Research Article

Calculation of a Health Index of Oil-Paper Transformers Insulation with Binary Logistic Regression

Weijie Zuo,[1] Haiwen Yuan,[1] Yuwei Shang,[2] Yingyi Liu,[1] and Tao Chen[3]

[1]*School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China*
[2]*Distribution System Department, China Electric Power Research Institute, Beijing 100192, China*
[3]*College of Electrical and Control Engineering, North China University of Technology, Beijing 100144, China*

Correspondence should be addressed to Yingyi Liu; liuyingyi@buaa.edu.cn

This paper presents a new method for calculating the insulation health index (HI) of oil-paper transformers rated under 110 kV to provide a snapshot of health condition using binary logistic regression. Oil breakdown voltage (BDV), total acidity of oil, 2-Furfuraldehyde content, and dissolved gas analysis (DGA) are singled out in this method as the input data for determining HI. A sample of transformers is used to test the proposed method. The results are compared with the results calculated for the same set of transformers using fuzzy logic. The comparison results show that the proposed method is reliable and effective in evaluating transformer health condition.

1. Introduction

Distribution transformer is one of the most important components in the power grid. The knowledge about the insulation health condition of a transformer is essential for determining appropriate asset management decisions. Recently, the health index (HI) is a useful tool, which combines the current information about a transformer (such as the results of operating observations, field inspections, laboratory testing, etc.) with a single quantitative index, providing its overall health condition, for planning routine maintenance strategies and reflecting the failure rate, and enables transformer operator to obtain a clear idea which is to the end of life [1, 2]. However, two questions have been encountered for computing the HI of in-service transformers in power distribution networks.

One question is to find an accurate HI with fewer condition monitoring tests in order to improve efficiency due to large amount of it. Various online and offline measurements, broadly categorized as electrical tests, mechanical tests, thermal tests, and oil tests, were carried out on a transformer and a quantitative HI was calculated according to the relevant industrial standards. Although the combination of all the measurement results can provide a better assessment

on the condition of the transformers [3], it is not suitable for distribution transformer because the online condition monitoring system may not be directly implemented in distribution transformers by considering economic aspects [4]. Moreover, it is not very efficient with large amount of offline measurement results since some of them could be correlated. For example, the moisture content and breakdown voltage of the transformer oil are highly relevant [5]. Therefore, it is necessary to identify the most significant measurement results, which clearly represent the health condition of the transformer, in order to improve the HI calculation efficiency.

The other challenge is how to incorporate the significant measurement results into the transformer HI calculation. One relatively easy way depends largely on the basis of industry standards such as IEC, IEEE standards, and CIGRE recommendations with the value and the weighting factor of each parameter [6, 7]. Fuzzy set theory, membership functions, and expert rules can also be used to produce the transformer HI [1]. Since the weighting factor and the membership functions are subjective; the calculated health indexes might not be consistent. The purpose of calculating the HI of transformer is to determine its health status. The health status of transformer is categorized as very good, good, moderate, bad,

and very bad in [1]. It is classified by EA Technology, a specialist corporation in asset management solutions for owners and operators of electrical asset in UK, as slightly aging, obviously aging, aging beyond the normal range, and extremely poor state. A broad classification, which is good, moderate, and bad, is also used in an asset management and health assessment consulting company called AMHA in [1]. In this paper, the HI of the transformer is simply categorized as healthy or unhealthy. Healthy means the transformer has just been put into operation and unhealthy means the transformer is in extremely poor state and it is necessary to replace it. Three main methods have been used for binary classification such as support vector machines, neural networks, and logistic regression, each with its own advantages and disadvantages [8]. Neural networks [9] and support vector machines [10] have been proposed to determine the transformer HI. As far as we are concerned, use of binary logistic regression for evaluating the health condition of power transformer has not been published in the literature.

In this paper, binary logistic regression is preferred to classify the data to get the health status of a sample of transformers rated under 110 kV for the following reasons:

(1) Binary logistic regression has been widely used in many fields like social science and medical field. Compared with support vector machines and neural networks, the advantage of it is that it gives a direct estimate of the class probabilities [11, 12]. The way a doctor analyzes the different symptoms to check a human being with or without disease and suggests cure, same way can be used to predict the transformers' operational status with different internal as well as external parameters. Accordingly, a proper maintenance decision can be made. Drawing a parallel to the medicine, binary logistic regression may also be used to determine the health condition of transformer.

(2) Binary logistic regression can be used for testing of a statistical hypothesis to determine whether the independent variables in the model are significantly related to the outcome variable [13]. On the basis of the correct evaluation of the transformer's health condition, the whole calculation efficiency is improved since those insignificant variables can be eliminated and the compelling information can be selected with this method.

(3) The cumulative probability distribution of failure of the electrical components in distribution networks summarized in [14] is consistent with logistic curves. The generalized logistic distribution function was assumed as the logarithmic time to the failure of the electrical insulation equipment [15]. The growth and death rates of human with aging are in accordance with logistic curve. Considering the tight correlation between the transformer HI and the transformer failure rate and the analogy between the health condition of transformer and human being, logistic regression may be used to determine the health condition of the transformer.

