Early Warning of Incipient Faults for Power Transformer Based on DGA Using a Two-Stage Feature Extraction Technique

Article *in* IEEE Transactions on Power Delivery · August 2021

DOI: 10.1109/TPWRD.2021.3103455

Yang Zhang
The Hong Kong Polytechnic University

Hongcai Chen
Southeast University (China)

Ya Ping Du
The Hong Kong Polytechnic University

Qingsha Cheng
Southern University of Science and Technology

Early Warning of Incipient Faults for Power Transformer Based on DGA Using a Two-Stage Feature Extraction Technique

Yang Zhang, Hong Cai Chen, *Member, IEEE*, Yaping Du, Min Chen, Jie Liang, Jianhong Li, Xiqing Fan, Ling Sun, Qingsha S. Cheng and Xin Yao, *Fellow, IEEE*

***Abstract*—Early warning for transformer faults is valuable for maintenance decision-making. However, limited work has been done in this area due to the difficulty of the model establishment. This paper proposes a two-stage feature extraction method for early warning of power transformer faults, where a novel feature extraction process is applied by combining feature ranking and genetic programming (GP). In the first stage, the data is labeled as normal and fault states and the feature extraction is evaluated on the data. Then, extracted key features and their growth rates are relabeled as normal and warning states, after which the feature extraction process is re-evaluated on the relabeled data. Obtained features and logic expression can finally be used for early warning. The proposed framework can implement an early warning with about 100 days in advance for transformer faults and is verified through 8 sequences of data. The comparisons with two recently reported methods show the superiority of the proposed method.**

***Index Terms*—dissolved gas analysis, transformer diagnosis, feature extraction, early warning, genetic programming.**

I. Introduction

TRANSFORMER is the key equipment for the reliability of power system. Transformer diagnosis and monitoring play an important role in transformer maintenance and have been a popular topic for a long period. Dissolved gas analysis (DGA) [1] is the main method for diagnosing the transformer faults based on the gases generated due to the dissolver of the transformer insulating oil.

Tradition techniques [2], such as Rogers ratios [3], Dornenberg ratios [4], IEC ratio [2], Duval triangle [5] and IEEE key gases [6], have been widely used in transformer fault diagnosis, which are largely based on personal experience and thus lead to unsatisfactory diagnosis accuracy sometime. To overcome such a limitation, artificial intelligence (AI) algorithms [7, 8] are developed to obtain more accurate diagnostic thresholds. Instances contain evidential reasoning [9], fuzzy logic [10, 11], self-organizing map [12], neural network [13, 14], wavelet network [15], belief network [16], fault info integration [17], naive classifier [18], fusion of classifiers [19] an so forth. However, these algorithms restrict the application on detecting the fault types rather than a warning in advance.

Recently, early warning [20] [21] [22] has been proposed for transformer status detection. Despite diagnosing between health and fault of the transformer, early warning focus on identifying the sub-health status of power transformers. The sub-health status is a state that fault did not occur but the condition of the transformer began to deteriorate rapidly. Identifying the sub-health status of power transformer can provide early warning for maintenance and decision-making before the occurrence of deficiency. Nevertheless, work related to the early warning of power transformer using DGA is limited. Hidden Markov Model [23] is used to convert transformer status to dynamic state changes, and to better understand the trends of dissolved gases in time series. The warning of sub-health status is implemented and dynamic fault prediction based on DGA data is achieved. A differentiated warning rule [24] for warning transformer health status is developed, investigating the distribution characteristics of gas concentration and gas increase rate, and finding their thresholds for early warning.

These attempts have two disadvantages. Firstly, the operation information of the transformer is not fully utilized. For instance, in [23], only five gases are used as the diagnosing criterion. In [24], CO and CO_2 are included and the gas increase rates are also considered. However, according to IEC [2] and IEEE [6], gas concentrations, gas ratios, combinations of the relative percentage of gases can be used as significant features for transformer diagnosis. Secondly, both two methods assume that the transformer deteriorates with the growth of gas

This work is supported by Science and Technology Project of State Grid Corporation of China (No. 5500-202019090A-0-0-00). The corresponding author is Hong Cai Chen.

Yang Zhang is with College of Advanced Interdisciplinary Studies, National University of Defense Technology, Changsha, China. (email: yangzhangsmile@outlook.com)

Hong Cai Chen is with School of Automation, Southeast University, Nanjing 210096, China. (email: hc.chen@live.com)

Yaping Du is with Department of Building Services Engineering, Hong Kong Polytechnic University, Hong Kong SAR. (email: ya-ping.du@polyu.edu.hk)

Min Chen is with Research Institute, State Grid Zhejiang Electric Power Co., Ltd., Hangzhou 310014, China. (email: 374550194@qq.com)

Jie Liang and Jianhong Li are Zhejiang Huayun Information Technology Co., Ltd., Hangzhou 310012, China. (email: 77398246@qq.com, 21902126@qq.com)

Xiqing Fan is with State Grid Zhejiang Electric Power Co., Ltd., Lishui Power Supply Bureau, Lishui, China. (email: fanxiqing2010@163.com)

Ling Sun is with State Grid Zhejiang Electric Power Co., Ltd., Shaoxing Power Supply Bureau, Shaoxing, China. (email: 165978515@qq.com)

Qingsha S. Cheng is with Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen 518055, China. (email: chengqs@sustech.edu.cn)

Xin Yao is with Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China. (email: xiny@sustech.edu.cn)

concentrations but in the practical situation, the concentrations of some gases may not increase even when the fault occur. Therefore, reported methods may lead to false warnings.

In order to solve the limitations of the previous DGA based early warning methods, a reliable power transformer incipient fault early warning method is proposed. The innovation and contribution of the paper are summarized as followed. Firstly, the features for power transformer sub-health state identification are investigated for the first time. The unrelated features were excluded and only the related features are extracted for early warning which increases the accuracy and efficiency of the algorithm. Secondly, the problem of data imbalance [25] which is rarely investigated but can largely reduce the diagnosis accuracy is tackled by using the two-stage method. The data of warning state is much smaller than that of health state (around 1:100). We reduce the dimension of features by labeling the DGA data to normal and fault states (with a relatively low imbalance ratio around 5:1) in stage one. Then, early warning indicators can be effectively obtained in stage two. Thirdly, the performances of the reported early warning methods are evaluated and their failure mechanisms are analyzed. The rule for early warning is interpretative and can be used or tested by others directly. Finally, Though the age, capacity or location of transformers are varied, the decomposed gases of the transformer with the same fault are similar [2] [6]. Extracted features in our work has a certain level of generality. This is also the reason why we use a logic synthesis despite of AI algorithms to construct the diagnosis rule.

The rest of the paper is organized as follows. Section II presents the whole framework of the proposed early warning, along with the three key processes, namely feature ranking, early warning construction and data relabeling. Each involved technique is verified in Section III and their advantages are also validated by comparing with other methods in Section VI. Finally, the conclusion is drawn in Section V.

