Check for updates

Rapid assessment of distribution network equipment status based on fuzzy decision making

Wang Qian[1,2]*, Li Yuquan[2], Wang Xiaohui[3], Yang Kun[1], Liu Yang[1], Xia Zhongyuan[3], Lan Guangyu[3], Liu Yunlong[3] and Tang Jiyuan[1]

[1]School of Electrical Engineering, Chongqing University, Chongqing, China, [2]State Grid Henan Electric Power Research Institute, Zhengzhou, China, [3]State Grid Henan Electric Power Company, Zhengzhou, China

Voltage instability, power imbalance, and unreliability are caused mainly by equipment failure in the distribution system, so it is important to accurately and quickly assess the status of distribution network equipment. However, it is challenging to detect equipment failures, the traditional XGBoost algorithm is unsuitable because some evaluation indices are incompetent to quantify. To address these issues, we propose a fast evaluation method for the state of electrical distribution equipment based on fuzzy decision-making. Firstly, key indices are selected from the multi-source equipment information. Secondly, this paper constructs the mapping between key indices and equipment status scores by combining the fuzzy iterative method and the XGBoost algorithm. Finally, the proposed assessment model is confirmed by using the distribution transformers as an example. The result shows that the proposed multi-source information assessment method can quickly and accurately determine the operation status of electrical distribution equipment, and the proposed method has better accuracy than the traditional method.

1 Introduction

With social and economic development, the scale of the power grid is gradually expanding, prompting higher reliability requirements. The distribution network is an essential part of the power system. It is responsible for the critical task of supplying power directly to customers (Liang et al., 2009). At the same time, the clear aims of operational status analysis improving the reliability of the power supply. It is an integral part of operations and maintenance in the power system (Guan, 2022). Through testing and evaluating equipment such as distribution transformers (DTs) and circuit breakers, system operators can ensure the safe and stable operation of distribution equipment and improve the economy of power supply companies.

In the research on equipment evaluation, there is more research on primary equipment such as generators, transformers, circuit breakers, etc., and less on medium- and low-voltage equipment. Most of them are limited to studying the remaining life of power transformers, and the distribution equipment still needs to form a set of quantitative evaluation methods (Fang et al., 2023). However, the location dispersion, the large amount of monitoring data, and the lack of uniform evaluation

standards have brought significant challenges in assessing equipment status (Yuan et al., 2019; Tamma et al., 2021).

The transformer is the main equipment for distributing power and transforming voltage to a wide range of customers, and people have always valued it (Guo and Liu, 2005). Methods such as fuzzy evaluation and artificial intelligence are widely used in transformer status assessment (Zhu et al., 2008; Xie et al., 2012; Chen, 2017; Zhou et al., 2020; Lv, 2022). Zhou et al. proposed a transformer condition assessment method based on an interval grey number dynamic grey target (Lv, 2022). The accuracy of the proposed transformer condition assessment method is verified by integrating the dynamic changes of transformer operation data and index information in multi-dimensional time phases. Zhu et al. used the transformer oil chromatography data as the operational condition assessment index, proposed a new method to transform the qualitative indexes into quantitative indexes, and finally used the obtained assessment results as the training set of the SVM and obtained the transformer's status level (Xie et al., 2012). However, the model consideration is relatively single, and the assessment results are not accurate. Reference (Zhou et al., 2020) studied the power transformer condition evaluation index system in depth, as well as the calculation method and model of the transformer health index based on fuzzy logic. Reference (Chen, 2017) established an insulation state evaluation system and proposed a transformer insulation state evaluation method based on fuzzy cloud theory. The affiliation degree cloud model was utilized to describe its fuzziness and randomness. Zhe et al. applied a conventional approach to evaluate transformers' condition and introduced a condition assessment model using support vector regression. However, this method heavily relies on the size of the transformer's sample capacity (Zhu et al., 2008). References (Ahmad and Senroy, 2020; Zhang et al., 2020) proposed a cloud model for transformer condition assessment considering the randomness of the data and the ambiguity of the evaluation level, which successfully realized the transformation between qualitative and quantitative indicators. Zhang et al. proposed a condition assessment index system based on the transformer test category, and the evaluation level was divided by solving the relative deterioration of the index. Finally, the confidence criterion was introduced for comprehensive judgment (Zhang et al., 2010). The model is more subjective, and new solutions need to be proposed to reduce the interference of human factors in the model. Zheng et al. introduced the grey assessment decision theory (Wang et al., 2012), but the index system was not comprehensive enough to consider the data of all the transformer components, and the results were more one-sided and lacked persuasive power.

The current evaluation method for distribution network equipment is not comprehensive. It lacks a quantitative assessment method and is greatly influenced by subjective factors, such as the experts' experience. Therefore, it is important to establish an assessment method that better aligns with the actual operating conditions of the electrical distribution equipment. The proposed method allows for the assessment of distribution transformers, providing a valuable reference for the economic and operational reliability of power system operators.

In this paper, we propose a state assessment model for distribution network equipment. This model integrates multi-source information derived from the operational data of the equipment, taking into account critical state variables. To establish the relationship between the multi-sources information and the equipment's state, we utilize a data-driven fuzzy iterative method and the XGBoost algorithm. This enables a more accurate evaluation of the equipment's condition. Additionally, the paper introduces a method for multi-source information to assess the operation status of distribution equipment, using DTs as an example. This method effectively determines the equipment's operation status by leveraging various types of information, offering a more comprehensive evaluation compared to other approaches.

2 Characteristic extraction and evaluation of key condition indicators of power distribution equipment

There are five categories of key equipment for the distribution network, which are DTs, switchgear, cables, overhead lines, and pole-mounted switches. During operation, a large amount of data is generated, including real-time and historical data, hardware information, and environ-mental conditions. To accurately determine the operating status of equipment, it is necessary to process data and extract key condition indicators that characterize the equipment's operation. Establishing a scientific and comprehensive evaluation system for condition indicators is significant for the status evaluation of distribution network equipment.

