High Voltage

Research Article

Application of Parzen Window estimation for incipient fault diagnosis in power transformers

ISSN 2397-7264
Received on 10th July 2018
Accepted on 10th August 2018
E-First on 17th September 2018
doi: 10.1049/hve.2018.5061
www.ietdl.org

Md Mominul Islam[1] ✉, Gareth Lee[1], Sujeewa Nilendra Hettiwatte[2]
[1]*School of Engineering and Information Technology, Murdoch University, Murdoch, Perth, Australia*
[2]*School of Engineering, National School of Business Management, Homagama, Sri Lanka*
✉ *E-mail: momin0355@yahoo.com*

Abstract: Accurate faults diagnosis in power transformers is important for utilities to schedule maintenance and minimises the operation cost. Dissolved gas analysis (DGA) is one of the proven and widely accepted tools for incipient fault diagnosis in power transformers. To improve the accuracy and solve the cases that cannot be classified using Rogers' Ratios, IEC ratios and Duval triangles methods, a novel DGA technique based on Parzen window estimation have been presented in this study. The model uses the concentrations of five combustible hydrocarbon gases: methane, ethane, ethylene, acetylene and hydrogen to compute the probability of transformers fault categories. Performance of the proposed method has been evaluated against different conventional techniques and artificial intelligence-based approaches such as support vector machines, artificial neural networks, rough sets analysis and extreme learning machines for the same set of transformers. A comparison with other soft computing approaches shows that the proposed method is reliable and effective for incipient fault diagnosis in power transformers.

1 Introduction

Power transformers are the most expensive and critical components in power transmission and distribution networks. A catastrophic failure of a transformer can jeopardise the stability of many power systems. Therefore, its reliable operation is essential to ensure continuous power supply and prevent a great financial loss for the utility companies. According to Dhote and Helonde [1], about 80% of transformer faults occur from incipient deterioration that could be identified through predictive maintenance and online monitoring techniques. Therefore, condition monitoring and early fault diagnosis techniques are gaining more attention among utilities for preventing unscheduled outages and minimise their operational risks.

Owing to continuous operation, faults and overloading, a power transformer is subjected to thermal, electrical, chemical and mechanical stresses throughout its operating life. These stresses may cause decay of insulating oil and release some gases which become dissolved in the dielectric fluid. The gas concentrations may be measured using gas chromatography [2] and analysed by different dissolved gas analysis (DGA) techniques to indicate the fault afflicting a transformer. The gas concentrations observed in transformers under incipient fault condition increase as a function of temperature, and their individual concentration depends on the type of fault allowing their prevalence to be used as a fault detector [3]. For instance, hydrogen (H_2) and methane (CH_4) start to form under low thermal stress at about 150°C and are an indicator of partial discharge (PD), while temperatures over 500°C lead to the formation of acetylene (C_2H_2) which is an indicator of arcing. Moreover, the concentration of carbon dioxide, carbon monoxide and their ratios can be used to assess the condition of paper insulation as they are produced by the degradation of solid insulation [4]. As DGA is a widely accepted, proven and non-invasive incipient faults detection method in transformers, its popularity increased over time.

Table 1 Permissible dissolved gases concentration (ppm) in healthy power transformers [4]

Gas	<4 years	4–10 years	>10 years
CH_4	70	150	300
C_2H_4	150	200	400
C_2H_6	50	150	1000
C_2H_2	30	50	150
H_2	150	300	300

To analyse the gas concentrations in transformer insulating oil, different DGA techniques such as modified Rogers' ratios, IEC ratios, Doernenburg, key gas method and Duval triangles have been used over recent decades [5, 6]. Some of the methods use gas ratios, while others use specific gas concentrations to indicate the condition of a transformer. Although the implementation of these conventional methods is easy, they have shortcomings, leading to incorrect diagnosis or unresolved diagnosis. According to Singh and Bandyopadhyay [4], the acceptable dissolved gases concentration in oil-filled transformers at different operating ages is shown in Table 1.

Previous investigations have employed a diverse range of artificial intelligence (AI) and machine learning (ML) techniques including artificial neural networks (ANNs) [7], support vector machines (SVMs) [8], fuzzy logic [9, 10], neuro fuzzy systems [11], along with the nearest neighbour clustering approach (NNCA) [12] to overcome the limitations associated with the established DGA methods. Results presented here are drawn from a DGA data set describing 376 transformers operated by a Western Australian utility company. The data set enables quantitative analysis of the critical cases, where Duval triangles and conventional ratio methods fail to correctly (or unambiguously) classify a transformer. The combustible gas concentrations have been analysed by a novel probabilistic density function based on the Parzen Window (PW) method [13], which is shown to be more effective when dealing with the critical cases that cause problems with established methods. The arrangement of this paper is as follows: Section 2 describes the motivation for the research. Section 3 introduces the concept of PW estimation. Section 4 describes the methodology, while Section 5 presents experimental results measuring the comparative performance of the new method. Section 6 evaluates and compares the proposed method with other methods. Section 7 presents a case study and Section 8 concludes with a summary of the results.

