIS

In this paper, the sample set of onboard traction transformers in "Good", "Fair", "Poor", and "Fault" states is established, and the initial fuzzy matrix is established combined with the test set. The fuzzy matrix is normalized, and a dynamic clustering diagram is formed by fuzzy clustering analysis. When the confidence factor λ is given a specific value, the sample set and the test set are matched and classified, to express the state diagnosis results intuitively in the case of small samples.

C. EVALUATION OF DIAGNOSIS RESULTS

To evaluate the effectiveness of the clustering algorithm, the evaluation indexes include accuracy rate (P), recall rate (R), and F-measure (F), and the calculation methods of the three indexes are shown in Equations (18) $\sim$ (20). TP is the actual cluster of the data set, and FP is the cluster obtained after clustering.

$$P = \frac{|TP \cap FP|}{|FP|} \times 100\% \tag{17}$$

$$R = \frac{|TP \cap FP|}{|TP|} \times 100\% \tag{18}$$

$$F = \frac{2 \times P \times R}{P + R} \tag{19}$$

Among the three evaluation indexes, F-measure is a comprehensive index of accuracy rate and recall rate, and the larger its value, the better the clustering effect will be.

IV. VALIDATION AND ANALYSIS

A. STATE FEATURE EXTRACTION AND ANALYSIS

1) AGING CHARACTERISTICS OF INSULATING OIL

To study the relationship between the aging characteristics of insulating oil and onboard traction transformers, this paper carries out oil sample detection on different levels of maintenance trains in Changke Maintenance Center. Taking the onboard traction transformers of new trains and 4.8 million km level-5 maintenance trains as examples, the oil sample detection results of the aging characteristics of insulating oil are shown in Tabs. 2 and 3.

TABLE 2. Aging characteristics of insulating oil of onboard traction transformers in new trains.

Feature code	Sample 1	Sample 2
f_1(Ω.m)	1.295E11	2E11
f_2 (mg/L)	0.001	0.001
f_3 (kV)	76.5	84
f_4 (ppm)	3	4
f_5 (°C)	17	17

TABLE 3. Aging characteristics of insulating oil of onboard traction transformers in level-5 maintenance trains.

Feature code	Sample 1	Sample 2
f_1(Ω.m)	5.577E8	5.672E8
f_2 (mg/L)	0.001	0.001
f_3 (kV)	40.9	41.2
f_4 (ppm)	106.5	114.6
f_5 (°C)	23	22

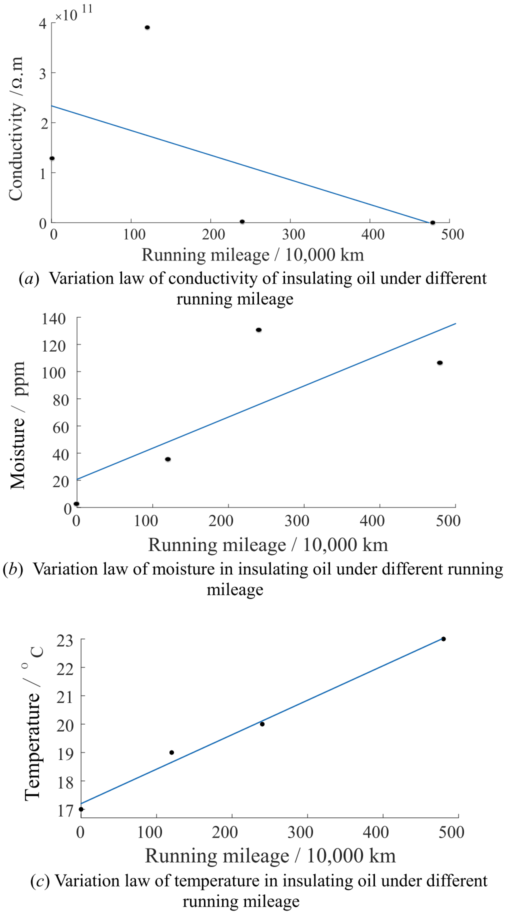
According to the requirements of GB/T 7595-2008 and DL/T 596-1996 on the characteristics of transformer insulating oil in Reference [26], the onboard traction transformer of the new train meets the standard. With the increase of operating mileage, the conductivity, moisture, and temperature of insulating oil of level-5 maintenance trains exceed the standard. According to the above analysis, curve fitting is conducted for conductivity, moisture, and temperature of insulating oil under different running mileage, and the results are shown in Fig. 2.