DT is an important equipment in the distribution network, and its operation status is closely related to the reliability of the power supply. The Table A1 in the appendix displays various types of faults in DT, including insulation, short-circuit, discharge, and mechanical drive operating mechanism faults. Since the DTs used in industry and large users are mainly step-down transformers and mostly oil-immersed transformers, the condition indicators in Table A1 are selected and classified according to the principle of selecting key condition indicators by referring to standards such as "Guidelines for Condition Evaluation of Equipment in Distribution Networks" (State Grid Corporation, 2011), and the results are shown in Table 1.

Once the condition indicators are selected, it is necessary to score them to further evaluate the state of the DT. According to the uniform regulations, the evaluation principles for each condition indicators are shown in the third column of Table 1. Before evaluation, the condition indicators of the transformer shown in Table 1 need to be normalized due to their qualitative and quantitative indicators varying in orders of magnitude and dimensions. The condition indicators that makes the status of the equipment better when the value gets smaller or lower, such as winding DC resistance and oil temperature, are processed by Eq. 1; The state quantities (withstand voltage, insulation resistance, etc.) that make the equipment state better when the value becomes larger or higher are handled

TABLE 1 Classification of key state variables of DT.

Classification	Specific parts	Condition indicators
Hardware situation	Sealing method	Sealing ability μ_1
	Degree of insulation	Withstand voltage test μ_2
	System contamination	Contamination μ_3
	Non-electricity protection device	Insulation resistance μ_{10}
	Winding and bushing	DC resistance μ_{11}
Operational situation	Oil level	Oil level μ_4
	Winding and bushing outer temperature	Temperature μ_5
	Grounding condition	Grounding down conductor appearance μ_6
	Respirator	Respirator status μ_7
	Load situation	Load rate μ_{12}
	Three-phase load balancing	Three-phase unbalance rate μ_{13}
Human factors	Equipment identity	Completeness of identification μ_8
	Tap changer	Tap changer performance μ_9

by Eq. 2. (Wang and Zhao, 2020). Empirical data gives the degree of deterioration in the qualitatively measured condition indicators such as running time and sealing performance.

$$\mu_{i,j} = \begin{cases} 0 & \mu_{i,j} \leq \mu_{i,j,0} \\ \dfrac{\mu_{i,j} - \mu_{i,j,0}}{\mu_{i,j,1} - \mu_{i,j,0}} & \mu_{i,j,0} < \mu_{i,j} \leq \mu_{i,j,1} \\ 1 & \mu_{i,j} > \mu_{i,j,1} \end{cases} \quad (1)$$

$$\mu_{i,j} = \begin{cases} 1 & \mu_{i,j} < \mu_{i,j,1} \\ \dfrac{\mu_{i,j,0} - \mu_{i,j}}{\mu_{i,j,0} - \mu_{i,j,1}} & \mu_{i,j,1} \leq \mu_{i,j} < \mu_{i,j,0} \\ 0 & \mu_{i,j} \geq \mu_{i,j,0} \end{cases} \quad (2)$$

Where the value of the subscript i of $\mu_{i,j} (i = 1, 2, ..., 9)$ is determined by the condition indicators; j indicates the relative deterioration degree of the condition indicators, $\mu_{i,j}$ is the observed value divided by the ideal value, and the range of values is $[0,1]$; $\mu_{i,j,0}$ is the baseline value, and $\mu_{i,j,1}$ denotes the attention value or the warning value, the values of $\mu_{i,j,0}$ and $\mu_{i,j,1}$ are obtained by (Wang and Zhao, 2020).

According to the evaluation criteria of condition indicators given in Table 1, combined with a large number of experts and long-term experience in the field (China Electric Power, 2008), the assessment set of DT key condition indicators in Table 1 is obtained, as shown in Table 2.

2.1 Weight determination based on fuzzy iteration and XGBoost

After selecting the key condition indicators for distribution network equipment, reasonable weights must be assigned to each status variable before conducting a comprehensive state assessment. In this paper, we use the eclectic fuzzy decision-making and multilevel fuzzy integrated evaluation model to analyze the pre-data of DT's. Then, the weight ratios of the assessment set are constantly updated by the XGBoost algorithm, which reduces the influence of subjective factors brought by experts and improves the reliability of data analysis. Finally, an expert database was established.

2.2 Solution process for eclectic fuzzy decision-making weights

The flow chart for eclectic fuzzy decision-making is illustrated in Figure 1. Beginning with the original sample data, First, virtualizing the fuzzy positive ideal and fuzzy negative ideal. The fuzzy positive ideal is composed of the maximum value of the fuzzy indicator in each indicator, while the fuzzy negative ideal is composed of the minimum value of the fuzzy indicator in each indicator (Zadeh, 1965). Next, the weighted Euclidean distance is used to calculate the distance between each alternative object and the fuzzy positive ideal and fuzzy negative ideal. Based on this, the degree of affiliation of each alternative object belonging to the fuzzy positive ideal is calculated. The greater the degree of affiliation, the more desirable the scheme is.

The basic solution steps for eclectic fuzzy decision-making are as follows.

Step 1: Transform the indicator data into triangular fuzzy numbers;

Suppose that $\mathbf{F}(R)$ is an overall fuzzy set on R and the set $\mathbf{M} \in \mathbf{F}(R)$. The affiliation function $\mu_{\mathbf{M}}$ of $\mathbf{M}$ is denoted as follows:

$$\mu_{\mathbf{M}}(x) = \begin{cases} \dfrac{x - l}{m - l}, x \in [l, m] \\ \dfrac{x - u}{m - u}, x \in [m, u] \\ 0, x < l \text{ or } x > u \end{cases} \quad (3)$$

TABLE 2 Key condition indicators assessment set of DT.

