Received: 30 January 2024 | Revised: 3 June 2024 | Accepted: 21 June 2024

IET Nanodielectrics

DOI: 10.1049/nde2.12088

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

A combined technique for power transformer fault diagnosis based on *k*-means clustering and support vector machine

Arnaud Nanfak[1] | **Abdelmoumene Hechifa**[2] | **Samuel Eke**[1] | **Abdelaziz Lakehal**[3] | **Charles Hubert Kom**[1] | **Sherif S. M. Ghoneim**[4]

[1]Laboratory of Energy, Materials, Modelling and Methods, National Higher Polytechnic School of Douala, University Douala, Douala, Cameroon

[2]LGMM Laboratory, Faculty of Technology, University of 20 August 1955-Skikda, Skikda, Algeria

[3]Laboratory of Research on Electromechanical and Dependability, University of Souk Ahras, Souk Ahras, Algeria

[4]Electrical Engineering Department, College of Engineering, Taif University, Taif, Saudi Arabia

Correspondence

Arnaud Nanfak, Laboratory of Energy, Materials, Modelling and Methods, National Higher Polytechnic School of Douala, University Douala, Douala 2701, Cameroon.

Email: nanfak.arnaud@yahoo.fr

Abstract

This contribution presents a two-step hybrid diagnostic approach, combining *k*-means clustering for subset formation, followed by subset analysis conducted by human experts. As the feature input vector has a significant influence on the performance of unsupervised machine learning algorithms, seven feature input vectors derived from traditional methods, including Duval pentagon method, Rogers ratio method, three ratios technique, Denkyoken method, ensemble gas characteristics method, Duval triangle method, and Gouda triangle method were explored for the subset formation stage. The seven proposed individual methods, corresponding to the seven feature input vectors, were implemented using a dataset of 595 DGA samples and tested on an additional 254 DGA samples. Furthermore, a combined technique based on a support vector machine was introduced, utilising the diagnostic results of the individual methods as input features. From training and testing, with diagnostic outcomes of 91.09% and 90.94%, the combined technique demonstrated the highest overall diagnostic accuracies. Using the IEC TC10 database, the diagnosis accuracies of the proposed diagnostic methods were compared to existing methods of literature. From the results obtained, the combined technique outperformed the proposed individual methods and existing methods used for comparison.

KEYWORDS

power transformer insulation, transformer oil

1 | INTRODUCTION

Power transformers are an important piece of equipment used to transmit and distribute electrical energy. The failure of these machines can lead to significant financial lost due to interruptions in the distribution of electricity and costly repairs or replacements [1]. Therefore, rapid detection and accurate assessment of emerging or existing internal faults in power transformers are essential to ensure the efficient and safe operation of the power system network [2]. In addition, it is one of the means to increase the operational reliability and the useful service life of these machines. To achieve this goal, several diagnostic methods have been proposed in the literature, such as partial discharge measurement, furans analysis, frequency response analysis, moisture analysis or dissolved gas analysis (DGA) [3]. Among the above techniques, DGA is the most often used to assess transformer conditions. Several DGA-based methods have been proposed in the literature and can be classified into two main categories: traditional and intelligent methods [4].

Traditional methods are based on rules produced by human experts. These rules rely on concentration, concentration ratios and/or percentages of fault-related gases to the various faults. In general, three approaches to the synthesis of traditional methods are used in the literature: The key gas approach, the gas ratios approach, and the graphical approach. According to IEEE C57.104-2019, key gas methods rely on the correlation between the fault types and the generated key gases [5]. Gas ratio methods consider the fault-related gases ratios as diagnostic criterion. These include, among others Rogers ratios method (RRM) [5], IEC 60599 ratios method [6] or Three Ratios Technique (TRT) [7]. Graphical methods are based on the projection in a two- or three-dimensional plane of a point representing the health state of the transformer. The most common graphical methods are Duval Triangle method [5], Mansour pentagon method [8] or Gouda Triangle method [9]. In intelligent methods, DGA data are interpreted using artificial intelligence (AI) tools. The literature suggests intelligent methods that rely on machine learning [10–12], artificial neural networks (ANN) [13] or deep learning [14], among others.