TABLE 1: Input data for oil-paper transformer insulation HI calculation.

Number	Item
1	Water content in oil
2	Total acidity of oil
3	Dissolved combustible gases (DCG)
4	Oil breakdown voltage (BDV)
5	Dissipation factor (DF)
6	2-Furfuraldehyde content

The remainder of this paper consists of five sections. Section 2 explains the key parameters representing the health condition of the transformer. Section 3 includes binary logistic regression and how to use it to calculate the HI of the transformer. Section 4 shows the case study and the discussions. Section 5 concludes the paper.

2. Input Data for HI Calculation

Various tests can be used to determine the insulation health condition of transformer such as dissolved gas analysis (DGA), oil quality tests, dissipation factor, and insulation resistance. In this paper, water content in transformer oil, total acidity of oil, dissolved combustible gases (DCG), oil breakdown voltage (BDV), dissipation factor (DF), and 2-Furfuraldehyde are selected as the input data, shown in Table 1, for HI calculation. The input data is the same as [1] and the reasons are as follows.

Over the past decades, DGA technique has been applied to detect the transformer fault conditions [16] such as overheating, internal arcing, bad electrical contacts, and partial discharge. It is always used to provide enough information to evaluate the integrity of a transformer. However, it is difficult to calculate the health status of the transformer with all of gas value. Dissolved combustible gases (DCG), summation of the levels of H_2, CH_4, C_2H_4, C_2H_2, and C_2H_6, are used for determining HI in this paper in order to improve the calculation efficiency.

Measuring the water content and breakdown voltage of the transformer oil gives a good indication of the overall health condition of transformer oil [1]. The contaminated particles and the free water in transformer oil will cause the reduction of BDV and may cause increased partial discharges that accelerate the aging of transformer [16, 17]. Therefore, the water content in oil and BDV are used for HI calculation.

DF measures the power lost in the transformer oil during its operation [1]. The dissipated power, transferred to the oil in the form of heat energy, increases the temperature of the insulating materials and accelerates its aging process. In oil-paper transformers, much attention must be paid to the condition diagnosis of the solid insulation (paper) [14] and the degradation of it can be considered as the primary reason for the end of life [9]. It is known that the 2-Furfuraldehyde content, an important indication to the degree of polymerization, directly assesses the health of the solid insulation [1]. Therefore, DF and 2-Furfuraldehyde content are considered as input data for determining HI. Since the deterioration of
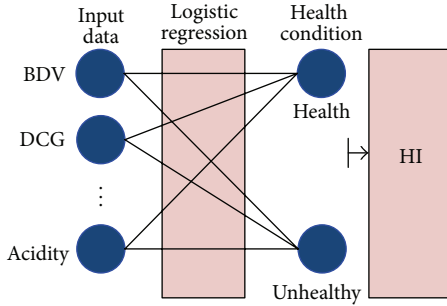
FIGURE 1: Logistic regression model for HI.

the transformer's overall insulation is very important and it also increases the acidity of the transformer oil [1], the total acidity of the transformer oil is also adopted in this paper to reflect to some extent the condition of the paper insulation.

3. Logistic Regression for HI Calculation

In this section, binary logistic regression, a type of probabilistic statistical classification model which can be used to measure the relationship between a categorical dependent variable and several independent variables by using probability scores as the predicted values of the dependent variable [18, 19], is used to explore the best fitting model to describe the relationship between the transformer HI and the set of independent variables like BDV, total acidity of oil, 2-Furfuraldehyde content, and DCG in order to determine the health condition of the transformer.

Assume we have a number of transformer classification samples from condition monitoring experiments. Each sample can be in one of the two classes: class 0 (healthy) and class 1 (unhealthy). As mentioned before, healthy means the transformer has just been put into operation and unhealthy means the transformer is in extremely poor state and it is necessary to replace it. A rule based on binary logistic regression is used to determine the probability that a sample belongs to one of the two classes.

Let HI be a variable indicating the health condition of the transformer: HI = 0 means the health condition of the sample transformer belongs to class 0 (healthy) and HI = 1 means that the health condition of sample transformer belongs to class 1 (unhealthy). So, HI is the probability that the transformer belongs to healthy or unhealthy. Let x_i be a variable indicating the set of independent variables like BDV, total acidity of oil, 2-Furfuraldehyde content, and so forth. The transformer health index for different input value x_i, illustrated in Figure 1, follows a logistic regression model expressed as [13]

$$HI = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{n} \beta_i x_i)}},$$ (1)

where β_0 is a constant and β_i is the coefficient reflecting the contribution of the independent variable x_i to the HI. That is to say, β_i is the assigned weighting factor for the input data x_i. Therefore, the weighting factor is objective and the proposed method has the ability to overcome the shortcoming that the calculated health indexes might not be consistent with fuzzy

set theory and expert system, which are subjective and based on the experience of the transformer expert, and might be different from one expert to another.