Fig. 2 shows that with the increase of running mileage, the conductivity, moisture, and temperature of insulating oil show a monotonous trend, which can also further characterize the state change of the onboard traction transformer.

2) AGING CHARACTERISTICS OF THE MAIN INSULATION

The insulation aging characteristics of windings of onboard traction transformers change under different operating conditions. To extract the aging characteristics of the main insulation, the frequency response of the main insulation is tested using the method of dielectric spectrum analysis. The tested dielectric spectrums of onboard traction transformers of new trains and level-4 maintenance trains in Changke Maintenance Center are shown in Fig. 3.

The test results show that in the low-frequency domain, there is no much different between the impedance of the new train and the level-4 maintenance train due to the interference of the environment. However, in the high-frequency stage, the impedance of the level-4 maintenance train increases significantly. To characterize the features of the dielectric spectrums in time-domain, dimensionless indexes such



(a) Variation law of conductivity of insulating oil under different running mileage

(b) Variation law of moisture in insulating oil under different running mileage

(c) Variation law of temperature in insulating oil under different running mileage

FIGURE 2. Changes in the aging characteristics of onboard traction transformers.

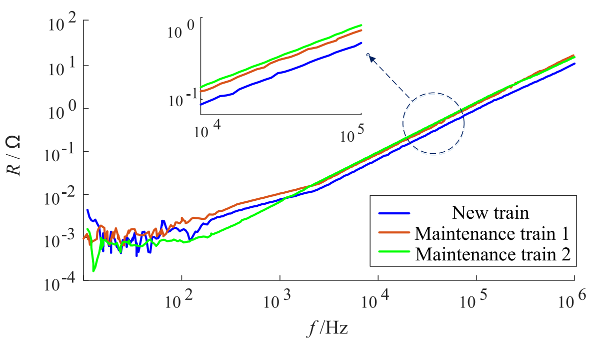


FIGURE 3. Dielectric spectrum curve of the main insulation of onboard traction transformers.

as RMS, peak factor, and kurtosis factor are selected as the feature parameters. They are calculated as follows:

RMS r_1:

$$r_1 = \sqrt{\frac{1}{N} x_i^2} \tag{20}$$

Peak factor r_2:

$$r_2 = \max_{i=1}^{N}(x_i) / \sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2} \tag{21}$$

Kurtosis factor r_3:

$$r_3 = \frac{E[(X-\mu)^4]}{\sigma^4} \tag{22}$$

where x_i is the signal sequence, μ is the mean value, and σ is the root mean square value.

The RMS reflects the energy of the dielectric spectrum, the peak factor measures the impact characteristics of the curve, and the kurtosis factor describes the probability density of the signal sequence and reflects the complexity of the waveform.

3) CHARACTERISTICS OF DISSOLVED GASES IN INSULATING OIL

To fully characterize the status feature of the onboard traction transformer under different running statuses, the DGA of the onboard traction transformer is added to the status feature set. In this paper, a traction transformer under level-4 maintenance in Changke Maintenance Center is taken as an example, the DGA data is shown in Tab. 4.

TABLE 4. DGA characteristics of the onboard traction transformer of the level-4 repair vehicle.

Feature code	State feature	Feature code	State feature
d_1 (μL/L)	7.8	d_5 (μL/L)	0.62
d_2 (μL/L)	214.23	d_6 (μL/L)	2.60
d_3 (μL/L)	1456.23	d_7 (μL/L)	6.06
d_4 (μL/L)	3.46	d_8 (μL/L)	26.19

According to GB/T 7252-2001 [27], the standard value of the onboard traction transformer C_2H_4 in the normal state should be zero ppm. In Table 4, the content of C_2H_4 significantly exceeds the standard, which further reflects the status change of the onboard traction transformer.