II. The Framework of Dynamic Early Warning

To implement an efficient early warning, a two-stage approach integrating feature ranking and logic synthesis is proposed as shown in Fig. 1. Measured dissolved gases are firstly mixed up and generate features containing various relations among them, followed by feature extraction aiming to find decisive features. The framework is divided into two stages, which adopt the same feature extraction process except for input features and labels. In the feature extraction, nonparametric test is firstly adopted for feature ranking and prescreening features, after which Genetic Programming (GP) is used to remove redundant features through logic synthesis. In the second stage, time-sequence data of these features and their growth rates are relabeled based on normal and warning states, and then fed into nonparametric test and GP again to find the location of sub-health status, so as to realize the early warning of transformer operation status. It is noted that the imbalance ratio of normal to warning is quite large, yet the proposed two-stage approach can significantly reduce the difficulty to deal with such an imbalanced ratio. The detailed procedures and techniques of feature extraction are explained in this section.

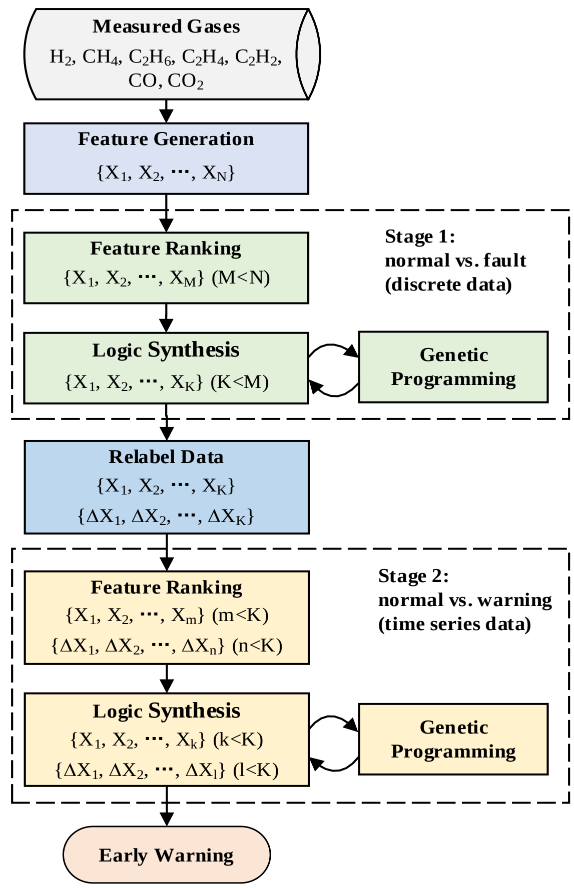


Fig. 1. The framework of the proposed two stage approach for early warning of incipient transformer faults.

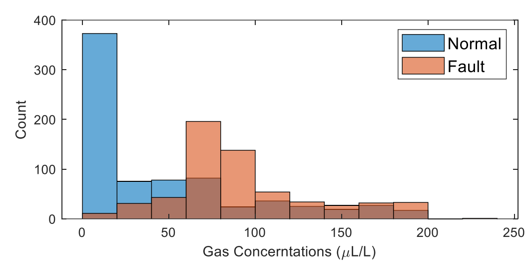


Fig. 2. The frequency distribution of CH_4 from normal and fault transformers.

A. *Feature generation*

In this work, eight gas concentrations are considered including Hydrogen (H_2), Methane (CH_4), Ethane (C_2H_6), Ethylene (C_2H_4), Acetylene (C_2H_2), Carbon Monoxide (CO), Carbon Dioxide (CO_2), and Total Hydrocarbon (TCH). Based on these gases, we generate 54 features including

- individual gas: H_2, CH_4, etc.
- the ratio of two gases: CH_4/ H_2, C_2H_4/ C_2H_6, etc.
- relative content of multiple gases: $H_2/(H_2+C_2H_4+C_2H_2)$, $C_2H_4/(CH_4+C_2H_2+C_2H_4)$, etc.

B. *Feature ranking using Wilcoxon rank sum test*

The histograms of the frequency distribution of CH_4 and C_2H_4 are shown in Fig. 2. It can be observed that the distribution of gases does not satisfy the normal distribution, and it is difficult to provide a distribution hypothesis. In order to avoid a biased estimation, obtaining the probability density estimation of parameters based on the idea of data-driven rather than parameter estimation is preferred. Therefore, nonparametric methods should be applied to analyze the data. We use Wilcoxon rank sum test [26] (a popular type of nonparametric test) to compare two independent samples, which can tell if they are drawn from populations having identical distribution based on ranks instead of the original observations.

The 'rank' used here is that defined in statistics despite the rank of a matrix. Suppose a group of samples $X = [x_1, x_2, \ldots, x_n]$. Rearrange these samples in ascending order as $X' = [x_{(1)}, x_{(2)}, \ldots, x_{(n)}]$. The rank of x_i is defined as R_i and is obtained as the index of the new samples when $x_i = x_{(R_i)}$.

The basic idea of Wilcoxon rank sum test is to combine two groups of samples and compare the probability density distribution of the sum of the ranks. Then the null hypothesis H_0 can be rejected or accepted according to significance. Suppose we have two groups of samples, denoted by X_1 and X_2. Combine two groups of samples into a single sequence, and then calculate the rank sum W_1 and W_2 for the two groups of samples respectively based on the ranking level (ascending order) in the combined sequence. The rank sum W_i is obtained by

$$W_i = \sum_{j=1}^{n_i} R_j \tag{1}$$

where n_i is the number of observations in samples X_i, and R_j is the rank for jth sample in X_i after pooling into a single sequence.

The rank sum statistic can be expressed using U-test statistic, where U counts the number of samples in X_2 precedes that in X_1 in an ordered arrangement of the elements and given by

$$U_i = W_i - \frac{n_i\left(n_i+1\right)}{2} \qquad i = 1,\ 2 \tag{2}$$

The smaller value of U_1 and U_2 is to consult significance. As the distribution of rank sum is close to the normal distribution when the sample size above 20, z-statistic is used to compute the approximate p-value of the test, which is defined as

$$z = \frac{U - \mu_U}{\sqrt{\sigma_U}} \tag{3}$$

where

$$\mu_U = \frac{n_1 n_2}{2} \qquad \sigma_U = \sqrt{\frac{n_1 n_2\left(n_1+n_2+1\right)}{12}}$$

in which μ_U and σ_U are the mean and standard deviation of U.

The probability of observing a test statistic (p-value of the test) is then can be checked once z is obtained based on the normal distribution as

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(t-\mu)^2}{2\sigma^2}} \quad \text{where} \ \ x = -|z| \tag{4}$$

The lower p indicates a better separation of the data for diagnosis. Therefore, we can determine the importance of the feature based on p-value and extract features with $p < 0.05$ (>95% confidence).

C. *Logic synthesis using Genetic Programming*

Obtained features from the prior step have superfluous terms and their association relations are undetermined. Logic synthesis with GP [27, 28] is successively applied to refine the obtained features.