The unknown parameters β_0 and β_i shown in formula (1) can be estimated by maximum likelihood criterion. Suppose the total number of the transformers is K, $y_k = 0$ or $y_k = 1$ represents the health condition of each transformer being healthy or unhealthy, respectively, and HI_k represents the health index of the kth transformer. The likelihood function used to determine the unknown parameters β_0 and β_i is as follows [8, 11–13, 20]:

$$L(\beta) = \prod_{k=1}^{K} \left[\left(HI_k\right)^{y_k} \cdot \left(1 - HI_k\right)^{(1-y_k)} \right] \longrightarrow \max! \quad (2)$$

Maximizing formula (2) is equivalent to maximizing the following formulation:

$$l(\beta) = \sum_{k=1}^{K} \left\{ \left[y_k \cdot \ln\left(HI_k\right) + \left(1 - y_k\right) \cdot \ln\left(1 - HI_k\right) \right] \right\}. \quad (3)$$

The Newton-Raphson algorithm can be used to get these values of unknown parameters β_0 and β_i according to the following likelihood equation [20, 21]:

$$l'(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \sum_{k=1}^{K} \left\{ \left[y_k - HI_k \right] \cdot x_{ik} \right\} = 0. \quad (4)$$

Wald value, similar to the t value in multiple regression, is used to assess the significance of each coefficient in logistic regression. More information about it can be got in chapter 2.4 of [13]. The larger the Wald value is, the more the variable is significantly related to the health index. On the basis of the correct evaluation of transformer's health condition, those insignificant variables with small Wald value can be eliminated and the compelling information can be selected with this method.

Statistical Product and Service Solutions (SPSS) can be employed to carry out the health index calculation. The whole validity of this health index calculation method can be represented by Nagelkerke-R^2, which ranges between zero and unity and measures the proportion of data variation explained by the independent variables [13, 22, 23]. The proposed method is effective when the Nagelkerke-R^2 is larger than 0.5. Classification table, shown in Table 3, is used in this paper to evaluate the goodness-of-fit of the health index calculation formula (1) [13, 22]. The larger the value of percentage correctly classified is, the better the goodness-of-fit of the health index calculation model is. For detailed information about Wald value, Nagelkerke-R^2, and classification table, please refer to the relevant literatures. The flowchart for calculating the health index is shown in Figure 2.

4. Case Study

A sample of 30 transformers shown in [1] is used as the input data and SPSS is employed to carry out the health index calculation using binary logistic regression. The transformer,
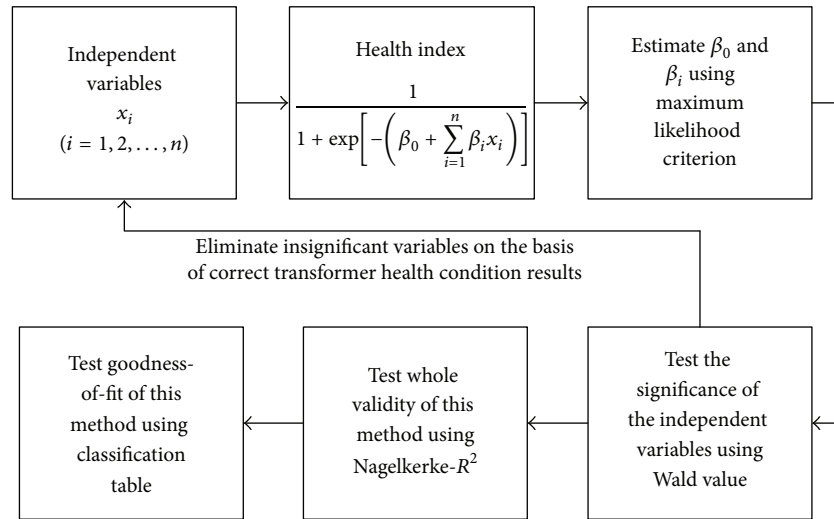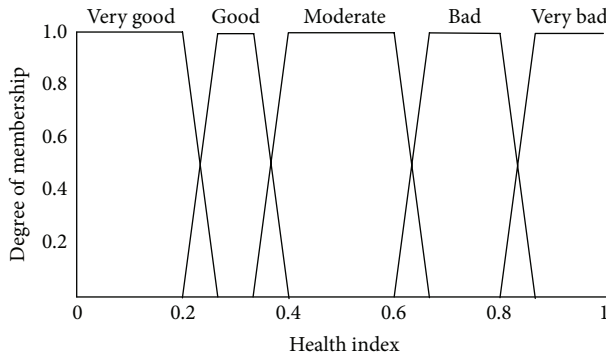
FIGURE 2: Flowchart for calculating the HI.