State variables	Description	Assessment set				
		Good	Normal	Attention	Abnormal	Serious
Sealing ability μ_1	Oil leakage situation $\mu_{1,1}$	0.2	0.2	0.3	0.2	0.1
	Oil dripping situation $\mu_{1,2}$	0	0	0.1	0.1	0.8
	Oil spilling situation $\mu_{1,3}$	0	0	0	0	1
Withstand voltage test μ_2	Pressure resistance $\mu_{2,1}$	0	0	0	0.1	0.9
Contamination μ_3	A small amount of Contamination $\mu_{3,1}$	0.9	0.1	0	0	0
	More pollution $\mu_{3,2}$	0.8	0.1	0.1	0	0
	Obviously damaged rust $\mu_{3,3}$	0.1	0.2	0.3	0.3	0.1
	Severely contaminated and blocked $\mu_{3,4}$	0	0	0.2	0.5	0.3
Oil level μ_4	Oil level gauge indicates Abnormality $\mu_{4,1}$	0.1	0.2	0.3	0.2	0.2
	Oil level gauge no Indication $\mu_{4,2}$	0.1	0.1	0.3	0.3	0.2
Temperature μ_5	Temperature of connector is too high $\mu_{5,1}$	0.1	0.3	0.4	0.2	0
	Rise of temperature is not normal $\mu_{5,2}$	0.1	0.2	0.3	0.3	0.1
Grounding down conductor appearance μ_6	Lack of connection $\mu_{6,1}$	0.1	0.2	0.3	0.3	0.1
	Insufficient depth $\mu_{6,2}$	0.2	0.3	0.4	0.1	0
Respirator Condition μ_7	The respirator is completely discolored by moisture $\mu_{7,1}$	0.3	0.3	0.3	0.1	0
	The respirator is completely breathless $\mu_{7,2}$	0.3	0.3	0.3	0.1	0
Identification Integrity μ_8	Lack of identification $\mu_{8,1}$	0	0.1	0.2	0.5	0.2
	Wrong identifies or no identified $\mu_{8,2}$	0	0	0.1	0.4	0.5
Tap changer Performance μ_9	Tap position power indicates abnormal $\mu_{9,1}$	0	0.5	0.5	0	0

Where $l \leq m \leq u$, $\mathbf{M}$ is triangular fuzzy number, denoted as $\mathbf{M} = (l, m, u)$. According to Eq. 3, the qualitative indicators, quantitative indicators, and weight data in the condition indicators are unified into a triangular fuzzy number.

The qualitative indicators $\mu_i (i = 1, 2, ..., 9)$ in DT are converted to quantitative indicators according to Table 3

The quantitative indicator values $\mu_i (i = 10, 11, ..., 13)$ of DT critical state quantities are expressed in the form of a triangular fuzzy number as shown in Eq. 4.

$$\mu_i = (\mu_i, \mu_i, \mu_i) \tag{4}$$

After transforming all indicators into triangular fuzzy numbers, the matrix of fuzzy indicators is obtained and denoted as $\mathbf{F} = (f_{ij})_{m \times n}$

The weighted triangular fuzzy number of quantitative indicators is obtained according to Eq. 4 and is expressed as Eq. 5:

$$\mathbf{w} = [(w_1, w_1, w_1), (w_2, w_2, w_2), ... (w_i, w_i, w_i)] \tag{5}$$

The weighted triangular fuzzy numbers of qualitative indicators were obtained according to the transformation method in Table 3.

Step 2: Fuzzy indicator matrix normalization process;

Assuming that there are N evaluation objects and the evaluation indicator $j (j \in n)$ corresponds to N fuzzy indicator values in $\mathbf{F}$, which are denoted as $x_i = (a_i, b_i, c_i)$, $(i = 1, 2, ..., N)$, the formula for the normalization of x_i is as follows:

①. If x_i is the value of the fuzzy indicator corresponding to the cost-based indicator, the normalization formula is Eq. 6:

$$y_i = \left(\frac{\min(a_i)}{c_i}, \frac{\min(b_i)}{b_i}, \frac{\min(c_i)}{a_i} \wedge 1 \right) \tag{6}$$

②. If x_i is the fuzzy indicator value corresponding to the income-based indicators, the normalization formula is

$$y_i = \left(\frac{a_i}{\max(c_i)}, \frac{b_i}{\max(b_i)}, \frac{c_i}{\max(a_i)} \wedge 1 \right) \tag{7}$$

The normalized fuzzy indicator matrix is denoted as $\mathbf{R} = (y_{ij})_{m \times n}$

Step 3: Constructing the fuzzy decision-making matrix $\mathbf{D} = (r_{ij})_{m \times n}$

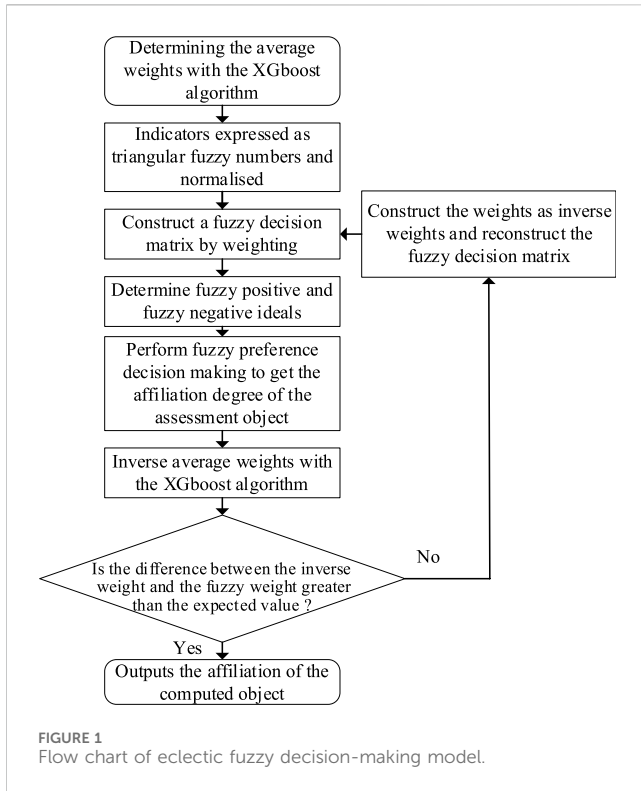The fuzzy decision-making matrix can be obtained by weighting $\mathbf{R}$, as shown in Eq. 8:

FIGURE 1
Flow chart of eclectic fuzzy decision-making model.

$$r_{ij} = \mathbf{w}\Theta y_{ij} \ (i = 1, 2, ..., N, j = 1, 2, ..., N) \tag{8}$$

Step 4: Determine the fuzzy positive ideal M^+ and the fuzzy negative ideal M^-, as shown in Eqs 9, 10;

$$M^+ = \left(M_1^+, M_2^+, ..., M_N^+\right) \tag{9}$$

$$M^- = \left(M_1^-, M_2^-, ..., M_N^-\right) \tag{10}$$

where component $M^+ = \max\left\{r_{1j}, r_{2j}, ..., r_{nj}\right\}$, $(j = 1, 2, ..., 15)$ is the fuzzy maximum value corresponding to the fuzzy indicator value of the j column in the fuzzy decision-making matrix $\mathbf{D}$; where component $M^- = \min\left\{r_{1j}, r_{2j}, ..., r_{nj}\right\}$, $(j = 1, 2, ..., 15)$ is the fuzzy minimum value corresponding to the fuzzy indicator value of the j column in the fuzzy decision-making matrix $\mathbf{D}$.

Step 5: Determine the distance d_i^+, d_i^- between the object i and M^+, M^-, as shown in Eqs 11, 12

$$d_i^+ = \sqrt{\sum_{j=1}^{N}\left(r_{ij} - M_j^+\right)^2}, i = 1, 2, \cdots, N \tag{11}$$

$$d_i^- = \sqrt{\sum_{j=1}^{N}\left(r_{ij} - M_j^-\right)^2}, i = 1, 2, \cdots, N \tag{12}$$

Step 6: Fuzzy optimal decision-making.

Let the assessment object i obeys the fuzzy positive ideal with affiliation degree μ_i, as shown in Eq. 13

$$\mu_i = \frac{d_i^-}{d_i^+ + d_i^-}, i = 1, 2, \cdots, N \tag{13}$$

Obviously $0 \le \mu_i \le 1$, the closer r_{ij} is to M^+, the closer μ_i is to 1. Utilizing the classification results of the degree of affiliation to rank the merits of the samples can get the fuzzy expert group assessment set of the multi-level fuzzy comprehensive evaluation model.

2.3 Multi-level fuzzy comprehensive assessment model based on XGboost algorithm

XGBoost (Chen and Guestrin, 2016) is an integrated learning algorithm based on gradient advancement that shows good performance in classification and regression problems. To reduce the influence of subjective factors brought about by expert experience and to avoid errors caused by data redundancy or error omission, this paper adopts a combination of eclectic fuzzy decision-making and the XGboost algorithm to improve the assessment accuracy of DT. This model integrates multiple weak learners together to build a strong learner, as follows:

For a dataset $\mathbf{D} = \{x_i, y_i\}$ containing n samples and m features, the output values of the integrated model with K weak learners are as shown in Eq. 14:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \tag{14}$$

TABLE 3 Index transformation of triangular fuzzy number method.

Quantitative value attributes	Cost indicator	Profitability indicator
(0,0,1)	Highest	Lowest
(1,1,2)	Very high	Very low
(2,3,4)	High	Low
(4,5,6)	General	General
(6,7,8)	Low	High
(7,8,9)	Very low	Very high
(9,10,10)	Lowest	Highest

Where f is a function of each classification and regression tree and $f(x) = \omega_{q(x)}(w \in R^T, q: R^M \to T)$. For each tree, q denotes the tree structure that maps samples to specific leaf nodes, T denotes the number of leaf nodes, and ω denotes the weights of the leaf nodes.

The XGBoost model is trained by additive approach, the optimal structure of this model is found by successively adding tree and segmentation features. Therefore, the predicted value of the nth tree is $\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i)$.

The objective function of the final model consists of two parts, the loss function l and the regularization term Ω, as shown in Eq. 15:

$$obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (15)$$

where the loss function represents the predictive power of the model and the regularization term restricts the structure of the tree, as shown in Eq. 16:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \quad (16)$$

where γ and λ are two parameters that control the complexity, and the smaller their values are, the more complex the tree structure is.

A second-order Taylor expansion of the objective function can be approximated as Eq. 17:

$$obj^{(t)} = \sum_{i=1}^{n} \left[l(y_i, \hat{y}^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$$
$$+ \text{constant} \quad (17)$$

Where $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ and $h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$ are the first and second order partial derivatives of the loss function, respectively, and the objective function after expanding the regularization term and removing the constant term is expressed as Eq. 18:

$$obj^{(t)} = \sum_{j=1}^{T} \left[\left(\sum_{i \in \mathbf{I}_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in \mathbf{I}_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \quad (18)$$

where $\mathbf{I}_j$ is the index set that assigns the data point to the j leaf node.

In Eq. 7, ω_j is independent, therefore, the minimal value of the objective function and the corresponding ω_j^* can be obtained by direct derivation of ω_j, as shown in Eqs 19, 20.

$$\omega_j^* = -\frac{\sum g_i}{\sum h_i + \lambda} \quad (19)$$

$$obj^* = -\frac{1}{2} \sum_{j=1}^{T} \frac{(\sum g_i)^2}{\sum h_i + \lambda} + \gamma T \quad (20)$$

After obtaining the weight distribution, it is replaced with step 3 in the basic solution step of eclectic fuzzy decision-making, and then the final expert group assessment set is obtained by repeated iterations.