2 Motivation for research

DGA is a widespread diagnostic technique that has gained worldwide acceptance for incipient fault diagnosis in transformers [14]. As a result of stress during operation (electrical, mechanical and thermal), the dielectric properties of oil, as well as solid insulation (paper and pressboard), become degraded over time. The decomposition of insulating material produces different combustible and non-combustible fault gases and increases the risk to transformers during operation. The DGA-based analysis involves measuring and monitoring the concentration and production rate of gases to assess the insulation condition of a transformer and indicate developing faults. To interpret dissolved gas concentrations, graphical techniques such as Duval triangles or ratio methods are used. Although relatively easy to implement, each of these established methods has different advantages and limitations. Therefore, comparison of the results from applying different methods on the same sample can lead to contradictions. Should this occur, there is no proven way to prioritise one method over another [15]. The accuracy of the IEC ratio is often compromised due to the incomplete coding (limited classifications) and strict ratio limits. It is unable to identify faulty samples that fall outside its defined ratio limits. In addition, the IEC low- and high-energy discharge categories (D1/D2, respectively) can interfere, resulting in a misleading classification [16]. Moreover, the classification from the Rogers' ratios method is unable to detect all faults precisely [16]. Its diagnosis is most accurate for the low thermal (T1) fault [12]. The Doernenburg ratio method can only provide three types of diagnoses by comparing its different limit values and cannot distinguish the severity of any thermal decomposition (The detailed procedure for this method is available in an IEEE standard [17].). All these ratio methods are empirical and lack a theoretical basis, so their accuracy is dependent on concentration thresholds and ratios that may vary from expert to expert. In some practical cases, the calculated ratios do not fall within any of the defined fault classes and hence remain unclassified. The Duval triangles method always provides a fault diagnosis even for transformers that are known to be healthy. The classical Duval triangle is unable to accurately detect both PD and thermal faults [12]. For transformers filled with mineral oil, and if the fault classification is either a thermal fault or a PD by the classical triangular method, then triangles 4 and 5 must be used to offer additional clarification. Unfortunately, triangles 4 and 5 can subsequently sometimes result in contradictory classifications. Moreover, all triangles contain an unclassified region. Consequently, the efficacy of fault classification with these methods is heavily dependent on an expert's past experience and ability to interpret their results. To improve the diagnostic capability of these standard methods, different AI and ML techniques such as ANNs, SVMs, fuzzy logic and nearest neighbour classification (NNCA) have been introduced [7–10, 12]. Although they can solve the problem of many unresolved and wrong diagnoses to a large extent, each has additional limitations. For instance, an ANN needs to be trained on a large set of examples to ensure reliable classification. Similarly, in fuzzy logic, the derivation of effective rules may prove difficult. Similarly, the wavelet network has high efficiency but low convergence [3]. Finally, in NNCA, detection of cluster centres and partitioning them into different fault categories are critical but hard problem.

3 Basic concepts of PW estimation

Probability density function (PDF) estimation is prevalent in many statistical techniques for analysing numerical data. Different PDF methods are currently available to estimate the density of unknown measurements. PW is a popular non-parametric method and is used here to estimate the probability of an unknown transformer fault class based on the distribution of known transformer measurements [18, 19]. It is a form of inductive learning that estimates the PDF from a finite set of examples drawn from the distribution. As it is non-parametric, it does not need to estimate the values of a large set of synaptic weights as would be the case with ANNs.

The PW method estimates the common PDF $p(x)$ for an independent and identically distributed finite observation of any measurements X. The shape of the density function $p(x)$ is entirely dependent on the sample data and its accuracy moves toward the true value with an increased number of observations [19]. Therefore, no assumed functional form is necessary to estimate the PDF of unknown measurements [13]. According to [20], the probability that a measurement x belongs to a region R that is a subset of the domain can be expressed by the equation below:

$$P = \int_R p(x)\,\mathrm{d}x \tag{1}$$

If the region $\mathscr{R}$ is assumed to be very small, the probability density $p(x)$ within $\mathscr{R}$ can be considered constant. Therefore, (1) can be approximated to the equation below:

$$P = p(x)V \tag{2}$$

where V is the volume of $\mathscr{R}$. If N samples $X_{1,}$, $X_{2,}$, $X_{3,}$, …, X_N are drawn from a distribution in D-dimensional space and each is a vector $\boldsymbol{X}_n = [x_{n,\,1,}, x_{n,\,2}, \ldots, x_{n,\,D}]$, then a probability density $p(x)$ predicts the number of samples K out of n fall inside the $\mathscr{R}$ region can be estimated by

$$K \simeq NP \tag{3}$$

Rearranging (2) and (3), the probability density can be approximated as

$$p(x) \cong \frac{K}{nV} \tag{4}$$

If the region $\mathscr{R}$ is treated as a hypercube centred on X_n with side length σ, then its volume will be $V = \sigma^D$. The number of samples (K) belonging to the region $\mathscr{R}$ can be calculated through a function $k(x)$ that meets the following conditions [19]:

$$k(\boldsymbol{x}) = \begin{cases} 1 & |x_i| \leq 0.5 \quad i = 1, 2, \ldots, D \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $\boldsymbol{x} = [x_{1,}, x_2, \ldots, x_D]$. Therefore, the value of K can be defined as

$$K = \sum_{n=1}^{N} k\left(\frac{x - X_n}{\sigma}\right) \tag{6}$$

The new expression for probability density at any sample point x can be calculated by substituting the value of K from (6) into (4), which is shown in the equation below:

$$p(x) = \frac{1}{N\sigma^D} \sum_{n=1}^{N} k\left(\frac{x - X_n}{\sigma}\right) \tag{7}$$