B. KPCA PROCESSING OF THE STATE FEATURE SET

Based on the aging characteristics of insulating oil and main insulation, combined with the consistency between DGA characteristics and the running state of the onboard traction transformer, the original feature matrix $\boldsymbol{Q}^{4\times4}$ is established. The aging characteristics of insulating oil are $\{f_1, f_2, f_3, f_4, f_5\}$, the aging characteristics of main insulation are $\{u_1, u_2, u_3\}$, and the DGA characteristics are $\{d_1, d_2 \ldots d_8\}$, there are a total of 16 characteristic indexes, and the original feature set $\boldsymbol{Q}^{4\times4}$ established is shown as follows:

$$\boldsymbol{Q} = \begin{bmatrix} f_1 & f_2 & f_3 & f_4 \\ f_5 & u_1 & u_2 & u_3 \\ d_1 & d_2 & d_3 & d_4 \\ d_5 & d_6 & d_7 & d_8 \end{bmatrix}$$

Considering the cooling system of the onboard traction transformer, the CO and CO_2 dissolved in insulating oil cannot be regarded as status characteristics. To eliminate signal redundancy and ensure the correlation between status features and the running status, further analysis of the feature set is needed. Through the above analysis, the state of an onboard traction transformer of a new train is taken as an example, the original feature set $\boldsymbol{Q}^{4\times4}$ established is:

$$\boldsymbol{Q} = \begin{bmatrix} 1.29e11 & 0.001 & 76.5 & 3 \\ 17 & 1.2 & 5 & 101 \\ 1 & 0 & 0.1 & 0 \\ 1 & 2.045 & 2.124 & 218.272 \end{bmatrix}$$

Some status features have poor relevance to changes in running statuses and cannot be used as inputs for fuzzy clustering analysis. To eliminate signal redundancy, feature sets need to be further analyzed. Based on the KPCA theory and the linear mapping relationship, the feature set $\boldsymbol{Q}$ is analyzed using KPCA, and the parameters with the cumulative contribution rate of more than 95% of the feature values are selected as the new feature set $\bar{\boldsymbol{Q}}$. After KPCA processing, the state feature vector of the onboard traction transformer of the new train is $\bar{\boldsymbol{Q}}$:

$$\bar{\boldsymbol{Q}} = [1.190e11 \;\; 16.58314.111 \;\; 3.694]$$

The feature vector $\bar{\boldsymbol{Q}}$ can satisfy the requirement of no loss of signal features and eliminate information redundancy, which greatly reduces the calculation of state diagnosis.

C. RUNNING STATE DIAGNOSIS OF ONBOARD TRACTION TRANSFORMERS

1) ESTABLISH A DIAGNOSTIC MODEL FOR THE RUNNING STATUS OF THE ONBOARD TRACTION TRANSFORMER

Based on the analysis of the status features of the onboard traction transformer, a set of characteristic vectors of different running statuses are taken as the sample object, and the characteristic vectors of "Good", "Fair", "Poor", and "Fault" are defined by $\boldsymbol{w}_1 \sim \boldsymbol{w}_4$. The fuzzy clustering algorithm is used to perform clustering analysis to establish a diagnosis model of the onboard traction transformer's running state. The specific implementation steps are as follows:

Step 1: Build the initial fuzzy matrix $\boldsymbol{X}$.

$$\boldsymbol{X} = [\boldsymbol{w}_1; \boldsymbol{w}_2; \boldsymbol{w}_3; \boldsymbol{w}_4] \tag{23}$$

Step 2: Use Equations (11) and (12) to obtain the standard fuzzy matrix $\boldsymbol{X}''$.

$$\boldsymbol{X}'' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.131 & 1 & 1 & 1 \\ 0.026 & 0.297 & 0.557 & 0.759 \\ 0 & 0.565 & 0.771 & 0.946 \end{bmatrix}$$

Step 3: The fuzzy similarity matrix $\boldsymbol{R}$ is established by using Equation (13).