3 Case study

In this section, the proposed state assessment approach is verified in the distribution transformers. We obtained the data,

FIGURE 2
Percentage of transformer health status in the data set.
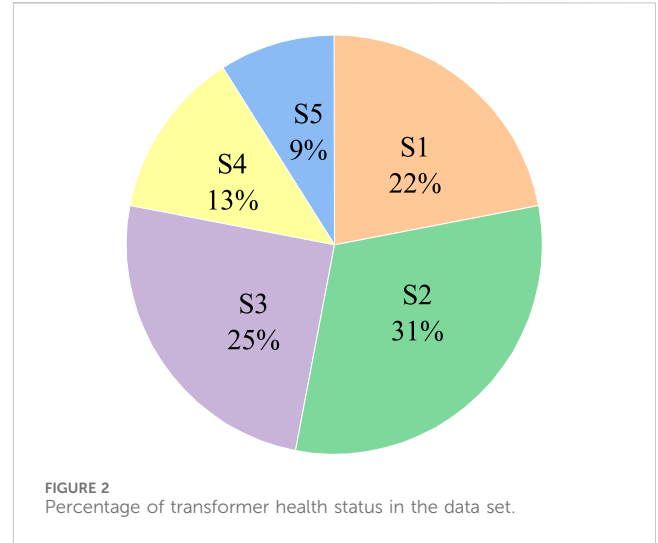
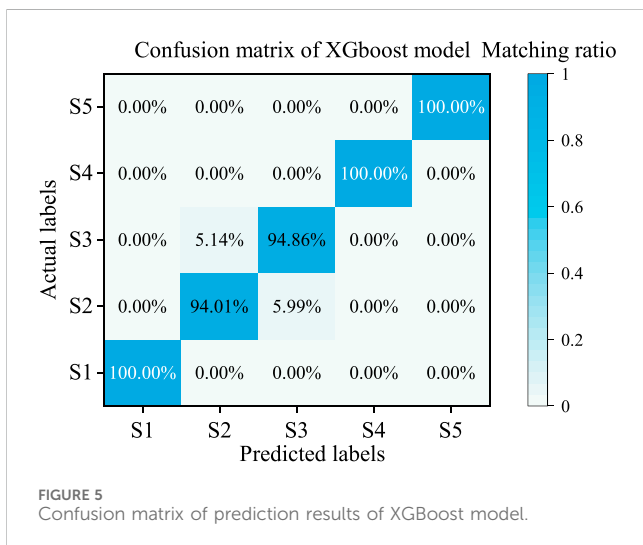including state parameters and health index, from DTs of 50 units in 20 maintenance periods.
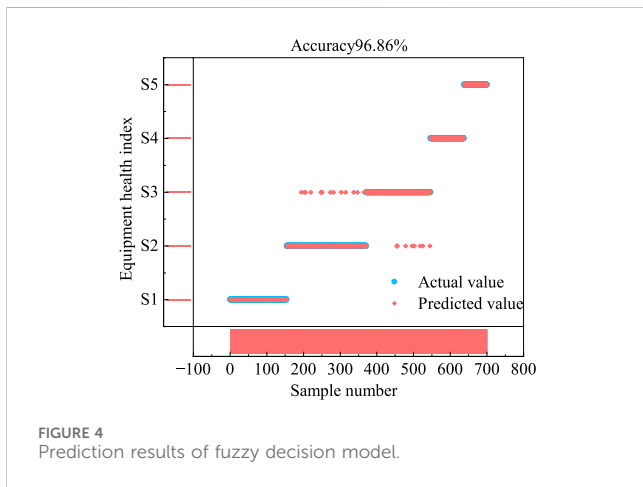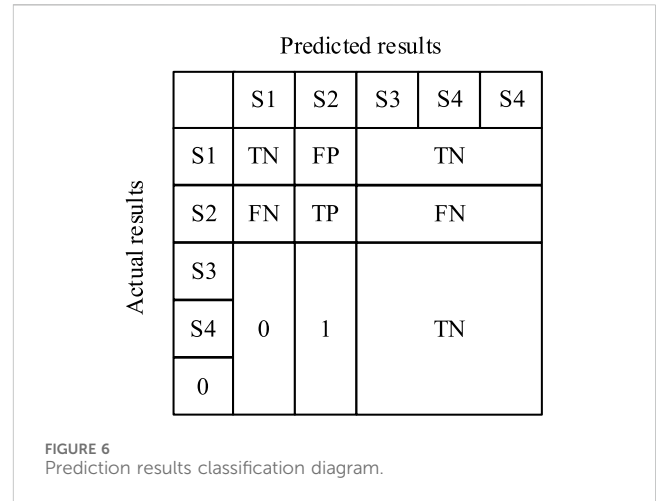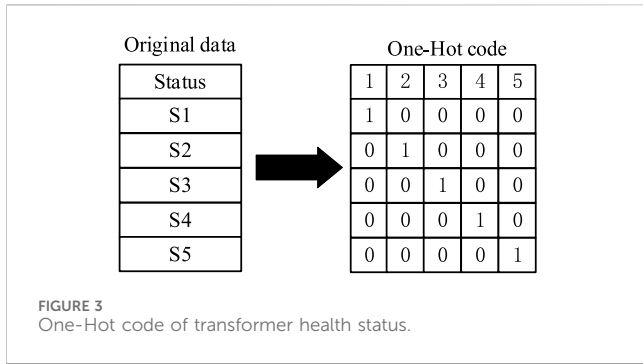
3.1 DT basic parameters

Table 2 displays the 13 important state quantities of the DT, which consist of four quantitative and nine qualitative markers. The rated values of the five major factor sets of DT are derived using the frequency statistics method in order to reduce data redundancy and clearly demonstrate the algorithm's accuracy. This method includes four qualitative metrics (grounding down conductor appearance μ6, sealing ability μ1, withstand voltage test μ2, and identification integrity μ8) in addition to one quantitative metric (winding DC resistance). The relevant data are divided into training and test sets according to the ratio of 7:3. All the cases were obtained by using Python 3.8 in a 3.4 GHz Intel Core i5-7500 computer with 8 GB of RAM, and the configuration of XGBoost-related parameters is shown in Supplementary Table S4.