Professionals involved in the maintenance of power transformers frequently employ traditional methods because they are straightforward, simple to understand, and easy to apply. They do, however, have some limits regarding uncertainty and precision. Indeed, the limitations of existing traditional methods, such as incomplete ratio ranges for gas ratio methods or misclassification of faults near the boundaries between adjacent regions for graphical methods, are mainly due to the strong non-linearity that exists between the fault types and fault-related gases produced [15, 16]. In addition to this reason, these limitations could also be explained by the one-step diagnostic approach and the size of training dataset size used in the implementation phase. These elements do not allow the human expert to consider all the characteristics of the dataset when implementing the diagnostic model. This paper addresses these issues using a hybrid two-step diagnostic approach based on a combination of clustering (*k*-means) and human expert analysis.

The proposed methods create subsets using the *k*-means clustering algorithm, and for each subset, a traditional sub-model is produced. The rules of these sub-models are based on the links between the different faults related to the subsets and the concentration ratios of hydrogen, methane, ethane, ethylene, and acetylene. Given that the feature input vector affects how well the algorithm finds intriguing and hidden patterns in unlabelled data [17], seven feature input vectors from traditional methods are studied for subset formation. For each input vector used, a diagnostic model is proposed. In addition to seven individual methods related to seven feature input vectors utilised, a combined technique is also proposed. This combined technique is based on diagnostic results of seven individual models and support vector machine (SVM) algorithm. Indeed, to integrate all the individual methods into the combined technique, an SVM classifier is trained with the diagnostic results of individual methods as feature inputs.

A total of eight diagnostic methods was proposed in this paper. All these methods were carried out using a dataset of 595 samples and tested on a dataset of 254 samples. Another dataset of 117 samples known as the IEC TC10 database will be utilised for validation and the diagnostic outcomes are compared with the DGA-based methods in the literature. The following key points characterise the originality of this contribution:

- This paper presents a two-step hybrid approach that combines human expertise and *k*-means clustering, significantly improving the accuracy of power transformer fault diagnostic.
- The proposal of a combined technique based on the SVM algorithm, which integrates the results of seven individual methods linked to the seven input vectors studied.

The remaining part of this paper is organised as follows: Section 2 presents the principle and the flow chart of the individual methods and combined technique that have been proposed in this paper. This section also introduces the seven feature input vectors and dataset that were utilised to construct the seven individual methods. Section 3 presents the evaluation of the performance and effectiveness of the proposed individual methods and combined technique. The metrics used for the evaluation and the comparison with other methods in the literature conducting using the IEC TC10 database are also presented in this section. Section 4 concludes the paper.

2 | MATERIALS AND METHOD

This section presents the principle and the flow chart of proposed methods. In addition, seven features input vectors and datasets used in this paper are presented.

2.1 | Principle of the individual methods

The principle of the proposed individual methods is based on the clustering and human expert analysis. These methods have two steps: clusters formation and analysis of subsets of dissolved gases. In cluster formation step, 120 clusters are generated using *k*-means algorithm. As a cluster may contain one or more fault types, during the subset's analysis step, human experts propose a traditional sub-model based on gas ratios approach to separate the different fault types associated with the same cluster. The different sub-models are combined to create the final diagnostic model. Figures 1 and 2 show, respectively, the schematic view of the methodology used to implement each individual method and the different steps in the diagnostic procedure once the methods have been implemented. Table 1 presents the ratios used by human expert to separate the different fault types associated with the same subset. In general, this approach could be applied to any data-based diagnostic system, considering certain system specific considerations. To this end, the two steps of clustering followed by cluster analysis to implement sub-models should be applied successively.