$$\boldsymbol{R} = \begin{bmatrix} 1 & 0.340 & 0.527 & 0.429 \\ 0.340 & 1 & 0.703 & 0.826 \\ 0.527 & 0.703 & 10.866 & \\ 0.429 & 0.826 & 0.866 & 1 \end{bmatrix}$$

Step 4: The fuzzy equivalent matrix $\boldsymbol{R}^*$ is constructed using Equation (14).

$$\boldsymbol{R}^* = \begin{bmatrix} 1 & 0.527 & 0.527 & 0.527 \\ 0.527 & 1 & 0.826 & 0.826 \\ 0.527 & 0.826 & 1 & 0.866 \\ 0.527 & 0.826 & 0.866 & 1 \end{bmatrix}$$

Step 5: Establish a diagnostic model of running status

Equation (15) is used to establish the corresponding equivalent Boolean matrix $\boldsymbol{R}_\lambda$. When the confidence factor λ changes from large to small, the same columns of the equivalent Boolean matrix $\boldsymbol{R}_\lambda$ are grouped into one class, and a dynamic clustering diagram is formed, to complete the formation of the onboard traction transformer's running state diagnosis model.

2) EXAMPLE DEMONSTRATION AND RESULT ANALYSIS

To verify the effectiveness of the diagnosis model of the onboard traction transformer's running status, two groups of onboard traction transformers in the "Good" and "Fault" state are randomly selected as test sets in the Changke Maintenance Center, and the labels are defined as $\boldsymbol{c}_1 \sim \boldsymbol{c}_2$, and $\boldsymbol{u}_1 \sim \boldsymbol{u}_2$ respectively. Combined with the diagnosis model of the onboard traction transformer in the previous section, the status feature vector to be tested is subjected to fuzzy clustering analysis to form a dynamic clustering diagram, as shown in Fig. 4.

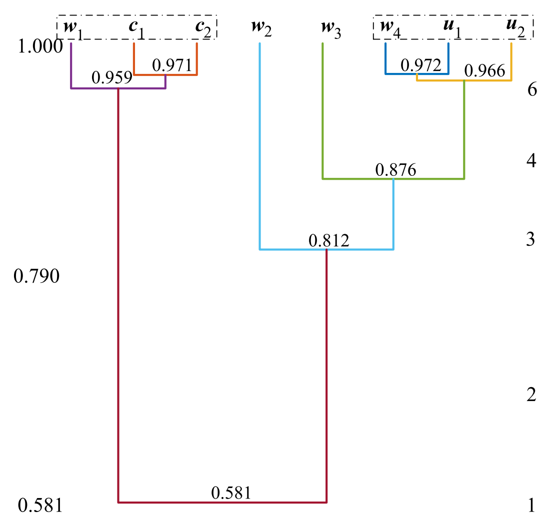


FIGURE 4. Dynamic clustering diagram of running status to be tested.

As can be seen from the dynamic clustering diagram of the running status, when the confidence factor λ on the left side declines from 1 to 0, the number of categories on the right side decreases and eventually falls into one category. When λ changes to 0.959 and 0.971, $\boldsymbol{w}_1$ and $\boldsymbol{c}_1 \sim \boldsymbol{c}_2$ are divided into one category; then the status to be tested $\boldsymbol{c}_1 \sim \boldsymbol{c}_2$ and sample $\boldsymbol{w}_1$ belong to the same running state, so $\boldsymbol{c}_1 \sim \boldsymbol{c}_2$ are diagnosed as the "Good" state. Similarly, when λ CHANGES to 0.966 and 0.972, $\boldsymbol{w}_4$ and $\boldsymbol{u}_1 \sim \boldsymbol{u}_2$ are divided into one category, and the status to be tested $u_1 \sim u_2$ are diagnosed as "Faulty", which is consistent with the on-site detection result.