Equation (7) is considered to be the basic formulation of PW estimation, where $k(x)$ is a statistical kernel [21] or window function. Different kernels such as rectangular (5) or Gaussian kernels can be applied to define a window function. As the Gaussian function is smooth, in this research, a multivariate Gaussian kernel is commonly applied to obtain a smoother density model. Moreover, in a special form of radially symmetrical Gaussian, the function can be completely specified by using a variance parameter only [21]. Thus, the PW density function using a Gaussian kernel function with a common covariance $\boldsymbol{\Sigma}$ can be written as

$$p(x) = \frac{1}{N\sqrt{(2\pi)^D \times |\boldsymbol{\Sigma}|}} \sum_{n=1}^{N} \exp\left(-\frac{(x - X_n)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(x - X_n)}{2}\right) \tag{8}$$

where $\boldsymbol{\Sigma}$ is a kernel covariance matrix (multivariate standard deviation) that decides the shape of the estimated PDF [13]. From

Table 2 Input and targeted output of the proposed method [12]

Input		Targeted fault category
1. $\%H_2 = \frac{H_2}{TCG} \times 100$	1.	PD
	2.	discharge of low energy (D1)
2. $\%CH_4 = \frac{CH_4}{TCG} \times 100$	3.	discharge of high energy (D2)
	4.	thermal fault, $t < 150°C$ (S)
3. $\%C_2H_4 = \frac{C_2H_4}{TCG} \times 100$	5.	thermal fault, $150°C < t < 300°C$ (O)
	6.	thermal fault, $300°C < t < 700°C$ (C)
4. $\%C_2H_6 = \frac{C_2H_6}{TCG} \times 100$	7.	thermal fault, $t > 700°C$ (T3)
5. $\%C_2H_2 = \frac{C_2H_2}{TCG} \times 100$		

(8), it can be seen that the estimated density function is obtained by summing the kernels of representative samples drawn from the set of measurements. The smoothness of the function increases with the increased value of $\boldsymbol{\Sigma}$ and gradually starts to lose information [20]. However, if the value is too small, the model becomes very sensitive to sample noise. The optimal value of $\boldsymbol{\Sigma}$ can be estimated by analysing the training set discussed in a later section. After estimating the optimal value of $\boldsymbol{\Sigma}$, the probability density of any test measurement can be calculated using (8).

4 Methodology

This section focuses on development of an effective PDF based on a limited number of transformer measurements to identify the fault category of a previously unseen transformer. (In this sense, it is like training an ANN from a training set and then testing with a second disjoint set of measurements). Different sections of the methodology such as data collection, pre-processing, density function estimation, training and feature selection have been discussed below.

4.1 Data processing and normalisation

As a basis for the experiment, concentrations of five combustible gases H_2, CH_4, ethylene (C_2H_4), ethane (C_2H_6) and C_2H_2 were measured for 376 power transformers. The concentrations of gases were measured in parts per million (ppm). The concentrations were collected by sampling the oil of each transformer's main tank and analysing it in a laboratory. In the case of an abrupt change in concentration between successive scheduled measurements, the sample was recollected and then re-examined in multiple laboratories to verify the actual concentration. To differentiate the faulty transformers and classify their fault categories, gas concentrations were analysed using conventional approaches such as Duval triangles, Doernenburg, Rogers' ratios and IEC ratio methods. Moreover, sophisticated software and expert judgement were also used to determine the most likely fault category of each transformer. In some cases, a faulty transformer had already been removed from service to investigate the ultimate fault category. In most cases, the findings exactly match the expert classification. Therefore, in this work, it is assumed that the final fault category of transformers decided from the combination of expert judgement and established methods are reliable and accurate. For the proposed PDF estimation technique, the amount of total combustible gases (TCGs) for each transformer has been calculated by adding the five combustible gases as in the equation below:

$$\mathrm{TCG} = H_2 + CH_4 + C_2H_4 + C_2H_6 + C_2H_2 \tag{9}$$

After calculating the TCG, individual percentages of the combustible gases were computed and used as an input to the proposed PW method. The percentage calculation procedure and the seven-targeted fault classification based on the PW method are summarised in Table 2.

Finally, the 376 collected measurements were divided into disjoint training (85%) and testing (15%) subsets. The testing subset has been purposely selected to include the critical cases that could not be classified unambiguously with conventional methods and cause a conflicting classification by using Duval triangles methods. Therefore, measurements from 318 transformers were used as a training set for the proposed model, whereas the remaining 58 transformer measurements were used to evaluate its performance.