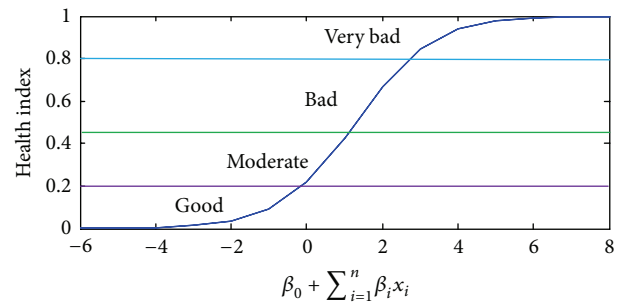FIGURE 3: Membership functions for HI in [1].

FIGURE 4: Health status classification.

TABLE 2: Model summary with six factors.

Step	$-2\log$ likelihood	Cox & Snell-R Square[a]	Nagelkerke-R Square
1	12.333	0.557	0.790

[a]Cox & Snell-R Square, similar to Nagelkerke-R Square, represents the whole validity of the logistic regression model and measures the proportion of data variation explained by the independent variables in logistic model; more information about it can be got from [23].

whose health index calculated in [1] is less (larger) than 0.5, is categorized as healthy (unhealthy) in this paper. The results of the diagnostic tests of a sample of 30 transformers in [1] are shown in the appendix.

Fuzzy set theory and six membership functions were used and 33 expert rules were determined in order to produce the transformer health index and its results were compared with AMHA [1]. The health index value is ranged between zero and unity. The transformer health index in [1] was classified into five conditions: very good, good, moderate, bad, and very bad shown in Figure 3. According to AMHA, the health index values between zero and 0.4 are good, 0.4 and 0.7 are moderate and 0.7 and unity are bad. In this paper, the health index value is also ranged between zero and unity and four health conditions are adopted according to EA Technology: zero to 0.2 is good, 0.2 to 0.45 is moderate, 0.45 to 0.8 is bad, and 0.8 to unity is very bad shown in Figure 4.

4.1. Six-Feature Factors for HI Calculation. The same as [1], the test results of the water content, acidity content, BDV, dissipation factor (DF), DCG, and 2-Furfuraldehyde content are taken as the input data for the proposed HI calculation method in this section. The unknown parameters β_0 and β_i in formula (1) and the Wald value are shown in Table 4. The Nagelkerke-R^2 and the classification table are shown in Tables 2 and 3, respectively. The HI calculation method using logistic regression is effective since the value of Nagelkerke-R^2 is 0.79, larger than 0.5 in Table 2. From Table 3 the percentage correct for the healthy (unhealthy) transformer is 90.5% (100%) and the overall percentage is 93.3%. The goodness-of-fit of the HI calculation model is good. According to [24], the dissipation factor and total acidity of oil are highly relevant and the relationship between them is log function. Therefore, in order to optimize and simplify the input data, the parameter with less importance may be deleted. Additionally, from Table 4, Wald value of dissipation factor is 0.053, the smallest value of

TABLE 3: Classification table with six factors.

Actual group	Predicted		Actual group size	Percentage correctly classified
	0	1		
0 (health)	19[a]	2[a]	21	90.5%[b]
1 (unhealthy)	0[a]	9[a]	9	100%[b]
Predicted group size	19	11	30	93.3%[b]

[a]The number of transformers correctly assigned to group 0 (healthy) is 19, whereas 2 transformers of group 0 (healthy) are incorrectly assigned to group 1 (unhealthy). Similarly, the number of correct classifications to group 1 (unhealthy) is 9, and the number of incorrect assignments to group 1 (unhealthy) is 0.
[b]Percentage correctly classified = (number correctly classified/total number of observations) × 100%.
For example, 90.5% = (19/21) × 100%; 93.3% = [(19 + 9)/30] × 100%.

TABLE 4: Variables evaluation with six factors.

Variables	B	SE	Wald	Df	Sig.[a]
Water x_1	−0.246	0.211	1.358	1	0.244
Acidity x_2	28.604	25.470	1.261	1	0.261
BDV x_3	−0.171	0.111	2.365	1	0.124
DF x_4	0.112	0.487	0.053	1	0.818
DCG x_5	0.01	0.009	1.514	1	0.219
2-Furfuraldehyde x_6	0.254	0.376	0.456	1	0.499
Constant	5.45	6.957	0.614	1	0.433

Note: B: the estimated coefficient; SE: standard error of the estimate; Wald: Wald value; Df: degree of freedom; Sig.: significance value.
[a]Sig. is also used to assess the significance of each coefficient in logistic regression [13]. The smaller the Sig. value is, the more the variable is significantly related to the logistic regression.

all the input six parameters, which means that it is the most unimportant parameter for calculating the insulation health index of transformer with the proposed method. Therefore, the dissipation factor is removed from the input data.

4.2. Five-Feature Factors for HI Calculation. Deleting the dissipation factor, the test results of the water content, acidity content, BDV, DCG, and 2-Furfuraldehyde content are used as the input data for the proposed HI calculation method. As shown in Tables 5 and 6, the Nagelkerke-R^2 is 0.782 and the overall percentage is 90%. The health index method is effective and the goodness-of-fit is good. Since the moisture content and oil breakdown voltage are highly relevant and the Wald value of the water content is 1.065 shown in Table 7, relatively smaller than the other four test results, the water content can be removed from the input data for the next calculation.