D. COMPARISON AND ANALYSIS

1) STATE FEATURE EXTRACTION ALGORITHM

Taking the fuzzy clustering state diagnosis model under the traditional DGA feature as an example, $c_1 \sim c_2$ and $u_1 \sim u_2$ of the onboard traction transformer in the "Good" and "Fault" states in the previous section are input into the diagnosis model to form a dynamic clustering diagram, as shown in Fig. 5

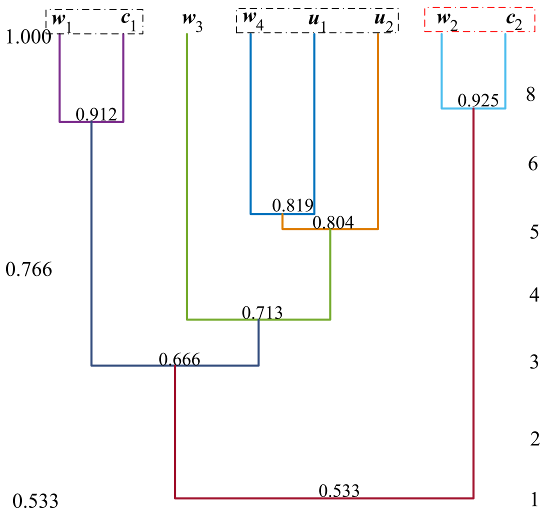


FIGURE 5. Dynamic clustering diagram based on traditional DGA features.

It can be seen from Fig. 5, when λ changes to 0.952, c_2 is misdiagnosed as a "Fair" state, which is inconsistent with the field test results. Only the DGA feature is used as state quantity, which makes it not sensitive to the weak change of state.

To further verify the effectiveness of the state feature extraction algorithm proposed in this paper, the F-measure of fuzzy clustering analysis is compared under the feature set without KPCA, the traditional DGA feature, and only with insulation aging feature. Taking the Changke Maintenance Center about the running status of the onboard traction transformer of different levels as the test set, a total of 40 groups are input into different diagnosis models, and the diagnosis results are shown in Tab. 5.

The field experimental results show that compared with other state feature extraction algorithms; the F-measure of the feature set processed by KPCA is 97.5%, which further verifies the effectiveness of the KPCA algorithm in state feature extraction of onboard traction transformer.

2) STATUS DIAGNOSIS ALGORITHM

Under the four different levels of running state, a total of 100 groups of datasets of onboard traction transformers are used as experimental data. Among the 100 groups of sets, the first 60 groups are used as training samples, while the remaining 40 groups are used as test sets. In this paper, the algorithms including support vector machine (SVM), back-propagation neural network (BPNN), and extreme learning machine (ELM) are added for comparison, with different sample sizes. Fig. 6 only presents the diagnosis result using SVM. As is shown in the figure, the state diagnosis accuracy of the test set of SVM is 82.5%, which is the second highest among different algorithms, but still inferior to the proposed method. For comparison, the diagnosis results of the different algorithms with different sample sizes are provided in Tab. 6.

TABLE 5. Diagnostic results under different state features.

Diagnostic model	F-measure (%)
Traditional DGA feature	85
Insulation aging feature	87.5
Feature set without KPCA	90
Feature set with KPCA	97.5

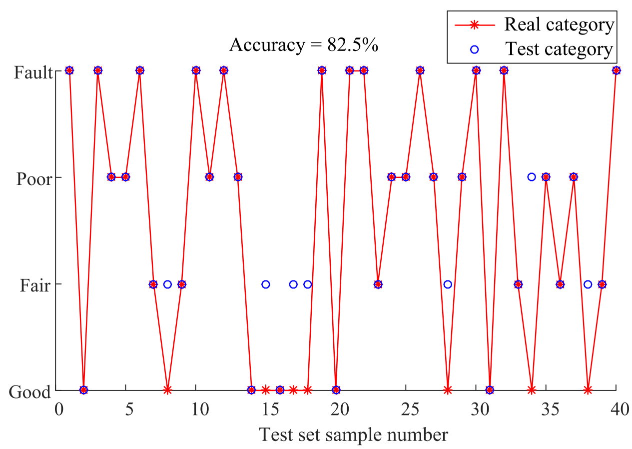


FIGURE 6. State diagnosis result with SVM.