To import to logic synthesis, features extracted using feature ranking should be firstly transformed into logic 0 or 1 representing normal and fault conditions as expressed:

$$x = \begin{cases} 1 & x \ge x_T \\ 0 & x < x_T \end{cases} \tag{5}$$

The threshold x_T to separate normal and fault transformers should be determined. The principle of selecting the threshold is to maximize the discrimination between normal and fault transformers. As shown in Fig. 3, the threshold is determined at which the difference between cumulative probabilities of normal and fault transformers is the largest [29]. As the samples of dissolved gases are discrete, empirical cumulative distribution is adopted, where the probability distribution is the count of occurrence of each value. The distribution is a stair function so that interpolation is employed to obtain a smooth distribution curve. Assume the cumulative distribution function as F_1 and F_2 for two cases respectively, the threshold can be obtained by

$$x_T = \arg\max_x \left\{\left|F_1(x) - F_2(x)\right|\right\} \tag{6}$$

Afterward, GP is evaluated. GP is a branch of evolutionary computation, where the individual in genetic algorithm is replaced by parse trees. GP is initially developed for symbolic regression, and generates several parse trees with two types of nodes, namely symbols and terminators. Many advanced GPs [30-33] have been proposed to deal with imbalanced feature extraction. Our method adopts a two-stage approach that the basic GP can provide satisfactory results. In this study, symbols are limited to logical operators (and &, or |) for discovering logical relationships. Features are terminators that end a branch on the parse tree.

GP evolves in a genetic style and makes the next generation as the diagram shown in Fig. 4(a). This consists of computing the "fitness" of every candidate solution in the current population and generating a new population using roulette wheel selection with elitism. In roulette wheel selection, solutions with higher fitness scores have more probability to be selected. Elitism means an identical copy of the solution with the highest fitness score. After scoring, a new generation is created using the genetic operators: replication, crossover, and mutation, and so on until convergence. This process promotes genetic diversity and removes redundancy from the population.

The fitness function for this study was chosen to be Sensitivity/Specificity (*SS*) fitness function [27]. Sensitivity reflects the probability of positives that are correctly identified, while specificity measures that of negatives with precise identification. This fitness function is designed for logic synthesis and efficient in situations where unbalanced training sets are being used. The Sensitivity/Specificity fitness is evaluated by the equation:

$$SS_i = k \cdot SE_i \cdot SP_i \tag{7}$$

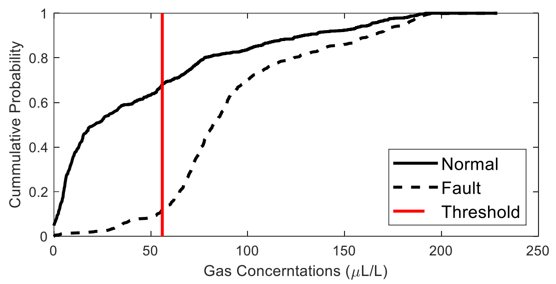


Fig. 3. The cumulative distribution of CH_4 for normal and fault transformers.

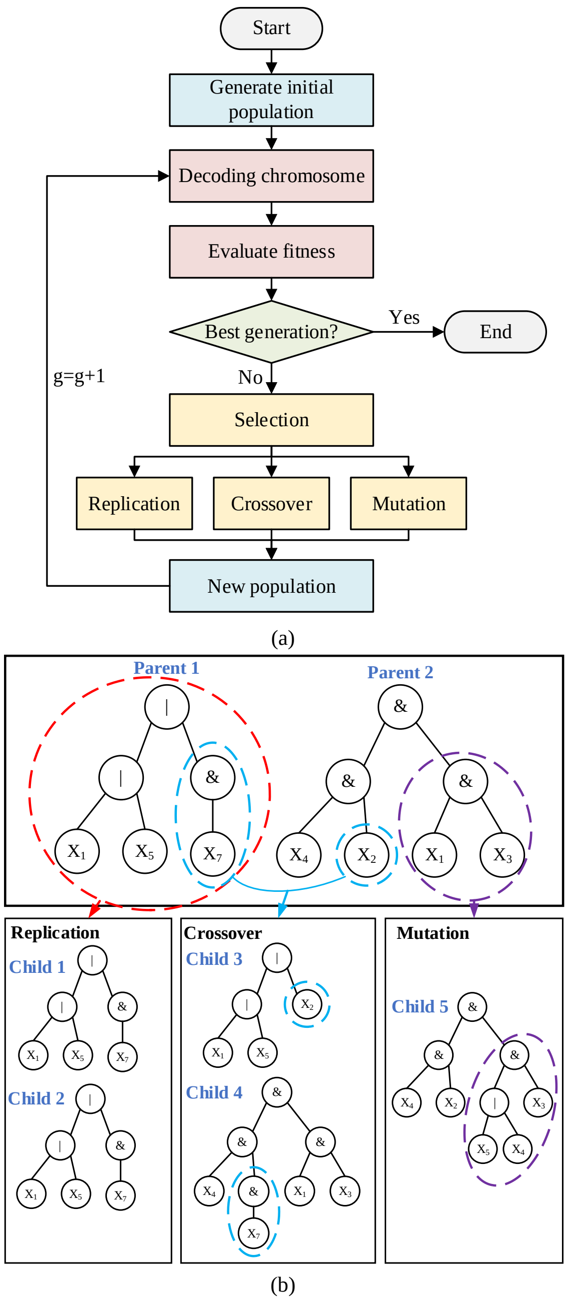


Fig. 4. The schematic diagram of GP. (a) The flowchart of the algorithm. (b) Genetic operations in GP.

where k = 1000 is the scaling factor, SE_i is the sensitivity and SP_i is the specificity of the individual i, and are given by the formulas:

$$SE_i = \frac{TP_i}{TP_i + FN_i} \quad \text{and} \quad SP_i = \frac{TN_i}{TN_i + FP_i}$$

where TP_i, TN_i, FP_i, and FN_i represent the number of true positives, true negatives, false positives, and false negatives, respectively. These four indices can be obtained by computing the confusion matrix. A confusion matrix [34] consists of information about actual and predicted classification returned by a classifier. Assuming the minority class to be positive class (P) and the majority class to be negative class (N), the confusion matrix is defined as Table I.

TABLE I THE DEFINITION OF THE CONFUSION MATRIX

	Positive prediction	Negative prediction
Positive class	True Positive (*TP*)	False Negative (*FN*)
Negative class	False Positive (*FP*)	True Negative (*TN*)

In our work, the configuration of the GP is summarized in Table II. Three types of genetic operators (as shown in Fig. 4(b)) are employed and can be briefly described as

- *Replication*: replicate the individuals to increase the number of the population, and improve the chance of the optimal individuals to be selected.
- *Crossover*: two individual trees randomly select their nodes to exchange branches. Then, two new individuals are generated.
- *Mutation*: Arbitrarily select a node in the tree, and then replace the mutation part of the original individual.

TABLE II THE CONFIGURATION OF GP IN THIS WORK

Operator/Parameter	**Strategy/Value**
Genetic operators	replication, crossover, mutation
Replication	roulette wheel selection
Crossover rate	0.3
Mutation rate	0.2
Population	100
Maximum number of iterations	1000
Symbols	logic operators & and |
Fitness function	Sensitivity/Specificity

D. Relabeling data for stage two

After the feature extraction in the first stage, early warning is developed based on these obtained features. As concluded in many articles [23, 24, 35], the transformer state is a cumulative change. Both gas features and their growth rates are the main indexes for monitoring statues. Either a higher gas concentration or a fast growth rate can reflect a sub-health or unhealthy status of the transformer. Therefore, this work utilizes moving growth rates, and gas features as the criteria to implement early warning.