4.3. Four-Feature Factors for HI Calculation. Deleting the water content, the test results of the acidity content, BDV, DCG, and 2-Furfuraldehyde content are used as the input data for the proposed health index method in this section. The unknown parameters β_0 and β_i in formula (1) and the Wald value are shown in Table 10. The Nagelkerke-R^2 and classification table are shown in Tables 8 and 9, respectively.

The health index method using binary logistic regression is effective since the value of Nagelkerke-R^2 is 0.754, larger than 0.5. From Table 9 the percentage correct for the healthy (unhealthy) transformer is 95.2% (88.9%) and the overall percentage is 93.3%. The goodness-of-fit of the health index

TABLE 5: Model summary with five factors.

Step	−2 log likelihood	Cox & Snell-R Square	Nagelkerke-R Square
1	12.571	0.552	0.782

calculation model is good. From Table 10, the unknown parameters β_0 and β_i are −1.291, 10.870, −0.08, 0.006, and 0.542, respectively. Therefore, the health index calculation model using binary logistic regression is as follows:

$$HI = \frac{1}{1 + e^{-(-1.291 + 10.87x_2 - 0.08x_3 + 0.006x_5 + 0.542x_6)}}, \quad (5)$$

where x_2, x_3, x_5, and x_6 represent acidity content, BDV, DCG, and 2-Furfuraldehyde content, respectively. The health index results of the 30 transformers are shown in Table 11.

4.4. Discussion

4.4.1. Discussion of the Calculation Results. Table 11 shows the health indices given by the proposed method, by fuzzy-logic method in [1], and by AMHA. To facilitate the comparison, the good health conditions produced by logistic regression method were compared with the very good and good category determined by fuzzy-logic decisions and good category from AMHA in [1]. The bad and very bad health conditions produced by logistic regression method were compared with bad and very bad category determined by fuzzy-logic decisions and bad category from AMHA in [1], respectively.

Although some numerical values of the health index are different, for example, the health index value of number 3 transformer is 0.007 and its value given by fuzzy-logic method is 0.3, they are not contradictory since the proposed method for health index calculation is based on the logistic curve. It is a common outcome on the transformers' health index, whose health conditions are good or very good because the initial stage of logistic curve is tending to zero.

From Table 11, three transformers' health conditions evaluated by the proposed method are different with the results shown in [1]: number 1, number 17, and number 25 transformers. It can be explained as follows.

For number 1 transformer, the health index calculated by AMHA is 0.377. However, according to AMHA, when the health index of the transformer is between 0.4 and 0.7, its health condition is moderate. The health condition of number 1 transformer tends to be moderate actually although it is

TABLE 6: Classification table with five factors.

Actual group	Predicted		Actual group size	Percentage correctly classified
	0	1		
0 (health)	19	2	21	90.5%
1 (unhealthy)	1	8	9	88.9%
Predicted group size	20	10	30	90%

TABLE 7: Variables evaluation with five factors.

Variables	B	SE	Wald	Df	Sig.
Water x_1	−0.189	0.183	1.065	1	0.302
Acidity x_2	19.532	16.515	1.399	1	0.237
BDV x_3	−0.149	0.105	2.022	1	0.155
DCG x_5	0.008	0.006	1.658	1	0.198
2-Furfuraldehyde x_6	0.413	0.242	2.903	1	0.088
Constant	4.539	6.934	0.428	1	0.513

TABLE 8: Model summary with four factors.

Step	−2 log likelihood	Cox & Snell-R Square	Nagelkerke-R Square
1	13.902	0.532	0.754

classified as good. Therefore, it might not be incorrect that the health condition of number 1 transformer is categorized as moderate with the proposed method.

For number 17 transformer, the health condition is categorized as bad with the proposed method. However, the health index calculated in [1] is 0.53, according to Figure 2, when the health index of the transformer is between 0.6 and 0.8. Its health condition is bad. The health condition of number 17 transformer tends to be bad actually although it is classified as moderate. Therefore, it might not be called incorrect that the health condition of number 17 transformer is categorized as bad with the proposed method.

For number 25 transformer, the health condition is classified as moderate in [1]. In this paper, the health index is 0.181 and the health condition is categorized as good. However, when the health index of the transformer is between 0.2 and 0.45 with the proposed method, its health condition is moderate. The health condition of number 25 transformer tends to be moderate actually although it is classified as good. Therefore, it might not be called incorrect that the health condition of number 25 transformer is categorized as good with the proposed method.

Table 11 also shows that, by using the proposed method, the evaluation results of the remaining 27 transformers health condition in this paper are consistent with the evaluation results in [1]. Therefore, binary logistic regression can be used to determine the correct health condition of the transformer.