4.2 Density function estimation

In this research, a non-parametric-based PW technique has been applied to estimate the density function. Each of the concentrations shown in Table 2 has been concatenated to form one point in a five-dimensional (5D) space. For PDF estimation, a Gaussian kernel function has been centred on each of the 318-point training set. The individual kernels are added together by determining a common width that is known as the smoothing parameter to estimate the probability density of each measurement. A mathematical expression of PDF is shown in (8).

The challenging task when applying the proposed method is to precisely estimate the value of the covariance matrix ($\boldsymbol{\Sigma}$) for 5D data providing a smoothing factor. The smoothing parameter is very important as the shape of PDF depends on it. Therefore, it has a great influence on the measured performance. Although a larger value of $\boldsymbol{\Sigma}$ will make the estimated PDF curve smoother, the estimated curve will also start to lose finer details. However, a smaller value of $\boldsymbol{\Sigma}$ may lead to false spikes in PDF curve depending on the specific distribution of the training points and thus becomes prone to noise. The optimal value of $\boldsymbol{\Sigma}$ depends on the size of the training measurement set and the amount of superimposed noise [18]. Probability distributions of H_2 are shown in Fig. 1, where the y-axis represents probability $p(x)$.

From Fig. 1, it is obvious that the probability distribution of H_2 becomes smoother with the increased value of sigma (covariance). To estimate the optimal value for $\boldsymbol{\Sigma}$, different methods such as Silverman's rule of thumb [22] or a leave-one-out estimator [18] can be used. Hu in [19] chose the optimal value of covariance $\boldsymbol{\Sigma}$ from $a/\sqrt{n}$, where a is a constant and n is the sample size. These methods proved not to be effective on the proposed higher-dimensional method. Therefore, the value of $\boldsymbol{\Sigma}$ has been estimated by search increments of a multiplicative factor (E) with a step size 0.001 in the range 0.001–1.0. With increased values of E, the accuracy of the model increased up to a certain level. When the value of E increased still further, the accuracy started to drop. As a result, the value at which maximum accuracy was found is considered as a near-optimal point for this experiment and which was 0.25.

However, to calculate the inverse of $\boldsymbol{\Sigma}$ following (8), in some fault classes, the covariance matrix becomes singular due to the limited number of features (some gases are absent). This means it is impossible to calculate the matrix inverse as needed by (8). To overcome this singularity problem, a regularisation technique [23] was applied. In this approach, a matrix with very small diagonal values (having the same dimensions as $\boldsymbol{\Sigma}$) was added to the covariance matrix. The diagonal values were controlled by a small multiplicative factor λ. The equation of covariance calculation for the proposed 5D case can be expressed by the equation below:

$$\Sigma = E\boldsymbol{\Sigma}_{\mathrm{S}} + \lambda I \tag{10}$$

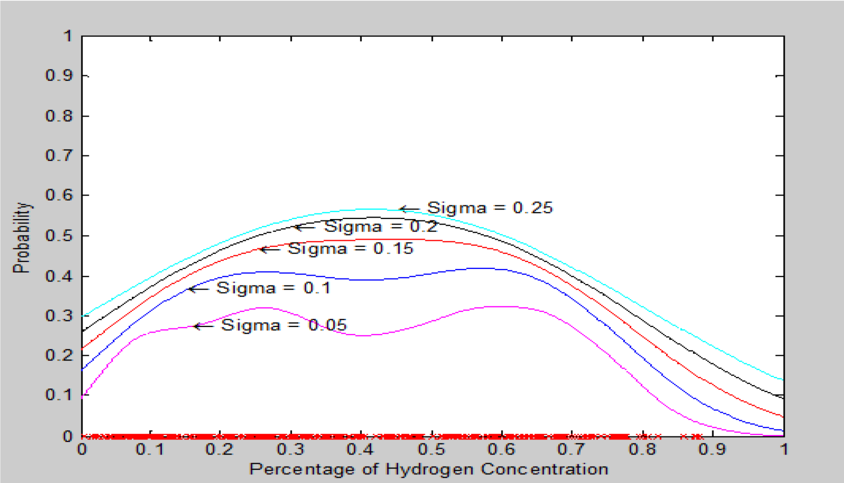

Fig. 1 *Probability distribution of H_2*

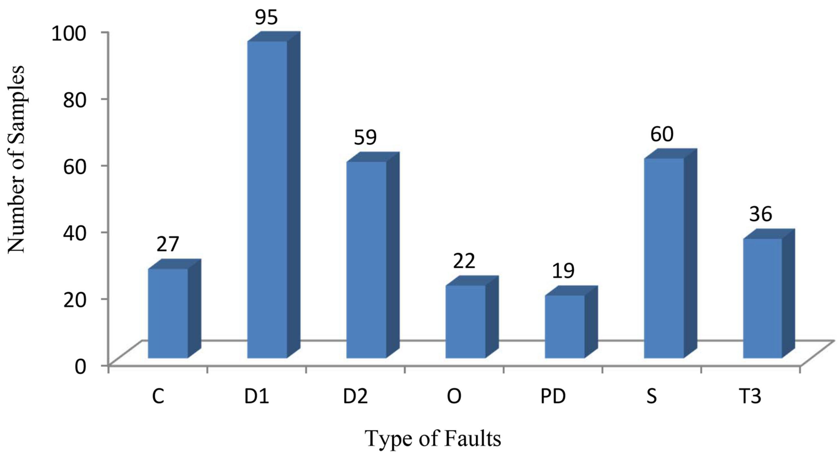

Fig. 2 *Training samples following the fault categories [12]*

where $\boldsymbol{\Sigma}_S$ is the covariance of the training samples, λ is a constant and $\boldsymbol{I}$ is an identity matrix having the same dimensions of the samples.