As shown in Fig. 5, by calculating the growth rate over a window of a fixed length and sliding the window, we obtain a new time series consisting of moving grow rates. We construct the warning framework based on offline gases. The growth rate is obtained by absolute gas feature variety rate as defined as

$$\Delta x_i = \frac{1}{m} \frac{x_i - x_{i-m}}{x_{i-m}} \tag{8}$$

where i is the time index, m is the length of the sliding window, x_i is the gas concentration at time i, and Δx_i is the growth rate in the window.

Before processing, data of various features (taking a four-feature case X_1-X_4 as an example) and their growth rates (ΔX_1-ΔX_4) in time sequences are firstly labeled as shown in Fig. 6. The labeling method is essential for the construction of early warning. This work aims to find a moment that can warn the fault in advance that only a short period is labeled, or called the drift [36] of fault state in the time sequence. The data is labeled as 1 at the period of warning moments, where waveforms show abnormal or sudden changes nearest to the fault as "Warning" marked in Fig. 6. The data after the fault are abandoned. Other parts represent the normal state which are labeled 0. The relabeling can be expressed by

$$state = \begin{cases} normal & = 0 & \boldsymbol{x} \le \boldsymbol{C}_N \\ warning & = 1 & \boldsymbol{C}_N < \boldsymbol{t} < \boldsymbol{C}_W \\ fault & NA & \boldsymbol{x} \ge \boldsymbol{C}_F \end{cases} \tag{9}$$

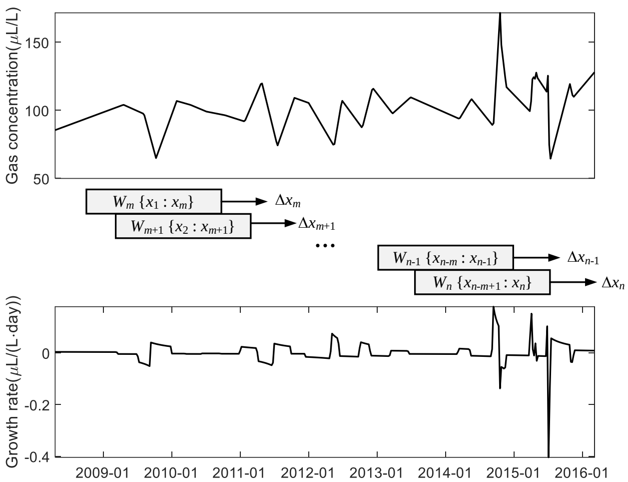


Fig. 5. The schematic diagram of the sliding window on the gas concentration of CO and related growth rates. W_{m+i} represents the ith time section with m time points namely {x_i, x_{i+1}, … x_{m+i}}.

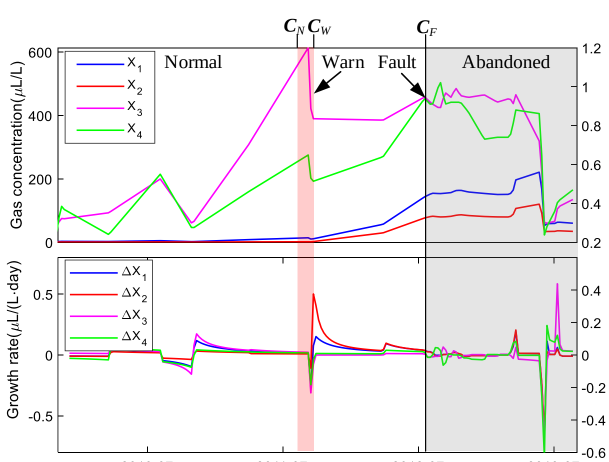


Fig. 6. The diagram of the gas features and growth rates and their labels in an example with four features.

where $\boldsymbol{C}_N$, $\boldsymbol{C}_W$ and $\boldsymbol{C}_F$ are time indicators of features for separating different conditions as marked in Fig. 6.

After labeling, the data are imported to feature ranking and then GP. Finally, the most related features for early warning are screened out and the diagnosis criterion is obtained. It is noted that the sample of warning state is very few as shown in Fig. 6, resulting in highly imbalanced data, and thus the proposed two-stage method is necessary.

III. Experimental Results of Practical Transformers

The interpretation of the DGA results includes thermal fault, partial discharge, and discharge. The discharge fault is usually caused by sudden faults such as nearby short-circuit which is hard to predict. The partial discharge, if not properly handled, might also result in low energy or high energy discharge in power transformer. Since this paper focuses on early warning of incipient faults, we limit our discussion on overheating and partial discharge faults. The discharge fault is beyond the range of the incipient faults and will not be discussed in this paper.

To show the effectiveness of the proposed framework, a dataset provided by Zhejiang power grid is studied. There are 8 sequences of data that evolve from normal to fault. To fully utilize the data and verify the robustness of the proposed model, leave-one-out cross validation is used to divide the data, where 1 sequence of data is used as the testing set and other k-1 sequences are used as the training set (k is the total number of data). The warning state extraction process of overheating fault is presented step-by-step to show the details of the method. The process of early warning for partial discharge which is similar to that of overheating is introduced briefly for comparison.

A. Early warning for overheating fault

In Stage One, the generated 54 features are firstly processed by Wilcoxon rank sum test to determine the confidence level of each feature. There are 37 features having a low p-value meeting the confidence level of 95%. Since higher confidence indicates higher discrimination, these features are rearranged in the ascending order of confidence level. As traditional interpretation methods can report the transformer fault using less than 5 features, we select the first 15 features (as X_1-X_{15} shown in Table III) as the input for the following logic synthesis. The selection of 15 features is about 3 times that used in traditional interpretation methods which can cover effective features. Taking advantage of cumulative distribution functions, the thresholds for these features are determined by maximizing the discrimination between the normal and fault transformers as shown in the same table.

These features are then used for logic synthesis. According to the thresholds, each feature of each data is converted to logical values 0 for normal and 1 for fault conditions. Obtained 0s and 1s of X_1-X_{15} are used as input, and GP is used to synthesize their relations. The ground truth value to build fitness is derived from labeled data introduced in Sec. III.

Since GP is a stochastic algorithm, 10 independent runs are evaluated and their convergence curves are shown in Fig. 7. It can be seen that GP converges rapidly to a high fitness score in less than 100 iterations. All these 10 runs give the same fitness score of 839.2, and result in the same solution after expression simplification. They all simplified to Y = (X_2 & X_5) | (X_2 & X_4 & X_{15}). These four features are finally selected and used for the feature extraction in Stage Two.

In Stage Two, the growth rates of extracted four features are firstly obtained. Extracted features and their growth rates are reprocessed to implement early warning. Six sequences of data that have a whole transformer life from normal, to warning and end with fault are tested.