4.4.2. Discussion of the Selected Input Variable. The first rule of 33 expert rules determined in [1] is if 2-Furfuraldehyde is very bad, then the health index is very bad. As mentioned before, 2-Furfuraldehyde content, an important indication to the degree of polymerization, directly assesses the health of

the solid insulation. From Table 10, the Wald value of the 2-Furfuraldehyde content is 3.393, maximum value in the four pieces of input data. It clearly shows that 2-Furfuraldehyde content is the most important parameter in the proposed method, which is the same conclusion as [1].

Dissipation factor and water content are removed in this proposed method. The reasons are as follows.

The moisture content and BDV are highly relevant. The dissipation factor and total acidity of oil are also highly relevant and the relationship between them is log function [24]. Since BDV and total acidity of oil are adopted in this paper, dissipation factor and water content can be removed. BDV and acidity of the transformer oil can provide a good symbol of the health condition of the oil. 2-Furfuraldehyde analysis gives a highly convincing transformer health index result since it is related to the degree of polymerization of the paper. Analyzing DCG in the oil also provides a good indicator of transformer aging. Therefore, 2-Furfuraldehyde, DCG, BDV, and acidity, selected by the binary logistic regression, have the compelling information for determining the oil-paper transformer insulation health index.

For reasons discussed above, binary logistic regression can be utilized to determine the correct health condition of the transformer insulation. It can also be used to identify the important available tests for calculating the distribution transformer health index in order to improve the calculation efficiency.

As far as we are concerned, logistic regression does not face the strict assumptions, such as multivariate normality compared with the discriminant analysis, and is much more robust when these assumptions are not met, making its application appropriate in many situations [22]. So, technically, many parameters can be handled with logistic regression. For example, in [22], 18 parameters were used as the input data for logistic regression. Therefore, it is possible that more parameters, not only just related to the oil-paper transformer insulation, can be used to determine the health condition of the transformers with the proposed method.

4.5. Application. Another 110 kV transformer, shown in [25], has been in service for 25 years until April 2014 in China. No accident has ever happened with good running environment and load rate is about 0.8. The test data is as follows: acid value is 0.172 mgKOH/g; BDV is 39 kV; DCG is 359.72 μL/L; 2-Furfuraldehyde content is 3.46 mg/L. The health index is 0.816 calculated with formula (5). According to Figure 4, the health condition of the transformer is very bad and nearly out of service, the same conclusion with higher calculation efficiency with [25].

TABLE 9: Classification table with four factors.

Actual group	Predicted		Actual group size	Percentage correctly classified
	0	1		
0 (health)	20	1	21	95.2%
1 (unhealthy)	1	8	9	88.9%
Predicted group size	21	9	30	93.3%

TABLE 10: Variables evaluation with four factors.

Variables	B	SE	Wald	Df	Sig.
Acidity x_2	10.87	16.851	0.416	1	0.519
BDV x_3	−0.08	0.08	1.003	1	0.317
DCG x_5	0.006	0.006	0.974	1	0.324
2-Furfuraldehyde x_6	0.542	0.294	3.393	1	0.065
Constant	−1.291	4.009	0.104	1	0.747

TABLE 11: Health index and health condition of transformer.

Number	HI		HI in [1]		AMHA HI	
1	0.434	M	0.36	G	0.377	G
2	0.17	G	0.3	G	0.334	G
3	0.007	G	0.3	G	0.29	G
4	0.826	VB	0.78	B	0.7	B
5	0.002	G	0.2	VG	0.102	G
6	0.004	G	0.3	G	0.274	G
7	0.023	G	0.3	G	0.316	G
8	0.006	G	0.3	G	0.29	G
9	0.003	G	0.22	VG	0.226	G
10	0.12	G	0.3	G	0.316	G
11	0.973	VB	0.94	VB	1	B
12	0.841	VB	0.93	VB	0.931	B
13	0.973	VB	0.94	VB	1	B
14	0.589	B	0.83	B	0.916	B
15	0.885	VB	0.78	B	0.732	B
16	0.03	G	0.3	G	0.354	G
17	0.594	B	0.53	M	0.45	M
18	0.037	G	0.3	G	0.291	G
19	0.997	VB	0.94	VB	1	B
20	0.005	G	0.3	G	0.347	G
21	0.077	G	0.15	VG	0.414	M
22	0.002	G	0.11	VG	0.241	G
23	0.895	VB	0.94	VB	0.953	B
24	0.018	G	0.3	G	0.368	G
25	0.181	G	0.51	M	0.45	M
26	0.004	G	0.11	VG	0.072	G
27	0.045	G	0.42	M	0.371	G
28	0.006	G	0.3	G	0.225	G
29	0.015	G	0.3	G	0.241	G
30	0.2	M	0.48	M	0.45	M

Note: M: moderate; G: good; VB: very bad; B: bad; VG: very good.

A total of one hundred 10 kV oil-paper transformers were collected in this paper. Their health conditions determined

TABLE 12: Comparison of the health conditions produced by the presented method and the health conditions produced by method in [26].