4.3 Training and feature extraction

The model was trained using the percentages of combustible gas concentrations from 318 transformers. The fault classifications of the transformers are labelled by the utility experts through a combination of different established methods, software analysis and their professional experience. The training sample based on fault categories is shown in Fig. 2, where the x-axis represents fault categories.

It can be seen from Fig. 2 that the utility experts have classified their transformers into seven fault categories. These categories have been used to form groups, where each group contains a certain number of faulty transformers. For instance, fault group C is formed from 27 faulty transformers. Following the proposed method, the probability density of each transformer in each group has been calculated for different values of E and λ to get an estimation of their optimal values. Transformers having the highest probability in a particular faulty group will be classified by the fault category of those transformers. From experiments, it has been found that with a small value of $E = 0.01$, the classification of the training data points exactly match with the expert's opinion but performance on the test set is unsatisfactory. The maximum accuracy (94.82%) of the model was found at $E = 0.25$. If the value of E is further increased, the accuracy again starts to drop. Bearing in mind that the proposed method specialises in cases, where the existing methods fail a workflow for combining the approach with the existing Duval triangles method has been shown in Fig. 3. This allows each of these methods to be used on the test measurements where they are most effective.

5 Results and discussion

To test the efficacy of the proposed PW method, the percentage of combustible gas concentrations from transformers have been analysed to generate a classification into one of seven fault categories as listed in Table 2. In each case, a domain expert has also classified the transformer using the same fault labels. In this work, critical cases have been chosen whose classification is ambiguous when using conventional ratio methods or the low-dimensional Duval triangles graphical approach. Therefore, a subset of 58 transformer measurements from 376 was selected as a test set, was excluded from the training phase, and subsequently classified using the PWs method. This allowed the proposed method's performance to be compared with the IEC ratios, modified Roger's ratios and previous work using the NNCA [12]. Test results have been summarised in Table 3.

The accuracy of the proposed method shown in Table 3 is higher than NNCA when compared with expert decisions (However, given the small test set, it is unclear whether this improvement is statistically significant.). The overall performance of this method is 95% accurate, which is much higher than the conventional ratio methods. In three cases, the method wrongly diagnoses the fault category of transformers but it can act as a decision-support mechanism for engineers responsible for assessing these critical cases. Moreover, this new approach shows probable advantages over earlier NNCA methods. In NNCA, which is a combination of the k-means algorithm (KMA), k-nearest neighbour algorithm and Linde–Buzo–Gray (LBG), determination of actual centre locations is critical. As KMA only converges to local minima, different positions of initial cluster centres lead to different final clusterings [24]. To overcome this problem partially, a hybrid approach was developed that combined Lloyd's and LBG algorithms with conventional KMA [25]. However, there is no

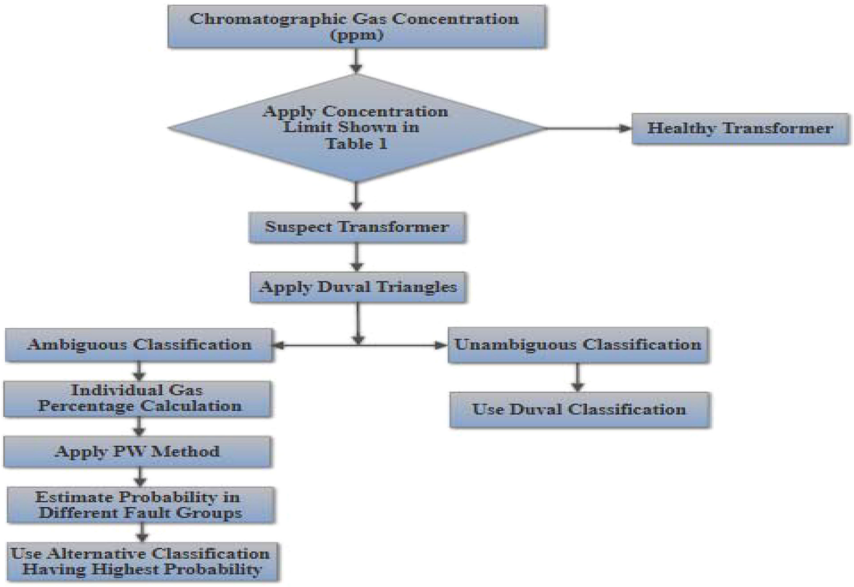

Fig. 3 *Workflow of the proposed model for practical application*

Table 3 Comparison of the proposed method with IEC ratios, Rogers' ratios and the NNCA methods [12]