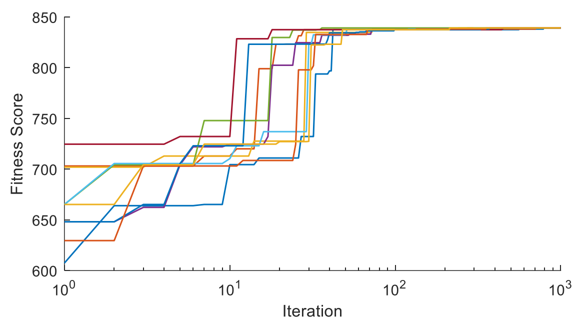


Fig. 7. The convergence curves of GP for 10 runs. (in logarithmic coordinates to show the convergence process)

TABLE III OBTAINED 15 SELECTED FEATURES AND THEIR THRESHOLDS

ID	Item	N	F	*p*-value
X_1	CH_4	≤55.93	>55.93	8e-77
X_2	THC	≤126.98	>126.98	6e-66
X_3	C_2H_6	≤17.13	>17.13	1e-61
X_4	C_2H_4	≤44.44	>44.44	6e-56
X_5	CO	≤1340	>1340	3e-54
X_6	CH_4/H_2	≤0.78	>0.78	2e-44
X_7	$H_2/(H_2+CH_4+C_2H_2)$	≥0.56	<0.56	2e-43
X_8	$CH_4/(H_2+CH_4+C_2H_4)$	≤0.44	>0.44	5e-43
X_9	C_2H_2	≤0	>0	8e-43
X_{10}	$H_2/(H_2+CH_4+C_2H_6)$	≥0.49	<0.49	3e-42
X_{11}	$H_2/(H_2+CH_4+C_2H_4)$	≥0.38	<0.38	2e-41
X_{12}	$CH_4/(H_2+CH_4+C_2H_6)$	≤0.39	>0.39	3e-41
X_{13}	$H_2/(H_2+CH_4+C_2H_2+C_2H_4+C_2H_6)$	≥0.35	<0.35	4e-41
X_{14}	C_2H_6/H_2	≤0.23	>0.23	1e-39
X_{15}	C_2H_4/CH_4	≤0	>0	2e-39

TABLE IV OBTAINED 4 FEATURES AND THEIR GROWTH RATES

ID	Item	N	F	*p*-value
X_2	THC	≤63.1	>63.1	2e-13
X_4	C_2H_4	≤28.8	>28.8	9e-14
X_5	CO	≤287.7	>287.7	0.33
X_{15}	C_2H_2/H_2	≤0	>0	1e-10
ΔX_2	THC growth rate	≤0.038	>0.038	2e-18
ΔX_4	C_2H_4 growth rate	≤0.034	>0.034	4e-18
ΔX_5	CO growth rate	≤0.001	>0.001	0.89
ΔX_{15}	C_2H_2/H_2 growth rate	≤0.02	>0.02	2e-9

The data is firstly processed using Wilcoxon rank sum test to screen out indecisive features, and then thresholds (Table IV) are determined according to cumulative distribution. It is noted that the thresholds determined in this part aim to discriminate normal and warning status, differing from that to separate

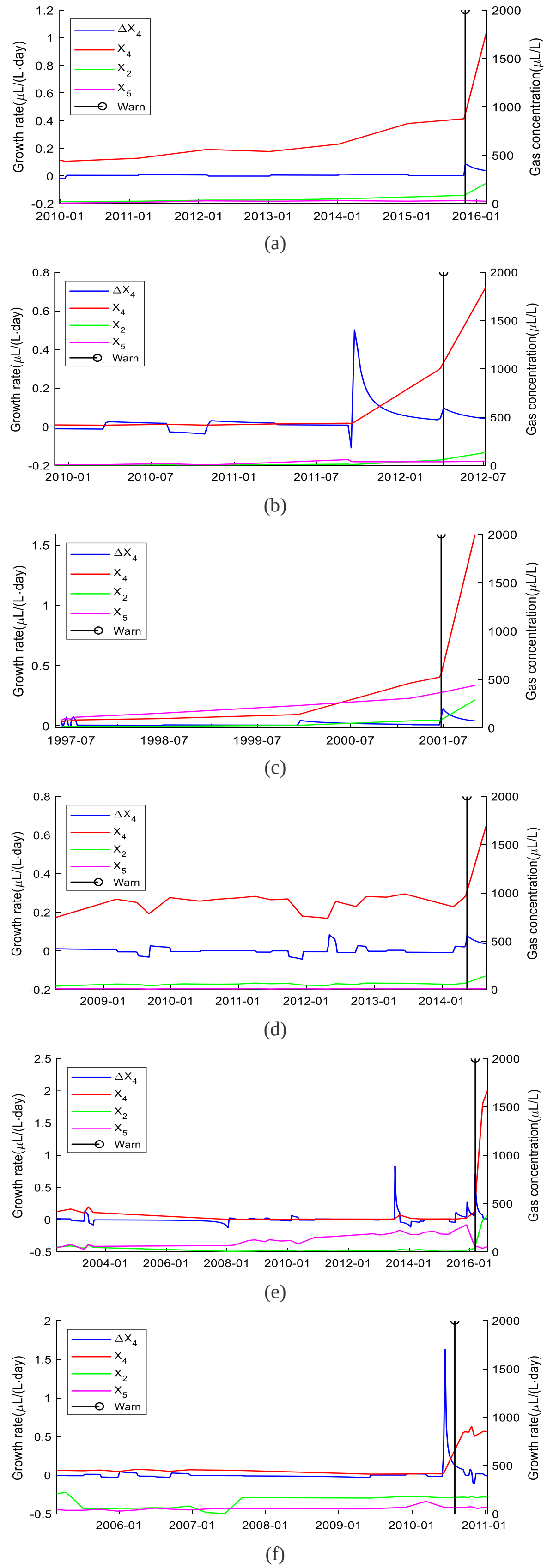


Fig. 8. The changes of extracted features for 6 transformers in time sequences. (a) - (f) indicates 6 transformers. The vertical black line indicates the time for the first warning. The fault occurs at the end of time which is not marked. (Left axis - ΔX_4 and X_4, right axis - X_2 and X_5)

TABLE V WARNING TIME BEFORE FAULT

ID	1	2	3	4	5	6
Warning Time (Day)	112	98	133	105	147	91

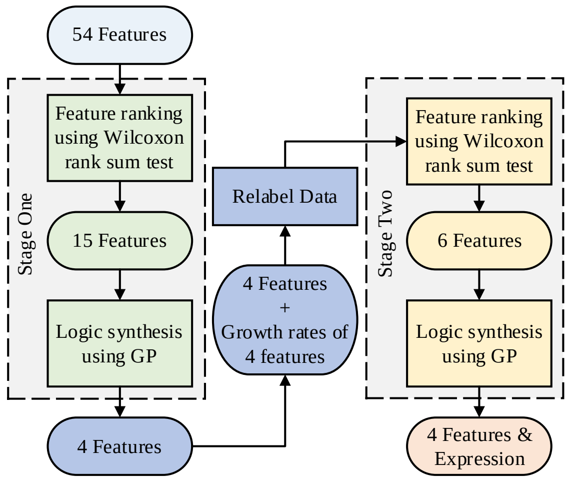


Fig. 9. The feature extraction process for tested overheating fault.

normal and fault status in Stage one. According to the *p*-value, X_5 and ΔX_5 are removed. Six features (as listed in Table IV) are converted into logical values and input to logic synthesis. 10 independent runs of GP are evaluated and all these runs result in the same fitness of 993.5. Finally, the result is obtained after expression simplification as

$$Y = X_2 \,\&\, X_4 \,\&\, X_5 \,\&\, \Delta X_4 \qquad (10)$$

The criterion (10) is used to find an early warning in a time sequence of transformer gas. The data are linearly interpolated with the same time interval (7 days). Six sequences of data are tested using the extracted criterion.