3.2 Data pre-processing

According to the health index of the DT, it is categorized into [S1,S2,S3,S4,S5] five states. S1 indicates that the transformer is in good condition with low risk of failure. S5 indicates that the transformer is in very poor condition with a high risk of failure. The lower the status number, the better the condition of the transformer. The percentage of transformers in various health states in the dataset is shown in Figure 2.

The dataset's character-labeled data must first undergo preprocessing in order to be used. There are five values that make up the health index: good, normal, attention, abnormal, and serious, which are converted into (Liang et al., 2009; Yuan et al., 2019; Tamma et al., 2021; Guan, 2022; Fang et al., 2023). Figure 3 shows the result of encoding the transformer health state using one-hot encoding.

FIGURE 3
One-Hot code of transformer health status.

FIGURE 4
Prediction results of fuzzy decision model.

FIGURE 5
Confusion matrix of prediction results of XGBoost model.

3.3 Analysis of example results

In this part, the health status of the transformer is evaluated by fuzzy iteration and XGBoost. The model is applied to the test set and the predictions of the model can be calculated as shown in Figure 4.

FIGURE 6
Prediction results classification diagram.

The actual and predicted categories are formed into rows and columns of a matrix, respectively, each element in the confusion matrix represents a categorization result, and the elements on the diagonal line indicate the number of correct predictions. Obviously, the lager the elements on the diagonal, the better the confusion matrix, and the confusion matrix is shown in Figure 5. The accuracy of S2 is 94.01%, the accuracy of S3 is 94.86%, and the accuracy of the other three categories is 100%. The model produces confusion only between S2 and S3, and all other categories are correctly categorized.

In the multiclassification problem, as shown in Figure 6, if S2 is set as a positive sample and the other categories are considered as negative samples, we can classify the results as True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). Similarly, the other categories can be divided in this way. We can calculate the situation of the evaluation indicators for each category based on the prediction results, and we can calculatethe the suituation of the evaluation indicators by the model through the macro-averaging method.

Accuracy is the percentage of correct predictions to the total sample, and is expressed as Eq. 21:

$$A = \frac{TP + TN}{TP + FP + TN + FN} \tag{21}$$

Precision is the percentage of true positive samples to the total positive samples in the predicted results, and is expressed as Eq. 22:

$$P = \frac{TP}{TP + FP} \tag{22}$$

Recall is the percentage of true positive samples to actual positive samples in the predicted results, and is expressed as Eq. 23:

$$R = \frac{TP}{TP + FN} \tag{23}$$

The F1 indicator (F1-score) is the average of precision and recall, and can be expressed as Eq. 24:

$$F1 = \frac{2PR}{P + R} \tag{24}$$

According to the above equation, we can calculate the precision index of the evaluation results of the proposed model. The recall rate

FIGURE 7
Weight of feature state.

FIGURE 8
Comparison of SC curves of each mode.
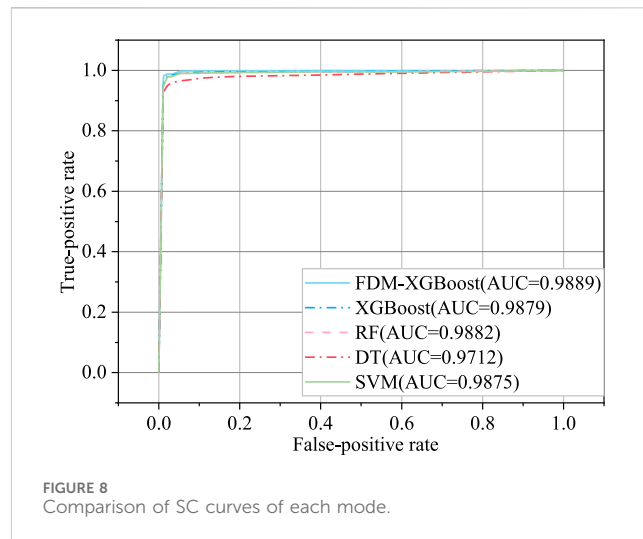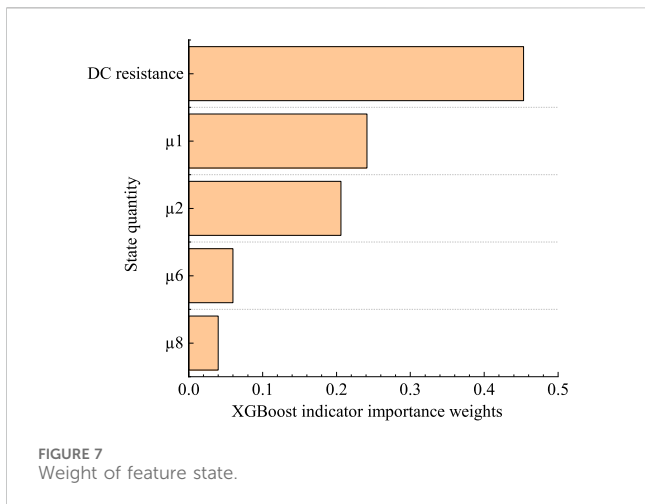
is 97.77%, the accuracy rate is 96.86%, the precision rate is 97.78%, and the F1 indicator is 97.77%. The evaluation indicators of the proposed method are all above 96%. Therefore, the method can accurately and comprehensively identify the aging condition of cables.

The importance weights of the characteristic state quantities are shown in Figure 7. Winding DC resistance, sealing ability and insulation performance are the three key variables for evaluating the health status of the transformer.