Test methods	Unresolved diagnosis	Wrong diagnosis	Accuracy, %
IEC ratio	27	8	74.19
Roger's ratios	21	9	75.67
NNCA	—	4	93.10
proposed method	—	3	94.82

Table 4 Comparison between decision tree, ratio and the proposed method (shown in bold) [12, 16]

Case number	H_2	CH_4	C_2H_4	C_2H_6	C_2H_2	Diagnosis result					
						Actual diagnosis	New approach DGA	Modified new approach DGA	Roger's ratios	IEC ratio	Proposed method
1	117	17	3	1	1	PD	PD, D1	PD	—	—	**PD**
2	32,930	2397	0	157	0	PD	PD, D1	PD	PD	PD	PD
3	78	20	13	11	28	D1	D1, D2	D1	—	D1	D1
4	1230	163	233	27	962	D1	D1, D2	D1	D2	D1	D1
5	8200	3790	4620	250	277	D2	D2, T1	D2	—	—	D2
6	130	140	120	2	0	T1	D2, T1	T1	T3	T3	T3
7	78	66	2.6	283	0	T1	D2, T1	T1	normal	PD	T1
8	30.4	117	138	44.2	0.1	T2	T2, T3	T2	T3	T2	T2
9	27	90	63	42	0.2	T2	T1, T2	T1	T2	T2	T2
10	1100	1600	2010	221	26	T3	T2, T3	T3	T3	T3	T3

guarantee that data will be optimally assigned to the various clusters. The advantages of the PW method are as follows:

i. Information on the cluster centres is not necessary.
ii. It can be used without knowledge of the probability distribution of a training set.
iii. No assumptions are required to classify a faulty class of transformers.
iv. The complexity and correlation of input random variables do not affect the performance of the proposed algorithm.

6 Evaluation and comparison with other methods

To better evaluate the performance of the proposed method, it has also been applied to the test sets generated by other researchers [16, 26, 27]. The authors of these papers have done a similar type of experiment but have focused on different approaches to identify the fault category of transformers. They have also trained their classifiers using training data, labelled by the utility experts. In [16], researchers diagnosed the category of a faulty transformer using a decision tree based on the concentration of combustible key gas limits and named it as a new approach DGA. They found that different faults having the same gas limit overlapped and were inseparable. To deal with the overlapping problem, an additional gas ratio was included into the decision mechanism and renamed it as the modified new approach DGA. The performance of the proposed PW method has been compared with the decision tree approach [16] and conventional ratio methods in Table 4.

From Table 4, it can be seen that the correct diagnosis rate of the proposed PW method is higher than those of the IEC and Roger's ratios methods. Moreover, the proposed method has accurately classified all the cases, except case number 6.

The proposed method was also evaluated on other test sets disclosed in the literature [26], where fault categories were determined by applying a rough sets (RSs) analysis technique and where ANNs were combined with *k*-means clustering (KMC) algorithms and RS, respectively. The comparative performance of these methods and other established methods versus the proposed method are summarised in Table 5.

Application of this linear approach to classify a transformer's fault category may not be effective for critical cases. To deal with the non-linear problem, different AI and ML techniques such as ANNs, a SVM, an extreme learning machine (ELM) and self-

Table 5 Fault diagnosis comparison between established methods and adapted methods [12, 26]

Case number	H_2	CH_4	C_2H_4	C_2H_6	C_2H_2	Diagnosis result							
						Expert diagnosis	Duval triangle	RS	RS–ANN	KMC–RS–ANN	Roger's ratios	IEC ratio	Proposed method
1	60	40	110	10	70	D2	D2	D2	D2	D2	D2	D2	D2
2	31	7	5	19	67	D2	D1	—	D2	D2	—	—	D1
3	293	50	15	13	120	D2	D2	D2	D2	D2	D2	D1	D2
4	57	7	4.5	19	71	D2	D1	—	D2	D2	—	—	D1
5	467	148	266	13	511	D2	D2	D2	DT	D2	D2	D1	D2
6	160	90	17	27	58	D1	D1	D1	D1	D1	—	—	D1
7	402	81	27	39	25	D1	D2	D1	D1	D1	—	—	S
8	4	79	312	112	0	T2	T1	—	T2	DT	T2	T2	T2
9	180	180	4	74	3	DT	T3	DT	DT	DT	—	—	T2
10	1300	740	2000	260	71	T3	T3	T3	T3	T3	—	—	T3
11	42	97	600	157	0	T3	T3	normal	normal	normal	T2	T2	T2
12	44	52	119	15	1	T3	T3	T3	T3	T3	T3	T3	T3
13	42	79	152	31	1	T3	T3	T3	T3	T3	T3	T3	T3
14	164	244	497	103	8	T3	T3	T3	T3	T3	T3	T3	T3
15	22	51	57	42	0	T2	T2	T2	T2	T2	T2	T2	T2
16	679	4992	3671	1823	0	T2	T1	T2	T2	T2	T2	T2	T2

Table 6 Comparison of different adapted technique and proposed methods [12, 27]