Fig. 8 displays the status changes and diagnosis results for all 6 transformers. It can be observed that early warning is implemented for all 6 transformers that can prove the efficiency of the proposed framework. The greatest advantage of the proposed framework is that it can warn the fault just before it occurs. No false warn occurs for all investigated cases which is significant for real-world maintenance. The warning times for 6 transformers are summarized in Table V. Early warning is implemented for all six sequences with an average of 114 days. This warning time is appropriate in practical maintenance.

Fig. 8 also reveals that the status change is not monotonous as shown in Fig. 8(b) and 8(e). Therefore, it is insufficient to use gas concentrations or growth rates directly as the criteria. If so, sudden changes in the local period before warning status may lead to false-warns. It also indicates that single feature cannot provide a reliable diagnosis of the transformer status, which means multiple features should be carefully selected.

In summary, the proposed feature extraction method contains two stages. Stage One extracts key features that can decimate between normal and fault conditions. Stage Two finds features for constructing early warning. The unimportant features are removed step-by-step as shown in Fig. 9. Feature ranking removes features with low confidence levels and logic synthesis finds key features as well as their relations.

TABLE VI OBTAINED 4 FEATURES AND THEIR GROWTH RATES

ID	Item	N	F
X_1	C_2H_2/C_2H_4	≥12.92	>12.92
X_5	$C_2H_2/(H_2+CH_4+C_2H_2)$	≤0	>0
X_{12}	C_2H_2/C_2H_6	≤0	>0
X_{14}	C_2H_4/C_2H_6	≥4.24	<4.24
ΔX_1	C_2H_2/C_2H_4 growth rate	≤0.0143	>0.0143
ΔX_5	$C_2H_2/(H_2+CH_4+C_2H_2)$ growth rate	≤0.0155	>0.0155
ΔX_{12}	C_2H_2/C_2H_6 growth rate	≤0.0239	>0.0239
ΔX_{14}	C_2H_4/C_2H_6 growth rate	≤0.0131	>0.0131

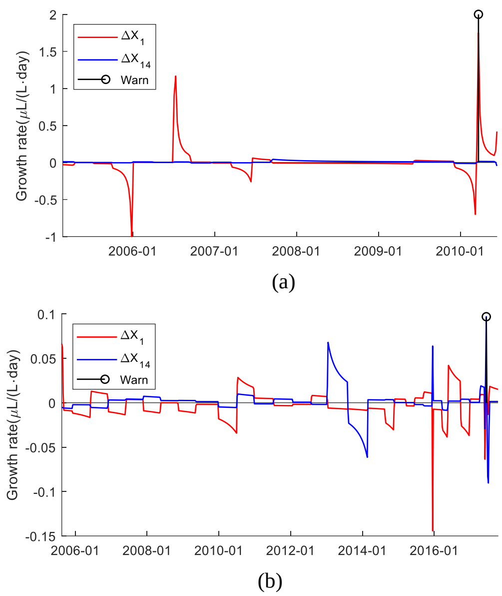


Fig. 10. The changes of extracted features for 2 transformers in time sequences. (a) - (b) indicates 2 transformers. The vertical black line indicates the time for the first warning. The fault occurs at the end of time which is not marked.

B. *Early warning for partial discharge fault*

The same process is evaluated for partial discharge fault. Firstly, 15 features are selected using Wilcoxon rank sum test which are then imported to GP and 10 independent runs are evaluated. Four features are obtained for all runs and their growth rates are calculated as listed in Table VI. This process is repeated on the relabeled data with normal and warning states. The early warning criterion is finally obtained as

$$Y = \Delta X_1 \,\&\, \Delta X_{14} \qquad (11)$$

The criterion (11) is tested on 2 sequences of data which result in partial discharge fault as shown in Fig. 10. The criterion successfully alarms the fault in advance which verifies the effectiveness of the proposed method. The average warning time is 89 days before the fault.

C. *Discussion on the advantage of two-stage approach*

To show the advantage of the proposed two-stage strategy over direct search (one-stage), feature extraction directly using GP is performed on six sequences of overheating faults. The data is labeled using normal and warning states. Total 108 features (54 features and their growth rates) are generated as the input of GP for logic synthesis. The configuration of the GP is the same as in Sec. II with the maximum number of iterations of 1000, the population as 100, the mutation rate as 0.2, and the crossover rate as 0.3. GP runs 10 times. The results are listed in Table VII, ‘Fail’ indicates the wrong warning and ‘Miss’ means no warning is reported. It shows that the diagnosis obtained by direct search shows nearly half fails for early warning. This is because the searching space of direct search is too large and the imbalance ratio of data is large. It confirms the advantage of our proposed two-stage feature extraction.

TABLE VII THE DIAGNOSIS RESULTS USING DIRECT SEARCH

Fault	Method	No.	Success	Fail	Miss
Over-heating	**Proposed**		**6**	0	0
	Direct Search (one-stage)	1	2	0	4
		2	3	0	3
		3	3	0	3
		4	4	0	2
		5	3	0	3
		6	3	0	3
		7	2	0	4
		8	3	0	3
		9	3	0	3
		10	3	0	3

IV. COMPARISON WITH OTHER METHODS

This section presents the comparison of the proposed method with two recently reported ones.

A. *Description of reported methods*

In [24], a differentiated warning rule (DWR) method is proposed for early warning of transformer faults. This method fits the distribution of data using Weibull function and separates states using thresholds at 99% confidence of the distribution. The warning is implemented by combining 8 gas concentrations and their growth rates. This method is compared in this part to demonstrate the superiority of the proposed method. In the original implementation [24], the determination of states requires to satisfy thresholds of all 8 gases and their growth rates, which is an extremely harsh criterion. In our implementation, we only use 5 gases and 90% confidence as the threshold.