In the implementation of the individual methods, the labelled dataset is separated into two groups of data, one for training and the other for testing. The training dataset, consisting of the absolute values of the fault gas concentrations associated to fault type, is first pre-processed. In the data pre-processing stage, the absolute concentrations are used to construct the input feature vector for the clustering algorithm. Once the training dataset has been pre-processed, the sample

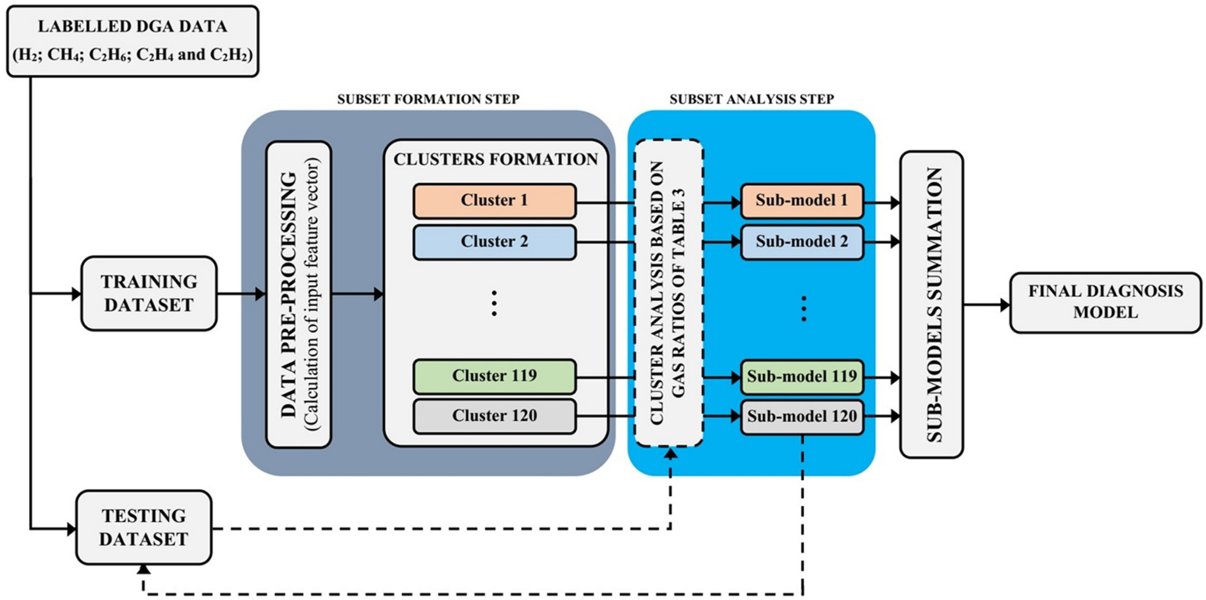

FIGURE 1 Schematic view of the approach used to implement the suggested individual methods into practice.

FIGURE 2 The various steps in the diagnostic procedure of the individual methods.

TABLE 1 Expressions of gas ratios used by human experts to implement diagnostic sub-models.

Ratio	Expression	Ratio	Expression	Ratio	Expression
R_1	$(CH_4 + C_2H_6)/THHG$	R_6	C_2H_2/C_2H_4	R_{11}	$C_2H_2/(C_2H_2 + C_2H_4 + CH_4)$
R_2	$(C_2H_4 + CH_4)/THHG$	R_7	$(C_2H_6 + C_2H_4)/(C_2H_2 + CH_4 + H_2)$	R_{12}	CH_4/H_2
R_3	$C_2H_6/(CH_4 + C_2H_4)$	R_8	$(CH_4 + C_2H_2)/C_2H_4$	R_{13}	C_2H_4/C_2H_6
R_4	$(CH_4 + H_2)/THHG$	R_9	$CH_4/(C_2H_4 + CH_4 + C_2H_2)$	R_{14}	$(C_2H_6 + C_2H_4)/(C_2H_2 + H_2)$
R_5	$(C_2H_4 + C_2H_2)/THHG$	R_{10}	$C_2H_4/(CH_4 + C_2H_4 + C_2H_2)$	R_{15}	$(C_2H_2 + C_2H_6)/C_2H_4$

Note: With THHG: Total hydro hydrocarbon gas $THHG = H_2 + C_2H_6 + C_2H_4 + CH_4 + C_2H_2$.

labels are removed, and 120 clusters are formed using the *k*-means algorithm. Once the clusters have been formed, the labels are put back in place before being analysed independently of each other using gas ratios of Table 2. The human expert proposes a diagnostic sub-model for each cluster based on these gas ratios. Once the 120 sub-models have been proposed, they are all integrated into the final diagnostic model.