TABLE 6. Diagnostic results with different algorithms and different sample size.

Algorithm	Number of samples		Diagnosis accuracy (%)
	Training sample	Test sample	
SVM	60/50	40/50	82.5/80.0
BPNN	60/50	40/50	77.5/77.5
ELM	60/50	40/50	80.0/77.5
Fuzzy cluster	/	40/50	95.0/95.0

The results show that the diagnostic accuracy of either the proposed method or the SVM is indeed high. In comparison, the proposed method is more dominant. This is mainly because the algorithm of SVM is based on structural risk and VC (Vapnik chervonenkis) theory. The size of the VC dimension directly affects the fitting degree, but the fuzzy clustering algorithm does not need to be trained in advance to achieve parameter optimization.

3) DIAGNOSTIC RESULTS WITH PUBLIC DATASET

According to the public dataset in IEC TC-10 (also see in IEC 60599) of power transformers, the running status of power transformer is summarized into 8 categories (Normal, Partial discharge, Discharges of low energy, Discharges of high energy, Thermal faults of temperature < 300 °C, Thermal faults of temperature 300 °C < T < 700 °C, Thermal faults of temperature T> 700 °C, and mixtures of electrical and thermal faults), which is different from the status categories for onboard traction transformer. For verification of the proposed method, the number of fault categories does not affect much. To verify the repeatability of the fuzzy clustering algorithm under the public dataset, a total of 48 groups of experimental data are taken from each category, in which 30 groups under different fault categories are used to establish the fault diagnosis model, and the last 18 groups are used as the test set. The experimental results are given in Tab. 7. It is shown that the K-measure of power transformer based on fuzzy clustering analysis is 95.8%, which verifies the applicability of the diagnosis algorithm under the public dataset.

TABLE 7. Diagnostic results with public datasets.

Algorithm	Number of samples		Diagnosis accuracy (%)
	Training sample	Test sample	
SVM	30	18	88.9 (16/18)
BPNN	30	18	83.3 (15/18)
ELM	30	18	72.2. (13/18)
Fuzzy cluster	/	48	95.8 (46/48)

V. CONCLUSION

Given the consistency between the running status of the onboard traction transformer and the aging characteristics of insulating oil and the main insulation, this paper proposes an algorithm to diagnose the running status of the onboard traction transformer.

a) Onboard traction transformer is different from the power transformer. The aging characteristics of insulating oil under different maintenance levels are analyzed, such as conductivity, furfural, conductivity, moisture, temperature, and so on. The status features of an onboard traction transformer are established by analyzing the dielectric spectrum of main insulation and combining it with the characteristics of dissolved gas in oil.

b) The fuzzy clustering is carried out to analyze the running status of the onboard traction transformer under with DGA features, without KPCA processing, and only with aging characteristics, respectively. Results show that the K-measure can be as high as 80% with traditional DGA features, and 90% for that without KPCA features, and 87.5% for that only with insulation aging features, while the K-measure of the algorithm proposed in this paper is high than 95%, which not only improves the diagnosis accuracy, but also provides a new guarantee for the equipment maintenance.

At present, the status features can only be extracted when the train is in maintenance. In the future, how to realize the full-cycle operation status monitoring of the onboard traction transformer becomes the future research direction.

REFERENCES

[1] S. Lin, X. Sun, and D. Feng, "A failure rate model for traction transformer based on PHM considering multiple factors," in *Proc. Prognostics Syst. Health Manage. Conf.*, Oct. 2016, pp. 1–6.

[2] S. Gao, Z. G. Liu, J. F. Xu, S.-Y. He, and S.-B. Gao, "Research on fault diagnosis for traction transformer on the basis of model-based diagnosis and expert system," *J. China Railway Soc.*, vol. 35, no. 7, pp. 42–49, 2013.

[3] W. Jing and Y. Z. Feng, "High-speed railway traction power supply system health management and fault early warning system," *China High-Tech*, vol. 50, no. 14, pp. 81–83, 2019.