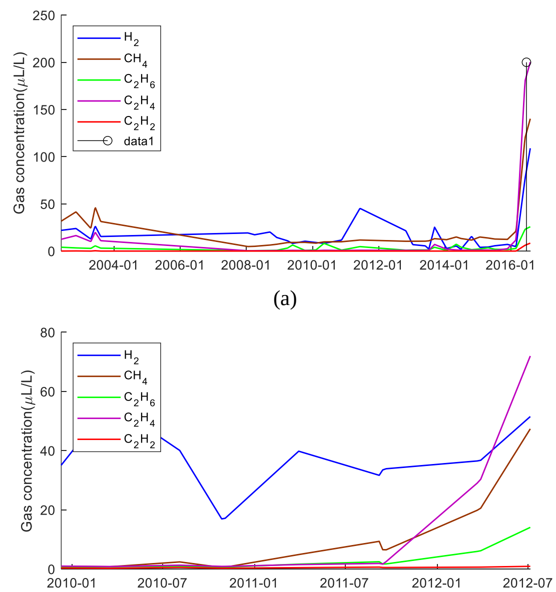
In [23], a dynamic fault prediction (DFP) method is proposed, where Gaussian mixture model (GMM) is firstly adopted to cluster health status. Then, the trained GMM and gas concentrations are imported to hidden Markov model (HMM) as the input to predict health status of the power transformer. In this approach, two GMMs are constructed, where the first one clusters the data into health, sub-health and fault states, and the other is trained only by health data resulting in one status. By combining the likelihood of these two GMMs, the status of the transformer is determined. This approach is evaluated using five gases and compared with the proposed method.

B. *Comparison result*

The diagnosis results of overheating and partial discharge faults are summarized in Table VIII. As presented in Section III, the proposed method can success warn the fault in advance for all cases. DWR can warn the fault only in one case for

TABLE VIII THE COMPARISON OF DIAGNOSIS RESULTS FOR THE PROPOSED METHOD, DWR AND DFP

Fault	Method	No.	Success	Fail	Miss
Over-heating	**Proposed**		**6**	0	0
	DWR		1	0	5
	DFP	1	4	2	0
		2	4	2	0
		3	4	1	1
		4	5	1	0
		5	5	1	0
Partial Discharge	**Proposed**		**2**	0	0
	DWR		0	0	2
	DFP	1	0	2	0
		2	0	2	0
		3	1	1	0
		4	0	2	0
		5	1	1	0



(a)

(b)

Fig. 11. The changes of 5 gases in time sequences. (a) The successful diagnosis corresponds to Fig. 8(e). (b) The miss- diagnosis corresponds to Fig. 8(b).

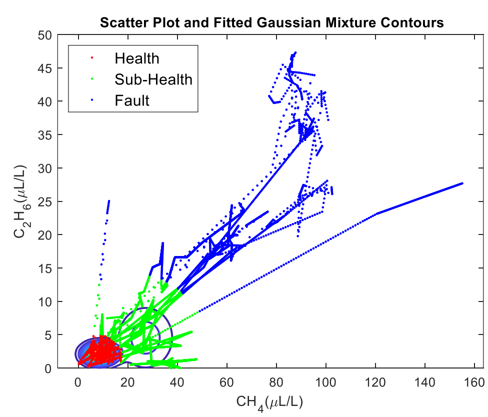


Fig. 12. The distribution of CH_4 and C_2H_6 for a sequence ending with overheating fault. The health index is obtained by GMM.

overheating fault and none for partial discharge fault. As the figures shown in Fig. 11, DWR has the capability of providing an efficient early warning only when the gas concentrations are nearly monotonous as well as having a sharp increase. The method frequently misses the diagnosis due to the extremely strict criterion. This indicates that we should not include all features in the diagnosis and implies the importance of feature extraction. The results also confirm the superiority of the proposed method.

Due to the DGA data has no clear distribution centers (as the data distribution shown in Fig. 12), DFP may result in distinct clusters in different runs. Therefore, five runs are evaluated for DFP and results for overheating and partial discharge cases are summarized in Table VIII. From the obtained results, it can be found that the method in [23] provides different results for each run. Meanwhile, the success rate of this method is better than DWR but also performs not as good as the proposed method.

The inefficiency of the reported methods can be explained by that both reported methods assume the increase of gas concentrations indicating the transformer deterioration. However, some gases do not follow this rule so that the diagnostic accuracy of these two methods is lower than that of the proposed method. Meanwhile, DWR uses a stringent criterion for the diagnosis, while that adopted by the dynamic fault prediction is soft. This is the reason why dynamic fault prediction has better performance.

V. Conclusion

This paper proposes a novel two-stage approach for transformer early warning. This method implements the feature extraction for imbalanced transformer data by using data of different imbalance ratios in two stages. In each stage, Wilcoxon rank sum test is firstly applied to select important features, and then GP is implemented for their logic synthesis. Finally, a reliable early warning for transformer faults is constructed, alerting about 114 days in advance for overheating fault and 89 days for partial discharge fault.

The proposed method successfully implements the early warning for practical transformer DGA analysis in State Grid Zhejiang Electric Power Co., Ltd. Compared to the other reported methods, it shows a significant superiority in the effectiveness of early warning. The method is data-driven so that no experts' efforts are required and it can be easily applied to other data.

Acknowledgment

The authors would like to thank State Grid Zhejiang Electric Power Co., Ltd., Research Institute for providing the valuable DGA data.

References

[1] N. Bakar, A. Abu-Siada, and S. Islam, "A review of dissolved gas analysis measurement and interpretation techniques," *IEEE Electrical Insulation Magazine,* vol. 30, no. 3, pp. 39-49, 2014, doi: 10.1109/mei.2014.6804740.

[2] *IEC 60599: Mineral Oil-Filled Electrical Equipment In Service - Guidance On The Interpretation Of Dissolved And Free Gases Analysis*, 2015.

[3] R. Rogers, "IEEE and IEC codes to interpret incipient faults in transformers, using gas in oil analysis," *IEEE transactions on electrical insulation*, no. 5, pp. 349-354, 1978.

[4] E. Dornenburg and W. Strittmatter, "Monitoring oil-cooled transformers by gas-analysis," *Brown Boveri Review,* vol. 61, no. 5, pp. 238-247, 1974.

[5] M. Duval and A. dePablo, "Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases," *IEEE Electrical Insulation Magazine,* vol. 17, no. 2, pp. 31-41, Mar-Apr 2001, doi: Doi 10.1109/57.917529.

[6] *IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers*, 2019.

[7] P. Mirowski and Y. LeCun, "Statistical Machine Learning and Dissolved Gas Analysis: A Review," *IEEE Transactions on Power Delivery,* vol. 27, no. 4, pp. 1791-1799, 2012, doi: 10.1109/tpwrd.2012.2197868.

[8] Y. Zhang *et al.*, "Power transformer fault diagnosis considering data imbalance and data set fusion," *High Voltage,* pp. 1-12, Dec 9 2020, doi: 10.1049/hve2.12059.

[9] W. H. Tang, K. Spurgeon, Q. H. Wu, and Z. J. Richardson, "An evidential reasoning approach to transformer condition assessments," *IEEE Transactions on Power Delivery,* vol. 19, no. 4, pp. 1696-1703, 2004, doi: 10.1109/tpwrd.2003.822542.

[10] A. Abu-Siada, S. Hmood, and S. Islam, "A New Fuzzy Logic Approach for Consistent Interpretation of Dissolved Gas-in-Oil Analysis," *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 20, no. 6, pp. 2343-2349, Dec 2013, doi: Doi 10.1109/Tdei.2013.6678888.