2.2 | Principle of the combined technique

The combined technique is based on the diagnostic results of seven individual methods constructed from the seven feature input vectors. To integrate the results of individual methods together in a unified diagnostic model, an SVM classifier is trained with as input features, the diagnostic results of individual methods, and as labels the fault types. Figures 3 and 4 show, respectively, the methodology used to implement the SVM-based combined technique and the different steps in the diagnostic procedure once the methods have been implemented.

2.3 | Data pre-processing and feature input vectors

During the clustering stage, the training data is pre-processed before being used. After this pre-processing, the gas concentration values are replaced by input feature vectors all inspired by traditional methods which are among the most significant and relevant in the literature. Table 2 shows the seven input feature vectors used in this paper. The significance and role of each vector are explained as follows:

TABLE 2 Equation of the seven feature input vectors used to form the clusters.

Vector	Components	Equation and reference
Vector 1	X = [%H_2, %CH_4, %C_2H_6, %C_2H_4, %C_2H_2]	$X(i) = \frac{C_i}{\sum_{j=1}^{5} C_j}$ Where C_i and C_j: C_1 to C_5 are the concentrations (in ppm) of H_2, CH_4, C_2H_6, C_2H_4, and C_2H_2 respectively.
Vector 2	X = [%R_1, %R_2, %R_3]	$\left\{ X(i) = \frac{R_i}{\sum_{j=1}^{3} R_j} \quad with \quad \begin{array}{l} R_1 = CH_4/H_2 \\ R_2 = C_2H_2/C_2H_4 \\ R_3 = C_2H_4/C_2H_6 \end{array} \right.$ Where R_i,R_j:R_1 to R_3 are the ratios of RRM [5]
Vector 3	X = [%R_1, %R_2, %R_3]	$\left\{ X(i) = \frac{R_i}{\sum_{j=1}^{3} R_j} \quad with \quad \begin{array}{l} R_1 = C_2H_2/C_2H_4 \\ R_2 = (C_2H_6 + C_2H_4)/(H_2 + C_2H_2) \\ R_3 = (CH_4 + C_2H_2)/C_2H_4 \end{array} \right.$ Where R_i,R_j:R_1 to R_3 are the ratios of TRT [7].
Vector 4	X = [%R_1, %R_2, %R_3, %R_4, %R_5]	$\left\{ X(i) = \frac{R_i}{\sum_{j=1}^{5} R_j} \quad with \quad \begin{array}{l} R_1 = H_2/H_{2\ \mathrm{lim}} \\ R_2 = CH_4/CH_{4\ \mathrm{lim}} \\ R_3 = C_2H_6/C_2H_{6\ \mathrm{lim}} \\ R_4 = C_2H_4/C_2H_{4\ \mathrm{lim}} \\ R_2 = C_2H_2/C_2H_{2\ \mathrm{lim}} \end{array} \right.$ Where R_i,R_j:R_1 to R_5 are relative concentrations of key gases proposed by the EGC method [18].
Vector 5	X = [%R_1, %R_2, %R_3, %R_4, %R_5]	$\left\{ X(i) = \frac{R_i}{\sum_{j=1}^{5} R_j} \quad with \quad \begin{array}{l} R_1 = H_2/C_{\max} \\ R_2 = CH_4/C_{\max} \\ R_3 = C_2H_6/C_{\max} \\ R_4 = C_2H_4/C_{\max} \\ R_2 = C_2H_2/C_{\max} \end{array} \right.$ Where $C_{\max}$ = max (H_2,CH_4,C_2H_6,C_2H_4,C_2H_2) and R_i,R_j:R_1 to R_5 are relative concentrations of key gases used in Denkyoken method [19].
Vector 6	X = [%CH_4, %C_2H_4, %C_2H_2]	$X(i) = \frac{C_i}{\sum_{j=1}^{3} C_j}$ Where C_i and C_j: C_1 to C_5 are the concentrations (in ppm) of CH_4, C_2H_4 and C_2H_2 respectively. This feature vector is based on the Duval triangle method [5].
Vector 7	X = [%R_1, %R_2, %R_3]	$\left\{ X(i) = \frac{R_i}{\sum_{j=1}^{3} R_j} \quad with \quad \begin{array}{l} R_1 = C_2H_2/C_2H_4 \\ R_2 = (C_2H_6 + C_2H_4)/(H_2 + C_2H_2) \\ R_3 = (CH_4 + C_2H_2)/C_2H_4 \end{array} \right.$ Where R_i,R_j:R_1 to R_3 are the ratios of Gouda triangle method [9].