Case number	H_2	CH_4	C_2H_4	C_2H_6	C_2H_2	Diagnosis result							
						ANN	SVM	ELM	SaE-ELM	Duval triangle	Roger's ratios	IEC ratio	Proposed method
1	103	5.8	7.3	5	0.7	T1	T3	T1	T1	S	—	—	T1
2	416	21	43.1	10.5	1	T1	T3	T1	T3	T3	—	—	T2
3	59	53	60.3	17.7	0.8	T2	T2	T2	T2	C	—	—	T2
4	10.5	4.8	4.8	5	2.2	D1	D1	D1	D1	DT	—	—	D1
5	137	97	29	12	1.5	T2	T2	T2	T2	C	Low-energy density arcing (LEDA)/PD	—	T2
6	89	73	6.8	6	5	D2	D2	D2	D2	DT	—	—	T2
7	240	157	127	98	0.8	T2	T2	T2	T2	C	LEDA/PD	—	T2
8	116	104	51	36	0	T2	T2	T2	T2	C	LEDA/PD	—	T2

adaptive evolutionary ELM (SaE-ELM) were proposed in [27–29]. The authors classified each test transformer using these four methods and a final decision is taken following the majority of the votes. A comparison of the performance of these assorted AI and ML techniques versus the PW probability density estimation method is shown in Table 6.

7 Case study

To provide a better insight, all the cases in the literature have been used in a case study. Except for case number 6, the proposed method has solved all the overlapping problems shown in Table 4. Case number 6 differs from the expert classification (T1), has been classified as T3. The method also has successfully classified 12 cases out of 16 as shown in Table 5. The wrongly classified cases are 2, 4, 7 and 9. The expert has classified those problems as D2, D2, D1 and Thermal and dielectric (DT), whereas the proposed method classifies them as D1, D1, S and T2, respectively. The Duval triangle method classified the faults as D1, D1, D2 and T3, respectively, but the other two established methods (Roger's ratio and IEC) failed to determine these cases as their ratios fell outside the defined limits. As the utility company that owns these assets predominantly use Duval triangles to classify their transformers faults, this could explain the misclassification of cases 2, 4 and 7. The DT fault combines aspects of thermal and dielectric faults and was, therefore, not included in the targeted categories of this research. Therefore, it has been classified as a T2 type thermal fault. A similar problem also arises in case number 6 as shown in Table 6. This case is classified by the SVM, ANN, ELM and SaE-ELM methods as suffering a D2 fault. Similarly, for the cases in Table 6, the IEC method did not detect any fault and the performance of Roger's ratio was also unsatisfactory. The failure of the proposed method could possibly be overcome by increasing the number of random samples drawn from the larger number of training measurements.

In this research, the optimal value of covariance for 5D data sets is carefully chosen through a continuously supervised iterative process. All the 5D gas concentration information has been deliberately preserved throughout the process, aiming at a better classification than with the Duval triangles. In Duval's approach, gas concentrations are mapped into 2D spaces prior to classifying each transformer's fault category. Deliberately preserved these extra dimensions has probably helped to accurately classify 95% of the cases. Moreover, the performance of this method is satisfactory on the published data which has been shown in Tables 3–5.

8 Conclusions

A new procedure for DGA based on PDF estimation by using the PW method is introduced in this paper. The method is specialised to deal with the difficult cases, where Duval's triangle fails to provide a definitive fault classification. To develop a more reliable and effective fault classifier, five key gas concentrations from 376 power transformers were normalised into percentage form. The new approach is straight forward and easy to apply without the knowledge of cluster centres. The comparative results shown in Tables 4–6 demonstrate that the performance of the PW method is much better than the conventional ratio-based diagnostic strategies and comparable with different AIs and ML techniques such as the ANN, SVM, ELM, SaE-ELM and NNCA. The experimental results in Table 3 show that it correctly classifies 94.8% of these difficult cases, where Duval's triangle provides an ambiguous classification. As the accuracy of this method is dependent on the number of training samples, the accuracy could be improved and the repair cost for a transformer may be reduced by having a larger number of training samples.

9 References

[1] Dhote, N.K., Helonde, J.B.: 'Diagnosis of power transformer faults based on five fuzzy ratio method', *WSEAS Trans. Power Syst.*, 2012, **6**, (3), pp. 114–125

[2] Agah, M., Lambertus, G.R., Sacks, R., *et al.*: 'High-speed MEMS-based gas chromatography', *J. Microelectromech. Syst.*, 2006, **15**, (5), pp. 1371–1378

[3] Khan, S.A., Equbal, M.D., Islam, T.: 'A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models', *IEEE Trans. Dielectr. Electr. Insul.*, 2015, **22**, (1), pp. 590–596

[4] Singh, S., Bandyopadhyay, M.N.: 'Dissolved gas analysis technique for incipient fault diagnosis in power transformers: a bibliographic survey', *IEEE Electr. Insul. Mag.*, 2010, **26**, (6), pp. 41–46

[5] Ashkezari, A.D., Saha, T.K., Ekanayake, C., *et al.*: 'Evaluating the accuracy of different DGA techniques for improving the transformer oil quality interpretation'. The 21st Australasian Universities Power Engineering Conf. (AUPEC 2011), Brisbane, Australia, 25–28 September 2011