3.4 Comparative analysis of algorithms

In order to validate the performance of the proposed model, this section compares the algorithm proposed in the paper with other classification models concluding traditional XGBoost model, random forest (RF) (Breiman, 2001), decision tree (DT) (Fürnkranz et al., 2011), and support vector machine (SVM) (Cortes and Vapnik, 1995). These classification algorithms are briefly described below.

The DT model classifies the instances based on the feature values, the nodes of the decision tree contain judgments on the features, and then, the model outputs the classification results based on the judgments of each node.

The RF model consists of multiple independent decision trees, each decision tree in the forest classifies the samples individually, and the category with the highest score among all the decision tree is used as the classification result of RF.

SVM is a linear classifier, the idea is to find a suitable hyperplane for sample classification, it is usually used to deal with binary classification problems, but it also can be used to deal with multiclassification problems using a one-to-one approach.

The model evaluation results of the fuzzy decision-based XGBoost algorithm and the other four models are shown in Table 4. The model evaluation accuracy of SVM is the lowest among all the models, which is due to the fact that traditional SVM is a linear classifier and cannot handle nonlinear problems well. RF and XGBoost are integrated learning models, while DT belongs to the weak learner model, its performance is weaker than of RF and XGBoost models.

The XGBoost algorithm shows better evaluation compared to all other methods. The regular XGBoost algorithm increases the regularization term compared to other algorithms, which improves the model accuracy and avoids overfitting. And among all the algorithms, the model proposed in the paper has the best performance. From the comparative results, it is clear that the use of fuzzy decision making method to optimize the qualitative state quantities by incorporating them into the classification model does improve the performance of the XGBoost model.

The sensitivity (sensitivity curve,SC) curve is plotted based on the false positive rate (FPR) and true positive rate (TPR) of the model. The goodness of the model can be quantified by the area under the SC curve (AUC) (Bradley, 1997). The TPR and FPR are expressed as Eq. 25.

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \tag{25}$$

TABLE 4 Evaluation effect indicators of all models.

Model name	Accuracy (%)	Precision (%)	Recall (%)	F1 indicator (%)
This article model	96.86	97.78	97.77	97.77
Traditional XGBoost model	95.60	96.57	96.85	96.71
RF model	95.20	96.17	96.51	96.33
DT model	93.87	94.91	94.22	94.06
SVM model	93.33	92.35	92.73	92.26

FIGURE 9
Comparison of P-R curves of each model.
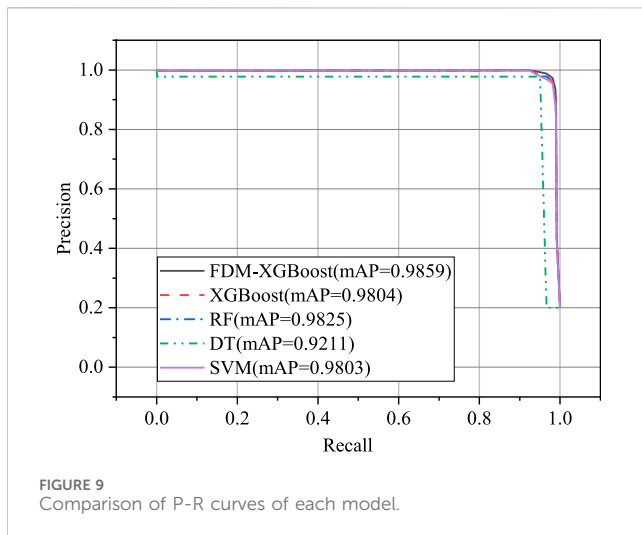
The SC curves for each model are shown in Figure 8. If the prediction is completely random, the curve is a straight line with slope 1. Among the multiple curves, the curve positions and AUC values allow a visual comparison of the model's performance. The larger the AUC value, the more accurate the prediction. The curve of the fuzzy decision-based XGBoost model is closer to the upper left corner with an AUC of 0.988 9, it indicates that the model has a good performance in prediction accuracy.

The P-R curve can be plotted on the basis of the checking accuracy and the checking precision. In multiple classification problems, the area under the P-R curve is called the mean of average precision (mAP) for each category, and this value describes the accuracy of classification. The P-R curve is shown in Figure 9. The fuzzy decision based XGBoost model has the best curve performance with the largest mAP value among all models at 0.985 9. The DT model with the smallest mAP, it is 0.9211.

4 Conclusion

Equipment breakdowns are the primary source of voltage instability, power imbalance, and unreliability in power systems. This work proposes a transformer health state evaluation model based on the fuzzy decision-making XGBoost algorithm in order to precisely analyze the DT health state. Nevertheless, the traditional XGBoost algorithm cannot quantitatively measure the assessment indexes of transformer health state, in order to overcome the difficulty, this paper combines the fuzzy iterative method with the XGBoost algorithm, constructs the mapping relationship between the key indexes and the state scores of the equipment, and puts forward a fuzzy decision-making based rapid assessment method of the state of the distribution equipment, which realizes the multi-source data fusion of systematic assessment.

The experimental results show that the accuracy, precision, recall and F1 indicator of the fuzzy decision-based XGBoost model are 96.86%, 97.78%%, 97.77%% and 97.77%%, respectively, the result are superior to the traditional XGBoost, RF, DT and SVM models mentioned in the paper. By comparing with the XGBoost model, which is constructed directly using quantitative parameters, the XGBoost model useing fuzzy decision theory does improve the evaluation performance. In addition,

the AUC and mAP values of the XGBoost model are larger than the other three models, indicating that the proposed model has better overall performance. The results show that the XGBoost transformer health state assessment model proposed in the paper is more accurate, and the model can effectively assess the DT state.

Author contributions

WQ: Conceptualization, Writing–original draft. LY: Data curation, Writing–review and editing. WX: Formal Analysis, Writing–original draft. YK: Funding acquisition, Writing–original draft. LY: Investigation, Writing–original draft. XZ: Methodology, Writing–review and editing. LG: Project administration, Writing–review and editing. LY: Resources, Writing–review and editing. TJ: Software, Writing–review and editing.