[4] Y. Xiao, W. Pan, X. Guo, S. Bi, D. Feng, and S. Lin, "Fault diagnosis of traction transformer based on Bayesian network," *Energies*, vol. 13, no. 18, p. 4966, Sep. 2020.

[5] N. A. Fauzi, N. H. N. Ali, P. J. Ker, V. A. Thiviyanathan, Y. S. Leong, A. H. Sabry, M. Z. B. Jamaludin, C. K. Lo, and L. H. Mun, "Fault prediction for power transformer using optical spectrum of transformer oil and data mining analysis," *IEEE Access*, vol. 8, pp. 136374–136381, 2020.

[6] X. Huang, X. Huang, B. Wang, and Z. Xie, "Fault diagnosis of transformer based on modified grey wolf optimization algorithm and support vector machine," *IEEJ Trans. Electr. Electron. Eng.*, vol. 15, no. 3, pp. 409–417, Mar. 2020.

[7] L. Zhou, D. Wang, Y. Cui, L. Zhang, L. Wang, and L. Guo, "A method for diagnosing the state of insulation paper in traction transformer based on FDS test and CS-DQ algorithm," *IEEE Trans. Transport. Electrific.*, vol. 7, no. 1, pp. 91–103, Mar. 2021.

[8] Z. Jiao-jiao, C. Te-fang, and F. Qiang, "The research and application of WNN in the fault diagnosis technology of electric locomotive main transformer," in *Proc. IET Int. Conf. Power Electron.*, 2014, pp. 1–6.

[9] Q. Fu, T.-F. Chen, and J.-J. Zhu, "Research on traction transformer faults diagnosis algorithm based on HPSO-WNN," *J. China Railway Soc.*, vol. 34, no. 9, pp. 26–32, 2012.

[10] C. Dai, Z. Liu, K. Hu, and K. Huang, "Fault diagnosis approach of traction transformers in high-speed railway combining kernel principal component analysis with random forest," *IET Electr. Syst. Transp.*, vol. 6, no. 3, pp. 202–206, Sep. 2016.

[11] J. Jiang, J. Chen, J. Li, X. Yang, Y. Bie, P. Ranjan, C. Zhang, and H. Schwarz, "Partial discharge detection and diagnosis of transformer bushing based on UHF method," *IEEE Sensors J.*, vol. 21, no. 15, pp. 16798–16806, Aug. 2021.

[12] X. M. Tao, L. Zhen, and Y. Zhang, "Acquisition system of wavelet energy spectrum entropy of transformer vibration signal based on DSP," *Electr. Power*, vol. 48, no. 8, pp. 37–41, 2015.

[13] J. Liu, Z. Wang, X. Fan, Y. Zhang, and J. Wang, "A novel curve database for moisture evaluation of transformer oil-immersed cellulose insulation using FDS and exponential decay model," *IEEE Access*, vol. 8, pp. 180728–180737, 2020.

[14] Z. Li, Y. Zhang, A. Abu-Siada, X. Chen, Z. Li, Y. Xu, L. Zhang, and Y. Tong, "Fault diagnosis of transformer windings based on decision tree and fully connected neural network," *Energies*, vol. 14, no. 6, p. 1531, Mar. 2021.

[15] H.-J. Lee, K.-H. Yoon, D.-J. Kweon, and J.-C. Kim, "A failure diagnosis method of oil-filled transformer using expert system based on Duval triangle method and IEC 60599," *Trans. Korean Inst. Electr. Eng.*, vol. 68, no. 10, pp. 1189–1194, Oct. 2019.

[16] Y. Zhang, X. Li, H. Zheng, H. Yao, J. Liu, C. Zhang, H. Peng, and J. Jiao, "A fault diagnosis model of power transformers based on dissolved gas analysis features selection and improved krill herd algorithm optimized support vector machine," *IEEE Access*, vol. 7, pp. 102803–102811, 2019.

[17] S. S. M. Ghoneim, K. Mahmoud, M. Lehtonen, and M. M. F. Darwish, "Enhancing diagnostic accuracy of transformer faults using teaching-learning-based optimization," *IEEE Access*, vol. 9, pp. 30817–30832, 2021.