[6] Bakar, N.A., Abu-Siada, A., Islam, S.: 'A review of dissolved gas analysis measurement and interpretation techniques', *IEEE Electr. Insul. Mag.*, 2014, **30**, (3), pp. 39–49

[7] Akbari, A., Setayeshmehr, A., Borsi, H., *et al.*: 'Intelligent agent-based system using dissolved gas analysis to detect incipient faults in power transformers', *IEEE Electr. Insul. Mag.*, 2010, **26**, (6), pp. 41–46

[8] Pal, M., Foody, G.M.: 'Feature selection for classification of hyperspectral data by SVM', *IEEE Trans. Geosci. Remote Sens.*, 2010, **48**, pp. 2297–2307

[9] Su, S., Mi, C., Lai, L.L., *et al.*: 'A fuzzy dissolved gas analysis method for the diagnosis of multiple incipient faults in a transformer', *IEEE Trans. Power Syst.*, 2000, **15**, (2), pp. 593–598

[10] Islam, S.M., Wu, T., Ledwich, G.: 'A novel fuzzy logic approach to transformer fault diagnosis', *IEEE Trans. Dielectr. Electr. Insul.*, 2000, **7**, (2), pp. 177–186

[11] Naresh, R., Sharma, V., Vashisth, M.: 'An integrated neural fuzzy approach for fault diagnosis of transformers', *IEEE Trans. Power Deliv.*, 2008, **23**, (4), pp. 2017–2024

[12] Islam, M.M., Lee, G., Hettiwatte, S.N.: 'A nearest neighbour clustering approach for incipient fault diagnosis of power transformers', *Electr. Eng.*, 2016, **99**, pp. 1109–1119

[13] Parzen, E.: 'On estimation of a probability density function and mode', *Ann. Math. Stat.*, 1962, **33**, pp. 1065–1076

[14] Saha, T.K.: 'Review of modern diagnostic techniques for assessing insulation condition in aged transformers', *IEEE Trans. Dielectr. Electr. Insul.*, 2003, **10**, (5), pp. 903–917

[15] IEC, 60599 Ed. 2.1 Mineral oil-impregnated electrical equipment in service – guide to the interpretation of dissolved and free gases analysis, in, 1999

[16] Ghoneim, S.S.M., Taha, I.B.M.: 'A new approach of DGA interpretation technique for transformer fault diagnosis', *Int. J. Electr. Power Energy Syst.*, 2016, **81**, pp. 265–274

[17] IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformer, IEEE Std. C57.104, in, 1991

[18] Archambeau, C., Valle, M., Assenza, A., *et al.*: 'Assessment of probability density estimation methods: Parzen window and finite Gaussian mixtures'. IEEE Int. Symp. Circuits and Systems, Island of Kos, Greece, 21–24 May 2006

[19] Hu, Z., Xiao, M., Zhang, L., *et al.*: 'Mahalanobis distance based approach for anomaly detection of analog filters using frequency features and Parzen window density estimation', *J. Electron. Test.*, 2008, **24**, (7), pp. 901–907

[20] Overton, L.M., Padovani, G., Girolami, M.A.: 'ParCrys: a Parzen window density estimation approach to protein crystallization propensity prediction', *Bioinformatics*, 2008, **24**, (7), pp. 901–907

[21] Babu, V.S., Viswanath, P.: 'An efficient and fast Parzen-window density based clustering method for large data sets'. First Int. Conf. Emerging Trends in Engineering and Technology, Nagpur, Maharashtra, India, 16–18 July 2008, pp. 531–536

[22] Silverman, B.W.: '*Density estimation: Chapman & Hall/CRC, number of iterations as parameter*' (London, 1986)

[23] Noschese, S., Reichel, L.: 'Inverse problems for regularization matrices', *Numer. Algorithms*, 2012, **60**, (4), pp. 531–544

[24] Jain, A.K.: 'Data clustering: 50 years beyond *k*-means', *Pattern Recognit. Lett.*, 2010, **31**, pp. 651–666

[25] Kanungo, T., Mount, D.M., Netanyahu, N.S., *et al.*: 'An efficient *k*-means clustering algorithm: analysis and implementation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (7), pp. 881–892

[26] Fei, S.-W., He, Y.: 'A multi-layer KMC-RS-SVM classifier and DGA for fault diagnosis of power transformer', *Recent Pat. Comput. Sci.*, 2012, **10**, (4), pp. 238–243

[27] Li, S., Wu, G., Gao, B.: 'Interpretation of DGA for transformer fault diagnosis with complementary SaE-ELM and arctangent transform', *IEEE Trans. Dielectr. Electr. Insul.*, 2016, **23**, (1), pp. 586–595

[28] Islam, M.M., Lee, G., Hettiwatte, S.N.: 'Application of a general regression neural network for health index calculation of power transformers', *Int. J. Electr. Power Energy Syst.*, 2017, **93**, pp. 308–315

[29] Islam, M.M., Lee, G., Hettiwatte, S.N.: 'Calculating a health index for power transformers using a subsystem-based GRNN approach', *IEEE Trans. Power Deliv.*, 2018, **33**, (4), pp. 1903–1912