Conflict of interest

Authors WX, XZ, LG, and LYl were employed by State Grid Henan Electric Power Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declare that this study received funding from State Grid Henan Electric Power Company. The funder had the following involvement in the study: collection and analysis of data.

References

Ahmad, T., and Senroy, N. (2020). Statistical characterization of PMU error for robust WAMS based analytics. *IEEE Trans. Power Syst.* 35, 920–928. doi:10.1109/tpwrs.2019.2939098

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* 30 (7), 1145–1159. doi:10.1016/s0031-3203(96)00142-2

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/a:1010933404324

Chen, T., and Guestrin, C. (2016). "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Chen, Z. (2017). *Fault prediction of distribution network equipment based on bp neural network.* Master's Thesis. Guangdong, China: Guangdong University of Technology.

China Electric Power (2008). *Q/GDW 168—2008 regulations of condition-based maintenance test for electric equipment.* Beijing, China: China Electric Power Press.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn* 20, 273–297. doi:10.1007/BF00994018

Fang, M., Huang, R., and Lv, T. G. (2023). "Research on state assessment method of key equipment in distribution network," in *Asia conference on power and electrical engineering* (Tianjin, China: ACPEE), 2614–2618.

Fürnkranz, J. (2011). "Decision tree," in *Encyclopedia of machine learning.* Editors C. Sammut and G. I. Webb (Boston, MA: Springer). doi:10.1007/978-0-387-30164-8_204

Guan, C. X. (2022). *Evaluation and prediction of distribution network operation status based on multi-source data fusion.* Master's Thesis. Shenyang, China: Shenyang University of Technology.

Guo, J., and Liu, C. S. (2005). Research on lightning accidents and lightning protection measures of distribution transformer. *J. Electr. Power Sci. Technol.* 3, 14–17.

Liang, J. L., Wang, Z. D., and Liu, X. (2009). State estimation for coupled uncertain stochastic networks with missing measurements and time-varying delays: the discrete-time case. *IEEE Trans. Neural Netw.* 20, 781–793. doi:10.1109/tnn.2009.2013240

Lv, B. (2022). *Health evaluation of main transformer in power plant based on AHP-fuzzy comprehensive evaluation metho.* Chongqing, China: Chongqing University.

State Grid Corporation (2011). *Q/GDW 645-2011 Guidelines for state evaluation of distribution network equipment.* Beijing, China: State Grid Corporation.

Tamma, W. R., Prasojo, R. A., and Suwarno, (2021). High voltage power transformer condition assessment considering the health index value and its decreasing rate. *High. Volt.* 6, 314–327. doi:10.1049/hve2.12074

Wang, N., and Zhao, F. (2020). An assessment of the condition of distribution network equipment based on large data fuzzy decision-making. *Energies* 13, 197. doi:10.3390/en13010197

Wang, Y. L., Zhao, X. P., and Bian, J. (2012). "Cloud model-based risk assessment of power transformer," in *International conference on high voltage engineering and application* (IEEE).

Xie, H. X., Shi, L. P., and Hui, Z. Y. (2012). Research on immune clustering algorithm for transformers fault diagnosis. *Electr. Meas. Instrum.* 49, 15–18.

Yuan, F., Guo, J., Xiao, Z. H., Zeng, B., Zhu, W., and Huang, S. (2019). A transformer fault diagnosis model based on chemical reaction optimization and twin support vector machine. *Energies* 12, 960. doi:10.3390/en12050960

Zadeh, L. A. (1965). Fuzzy sets. *Inf. Control* 8 (3), 338–353. doi:10.1016/s0019-9958(65)90241-x

Zhang, Z., Zhao, W. Q., and Zhu, Y. L. (2010). State evaluation of power transformer based on support vector regression. *Electr. Power Autom. Equip.* 30, 81–84.

Zhang, J. W., Yang, Y., Weng, Y., and Zhang, N. (2020). Topology identification and line parameter estimation for non-PMU distribution network: a numerical method. *IEEE Trans. Smart Grid* 11, 4440–4453. doi:10.1109/tsg.2020.2979368

Zhou, J. S., Yu, J. F., and Yang, H. H. (2020). A condition assessment method of transformers based upon the dynamic grey target with theinterval grey number. *J. Electr. Power Sci. Technol.* 35, 133–140.

Zhu, Y. L., Shen, T., and Li, Q. (2008). Transformer condition assessment based on support vector machine and DGA. *J. Electr. Power Syst. Automation* 24, 47–50.

Appendix A

TABLE A1 Critical state evaluation set of DT.

Component	State quantity	Reflected state
Winding and bushing	DC resistance	DC resistance exceeds the range
	Insulation resistance	Insulation resistance is not normal
	Temperature	The temperature of the joint is abnormal and the temperature rise is abnormal
	Load rate	Overload
	Degree of contamination	Severely contaminated or rusted appearance
	Appearance integrity	Damaged appearance
	The temperature of respirator	Exceed the factory defaults
	Three-phase unbalance rate	Three-phase unbalance rate is not normal
Tap changer	Performance	Operation is not Normal
Cooling system	Mechanical properties	Dry change fan vibration is not normal
	Temperature	Temperature control device is abnormal
Tank	Ground distance of the bench	The distance to the ground is not enough
	Sealing	Finishing seal aging
	Oil level	Oil level is not normal
	Oil temperature	Oil temperature is abnormal
Non-electricity protection device	Insulation resistance	Unqualified insulation
Ground wire	Exterior	Insufficient connection or insufficient depth of grounding body
Insulation	Grounding resistance	Grounding resistance is abnormal
	Withstand voltage test	Pressure resistance is unqualified
Identification	Identification integrity	Equipment identification is vague, incomplete, wrong, etc.