Method in [26]	Proposed method				Total by method in [26]
	G	M	B	VB	
Normal condition	30	3			33
Attention condition	2	25	1		28
Abnormal condition		2	22		24
Severe condition				15	15
Total by proposed method	32	30	23	15	100

Note: G: good; M: moderate; B: bad; VB: very bad.

by the proposed method are shown below: 32 transformers in good condition, 30 transformers in moderate condition, 23 transformers in bad condition, and 15 transformers in very bad condition. Their health conditions were also categorized as four types: normal condition, attention condition, abnormal condition, and severe condition according to the enterprise standard of State Grid Corporation of China (SGCC) in 2011 [26]. The results are shown in the last column in Table 12. To facilitate the comparison with these two methods, the good health index condition produced by logistic regression is compared with normal condition category from the enterprise standard of SGCC. Similarly, moderate, bad, and very bad conditions produced by this proposed method are compared with attention condition, abnormal condition, and severe condition category from the enterprise standard of SGCC, respectively. Table 12 summarizes this comparison and it shows eight cases from the results of the proposed method are not identical to the enterprise standard of SGCC method. For example, 2 transformers, shown in the third row of column 2 in Table 12, were categorized as attention condition by the enterprise standard of SGCC method. However, these conditions were classified as good condition by the proposed method. So, the comparison result is a 92% match.

5. Conclusion

A procedure of calculating health index for determining the insulation health condition of oil-paper transformer using binary logistic regression is introduced in this paper. Acid value, BDV, 2-Furfuraldehyde, and DGA are singled out as the inputs for binary logistic regression for calculating

TABLE 13: Data in [1].

Number	x_1^* (ppm)	x_2^* (mgKOH/g)	x_3^* (kV)	x_4^* (%)	x_5^* (ppm)	x_6^* (mg/L)
1	21.7	0.024	32.5	0.075	483	0.86
2	26.9	0.098	40.5	0.894	254	0.65
3	14.5	0.033	58	0.14	78	0.26
4	21.2	0.226	48.7	0.424	215	5.53
5	10	0.01	75	0.111	126	0.06
6	15.5	0.075	71	0.143	38	0.53
7	16.8	0.167	70.1	0.255	149	0.78
8	15	0.092	67.8	0.211	28	0.69
9	17	0.035	62.7	0.113	9	0.21
10	30	0.088	37.6	0.353	197	0.31
11	16.2	0.181	25.5	0.201	35	8.76
12	15	0.155	37.5	0.182	53	7.29
13	16.8	0.115	25.6	0.174	78	9.6
14	15	0.21	57.2	0.22	53	6.69
15	27.6	0.089	30.4	0.128	336	5.12
16	23.5	0.106	45.8	0.207	30	0.24
17	24.8	0.012	29.9	0.068	504	1.68
18	23.6	0.07	39.2	0.203	22	0.5
19	30.5	0.073	28.7	67	30	15.7
20	18.4	0.063	64.5	0.243	69	0.15
21	21.1	0.019	28.4	0.025	144	0.02
22	8.1	0.01	66.9	0.042	71	0.05
23	19.6	0.216	41.1	0.264	48	7.54
24	6	0.01	67.6	0.126	427	0.08
25	18.4	0.152	37.2	0.299	81	1.14
26	11.1	0.032	67.2	0.089	119	0.04
27	21.5	0.147	60.8	0.938	168	0.92
28	7.5	0.16	70.1	0.448	10	0.06
29	13	0.091	51.6	0.369	8	0.32
30	35.7	0.229	41.4	0.639	24	1.07

$^*x_1, x_2, x_3, x_4, x_5$, and x_6 represent water content in oil, total acidity of oil, oil breakdown voltage, dissipation factor, DCG, and 2-Furfuraldehyde content, respectively.

the sample of oil-paper transformers' health index. The presented method is tested using a sample of 30 transformers' field data in [1]. The health index results are similar with [1] which shows the effectiveness of the proposed method. The calculation efficiency may be improved and the overhaul cost for transformer may be reduced, since the proposed method deletes two pieces of input data, that is, dissipation factor and water content, compared with [1]. Next, further study to test the method with more field data will be conducted.

Appendix

See Table 13.

References

[1] A. E. B. Abu-Elanien, M. M. A. Salama, and M. Ibrahim, "Calculation of a health index for oil-immersed transformers rated under 69 kV using fuzzy logic," *IEEE Transactions on Power Delivery*, vol. 27, no. 4, pp. 2029–2036, 2012.

[2] A. N. Jahromi, R. Piercy, S. Cress, J. R. R. Service, and W. Fan, "An approach to power transformer asset management using health index," *IEEE Electrical Insulation Magazine*, vol. 25, no. 2, pp. 20–34, 2009.

[3] Y. Z. Yang Ghazali, M. A. Talib, and H. Ahmad Rosli, "TNB experience in condition assessment and life management of distribution power transformers," in *Proceedings of the 20th International Conference and Exhibition on Electricity Distribution (CIRED '09)*, pp. 1–4, Prague, Czech Republic, June 2009.