[11] A. Abu-Siada and S. Hmood, "A new fuzzy logic approach to identify power transformer criticality using dissolved gas-in-oil analysis," *International Journal of Electrical Power & Energy Systems,* vol. 67, pp. 401-408, 2015, doi: 10.1016/j.ijepes.2014.12.017.

[12] K. F. Thang, R. K. Aggarwal, A. J. McGrail, and D. G. Esp, "Analysis of power transformer dissolved gas data using the self-organizing map," *IEEE Transactions on Power Delivery,* vol. 18, no. 4, pp. 1241-1248, 2003, doi: 10.1109/tpwrd.2003.817733.

[13] V. Miranda and A. R. G. Castro, "Improving the IEC table for transformer failure diagnosis with knowledge extraction from neural networks," *IEEE Transactions on Power Delivery,* vol. 20, no. 4, pp. 2509-2516, Oct 2005, doi: 10.1109/Tpwrd.2005.855423.

[14] X. Hao and S. Cai-xin, "Artificial Immune Network Classification Algorithm for Fault Diagnosis of Power Transformer," *IEEE Transactions on Power Delivery,* vol. 22, no. 2, pp. 930-935, 2007, doi: 10.1109/tpwrd.2007.893182.

[15] C. Weigen, P. Chong, Y. Yuxin, and L. Yilu, "Wavelet Networks in Power Transformers Diagnosis Using Dissolved Gas Analysis," *IEEE Transactions on Power Delivery,* vol. 24, no. 1, pp. 187-194, 2009, doi: 10.1109/tpwrd.2008.2002974.

[16] J. Dai, H. Song, G. Sheng, and X. Jiang, "Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network," *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 24, no. 5, pp. 2828-2835, 2017, doi: 10.1109/TDEI.2017.006727.

[17] X. L. Wang, Q. M. Li, C. R. Li, R. Yang, and Q. Su, "Reliability Assessment of the Fault Diagnosis Methodologies for Transformers and a New Diagnostic Scheme Based on Fault Info Integration," *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 20, no. 6, pp. 2292-2298, Dec 2013, doi: Doi 10.1109/Tdei.2013.6678882.

[18] B. Zhao, M. Yang, H. R. Diao, B. An, Y. C. Zhao, and Y. M. Zhang, "A novel approach to transformer fault diagnosis using IDM and naive credal classifier," *International Journal of Electrical Power & Energy Systems,* vol. 105, pp. 846-855, 2019, doi: 10.1016/j.ijepes.2018.09.029.

[19] J. I. Aizpurua, V. M. Catterson, B. G. Stewart, S. D. J. McArthur, B. Lambert, and J. G. Cross, "Uncertainty-Aware Fusion of Probabilistic Classifiers for Improved Transformer Diagnostics," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* pp. 1-13, 2020, doi: 10.1109/tsmc.2018.2880930.

[20] J. C. S. Sousa, O. R. Saavedra, and S. L. Lima, "Decision Making in Emergency Operation for Power Transformers With Regard to Risks and Interruptible Load Contracts," *IEEE Transactions on Power Delivery,* vol. 33, no. 4, pp. 1556-1564, 2018, doi: 10.1109/tpwrd.2017.2764466.

[21] J. I. Aizpurua, S. D. J. McArthur, B. G. Stewart, B. Lambert, J. G. Cross, and V. M. Catterson, "Adaptive Power Transformer Lifetime Predictions Through Machine Learning and Uncertainty Modeling in Nuclear Power Plants," *IEEE Transactions on Industrial Electronics,* vol. 66, no. 6, pp. 4726-4737, 2019, doi: 10.1109/tie.2018.2860532.

[22] J. Foros and M. Istad, "Health Index, Risk and Remaining Lifetime Estimation of Power Transformers," *IEEE Transactions on Power Delivery,* pp. 1-1, 2020, doi: 10.1109/tpwrd.2020.2972976.

[23] J. Jiang, R. Chen, M. Chen, W. Wang, and C. Zhang, "Dynamic Fault Prediction of Power Transformers Based on Hidden Markov Model of Dissolved Gases Analysis," *IEEE Transactions on Power Delivery,* pp. 1-1, 2019, doi: 10.1109/tpwrd.2019.2900543.

[24] B. Qi, P. Zhang, Z. Rong, and C. Li, "Differentiated warning rule of power transformer health status based on big data mining," *International Journal of Electrical Power & Energy Systems,* vol. 121, 2020, doi: 10.1016/j.ijepes.2020.106150.

[25] V. Tra, B.-P. Duong, and J.-M. Kim, "Improving diagnostic performance of a power transformer using an adaptive over-sampling method for imbalanced data," *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 26, no. 4, pp. 1325-1333, 2019, doi: 10.1109/tdei.2019.008034.

[26] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference: Revised and Expanded*. New York: Marcel Dekker, Inc., 2003.

[27] C. Ferreira, *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence* (Studies in Computational Intelligence). Springer, 2006.

[28] R. Poli, W. B. Langdon, N. F. McPhee, and J. R. Koza, *A Field Guide to Genetic Programming*. Lulu.com, 2008.

[29] C. Wei, Q. Wu, and W. Tang, "Dissolved gas analysis method based on novel feature prioritisation and support vector machine," *IET Electric Power Applications*, vol. 8, no. 8, pp. 320-328, 2014, doi: 10.1049/iet-epa.2014.0085.

[30] P. G. Espejo, S. Ventura, and F. Herrera, "A Survey on the Application of Genetic Programming to Classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* vol. 40, no. 2, pp. 121-144, 2010, doi: 10.1109/tsmcc.2009.2033566.

[31] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A Survey on Evolutionary Computation Approaches to Feature Selection," *IEEE Transactions on Evolutionary Computation,* vol. 20, no. 4, pp. 606-626, 2016, doi: 10.1109/TEVC.2015.2504420.

[32] U. Bhowan, M. Johnston, Z. Mengjie, and Y. Xin, "Reusing Genetic Programming for Ensemble Selection in Classification of Unbalanced Data," *IEEE Transactions on Evolutionary Computation,* vol. 18, no. 6, pp. 893-908, 2014, doi: 10.1109/tevc.2013.2293393.

[33] E. Fernandes, A. C. P. d. L. F. De Carvalho, and X. Yao, "Ensemble of Classifiers based on MultiObjective Genetic Sampling for Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering,* pp. 1-1, 2019, doi: 10.1109/tkde.2019.2898861.

[34] H. Haibo and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering,* vol. 21, no. 9, pp. 1263-1284, 2009, doi: 10.1109/tkde.2008.239.

[35] Y. Kim, A. Goyal, and T. Kumar, "Predictive Modeling of Dissolved Gas Concentration in Oil-Immersed Substation Transformers," presented at the 2016 the 4th IEEE International Conference on Smart Energy Grid Engineering, 2016.

[36] L. L. Minku and X. Yao, "DDD: A New Ensemble Approach for Dealing with Concept Drift," *IEEE Transactions on Knowledge and Data Engineering,* vol. 24, no. 4, pp. 619-633, Apr 2012, doi: Doi 10.1109/Tkde.2011.58.