[18] S. A. Ward, A. El-Faraskoury, M. Badawi, S. A. Ibrahim, K. Mahmoud, M. Lehtonen, and M. M. F. Darwish, "Towards precise interpretation of oil transformers via novel combined techniques based on DGA and partial discharge sensors," *Sensors*, vol. 21, no. 6, p. 2223, 2021.

[19] A. R. Aarathy, S. S. Pillai, D. Gopinath, S. S. Sajan, A. R. Nair, and B. Aryanandiny, "Spectral analysis for condition assessment of transformer oil under different aging conditions," in *Proc. IEEE 1st Int. Conf. Smart Technol. Power, Energy Control (STPEC)*, Sep. 2020, pp. 1–5.

[20] C. Li, M. Cerrada, D. Cabrera, R. V. Sanchez, F. Pacheco, G. Ulutagay, and J. V. de Oliveira, "A comparison of fuzzy clustering algorithms for bearing fault diagnosis," *J. Intell. Fuzzy Syst.*, vol. 34, no. 6, pp. 3565–3580, Jun. 2018.

[21] J. Rabcan, V. Levashenko, E. Zaitseva, M. Kvassay, and S. Subbotin, "Non-destructive diagnostic of aircraft engine blades by fuzzy decision tree," *Eng. Struct.*, vol. 197, Oct. 2019, Art. no. 109396.

[22] C. Liu, Y. J. Wu, and C. G. Zhen, "Rolling bearing fault diagnosis based on variational mode decomposition and fuzzy C means clustering," *Proc. Chin. Soc. Electr. Eng.*, vol. 35, no. 13, pp. 3358–3365, 2015.

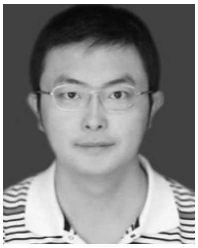
[23] H. Sun, Y. Guo, and W. Zhao, "Fault detection for aircraft turbofan engine using a modified moving window KPCA," *IEEE Access*, vol. 8, pp. 166541–166552, 2020.

[24] L. Qingsheng and S. Haifu, "China's heavy-haul railway development and technical standards," *Eng. Construct. Standardization*, vol. 4, no. 4, pp. 53–56, 2015.

[25] S. Shi, K. Wang, L. Chen, F. Yang, P. Liu, D. Tao, and G. F. Lu, "Power transformer status evaluation and warning based on fuzzy comprehensive evaluation and Bayes discrimination," *Electr. Power Automat. Equip.*, vol. 36, no. 9, pp. 60–66, 2016.

[26] F. Zhao and S. Li, "Fault diagnosis for traction transformer based on DGA and improved grey correlation analysis model," *High Voltage App.*, vol. 51, no. 1, pp. 41–45, 2015.

[27] L. Cheng and T. Yu, "Dissolved gas analysis principle-based intelligent approaches to fault diagnosis and decision making for large oil-immersed power transformers: A survey," *Energies*, vol. 11, no. 4, p. 913, 2018.



JUNMIN ZHU received the M.S. degree from the Department of Automation, Lanzhou Jiaotong University, in 2021, where he is currently pursuing the Ph.D. degree with the School of Control Engineering. His research interests include intelligent fault diagnosis, power system optimization control, and intelligent control research.

SHUAIBING LI (Member, IEEE) was born in 1989. He received the Ph.D. degree from Southwest Jiaotong University, in 2018. He is currently an Associate Professor with Lanzhou Jiaotong University. He is mainly engaged in the research work of condition assessment and nondestructive testing of high-voltage electrical equipment and the detection and diagnosis of new energy power systems.

HAIYING DONG (Member, IEEE) was born in 1966. He received the Ph.D. degree from Xi'an Jiaotong University, in 2003. He is currently a second-class Professor and a Ph.D. Supervisor with Lanzhou Jiaotong University. He is mainly engaged in power system optimization and intelligent control and new energy power generation optimization control.