[4] A. A. Nelson, G. C. Jaiswal, M. S. Ballal, and D. R. Tutakne, "Remote condition monitoring system for distribution transformer," in *Proceedings of the 8th National Power Systems Conference (NPSC '14)*, pp. 1–5, IEEE, Guwahati, India, December 2014.

[5] A. D. Ashkezari, H. Ma, T. K. Saha, and Y. Cui, "Investigation of feature selection techniques for improving efficiency of power transformer condition assessment," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 21, no. 2, pp. 836–844, 2014.

[6] I. G. N. S. Hernanda, A. C. Mulyana, D. A. Asfani, I. M. Y. Negara, and D. Fahmi, "Application of health index method for transformer condition assessment," in *Proceedings of the IEEE Region 10 Conference (TENCON '14)*, pp. 1–6, Bangkok, Thailand, October 2014.

[7] A. Naderian, S. Cress, R. Piercy, F. Wang, and J. Service, "An approach to determine the health index of power transformers," in *Proceedings of the IEEE International Symposium on Electrical Insulation (ISEI '08)*, pp. 192–196, Vancouver, Canada, June 2008.

[8] J. Petersen, D. Austin, R. Sack, and T. L. Hayes, "Actigraphy-based scratch detection using logistic regression," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 2, pp. 277–283, 2013.

[9] A. E. B. Abu-Elanien, M. M. A. Salama, and M. Ibrahim, "Determination of transformer health condition using artificial neural networks," in *Proceedings of the International Symposium on INnovations in Intelligent SysTems and Applications (INISTA '11)*, pp. 1–5, IEEE, Istanbul, Turkey, June 2011.

[10] A. D. Ashkezari, H. Ma, T. Saha, and C. Ekanayake, "Application of fuzzy support vector machine for determining the health index of the insulation system of in-service power transformers," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 20, no. 3, pp. 965–973, 2013.

[11] L. Shen and E. C. Tan, "Dimension reduction-based penalized logistic regression for cancer classification using microarray

data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 2, no. 2, pp. 166–175, 2005.

[12] X. Gu, F.-L. Chung, H. Ishibuchi, and S. Wang, "Multitask coupled logistic regression and its fast implementation for large multitask datasets," *IEEE Transactions on Cybernetics*, vol. 45, no. 9, pp. 1953–1966, 2015.

[13] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, John Wiley & Sons, Hoboken, NJ, USA, 3rd edition, 2013.

[14] X. Zhang, E. Gockenbach, V. Wasserberg, and H. Borsi, "Estimation of the lifetime of the electrical components in distribution networks," *IEEE Transactions on Power Delivery*, vol. 22, no. 1, pp. 515–522, 2007.

[15] H. Hirose, T. Sakumura, and N. Tabuchi, "Optimum and semi-optimum life test plans of electrical insulation for thermal stress," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 1, pp. 488–494, 2015.

[16] A. D. Ashkezari, H. Ma, C. Ekanayake, and T. K. Saha, "Multivariate analysis for correlations among different transformer oil parameters to determine transformer health index," in *Proceedings of the IEEE Power and Energy Society General Meeting (PES '12)*, pp. 1–7, IEEE, San Diego, Calif, USA, July 2012.

[17] "IEEE guide for acceptance and maintenance of insulating oil in equipment," IEEE Standard C57. 106-2006, 2006.

[18] X. Zhang, B. Hu, X. Ma, and L. Xu, "Resting-state whole-brain functional connectivity networks for MCI classification using L2-regularized logistic regression," *IEEE Transactions on Nanobioscience*, vol. 14, no. 2, pp. 237–247, 2015.

[19] H. Zhang, X. Han, and S. Dai, "Fire occurrence probability mapping of northeast china with binary logistic regression model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 1, pp. 121–127, 2013.

[20] R. Xi, N. Lin, and Y. Chen, "Compression and aggregation for logistic regression analysis in data cubes," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 4, pp. 479–492, 2009.

[21] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, NY, USA, 2001.

[22] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, China Machine Press, Beijing, China, 7th edition, 2011.

[23] K. Backhaus, B. Erichson, W. Plinke, X. Wang, and R. Weiber, *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*, Springer, Heidelberg, Germany, 2005.

[24] P. Dou, D. Yang, L. Jinglu, H. Wang, and M. Li, "Research on relationship between acid value and dielectric dissipation factor of transformer oil," *Transformer*, vol. 45, no. 7, pp. 47–50, 2008.

[25] L. En-Wen and S. Bin, "Transformer health status evaluation model based on multi-feature factors," in *Proceedings of the International Conference on Power System Technology (POWERCON '14)*, pp. 1417–1422, Chengdu, China, October 2014.

[26] "Guidelines of condition assessment for electric distribution network equipments," China State Grid Corporation Standard Q/GDW 645-2011, China Electric Power Press, Beijing, China, 2011.