- **Vector 1**: The vector consists of the percentage concentrations of the fault-related gases (H_2, CH_4, C_2H_6, C_2H_4 and C_2H_2). This vector eliminates the effect of individual gas concentration values and ensures that each gas is considered according to its proportion in the whole.
- **Vector 2**: The vector consists of the percentages of the three gas ratios proposed by the RRM. The gas ratios of this method have been used as diagnostic criteria in other methods for diagnosing faults in power transformers, such as the IEC 60599 ratio method or the SOU-N EE 46.501:2006 method.
- **Vector 3**: The vector consists of the percentages of the three gas ratios proposed by the TRT. This ratio method uses three new gas ratios selected based on their ability to distinguish faults according to their severity.
- **Vector 4**: The vector consists of the percentages of the relative concentrations of the key gases proposed by the

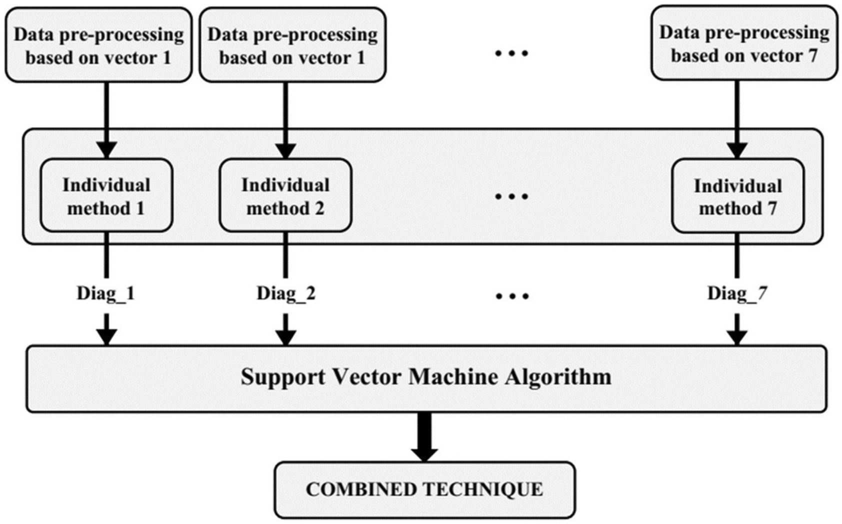

FIGURE 3 Schematic view of the approach used to implement the suggested combined technique into practice.

FIGURE 4 The various steps in the diagnostic procedure of the combined technique.

ensemble gas characteristics (EGC) method [18]. As the relative concentrations are calculated in relation to the permissible value of each key gas respectively, they allow the contribution of each gas to be considered in the interpretation of dissolved gases.

- **Vector 5**: The vector consists of the percentages of the relative concentrations of the key gases proposed by the Denkyoken method [19]. The relative concentrations proposed in this method allow the major gas and the proportions of the other gases in relation to the major gas to be considered in the process of interpreting dissolved gases.
- **Vector 6**: The vector consists of the percentages of the three key gases used in the Duval's triangle method. Although the Duval's triangle method has been criticised in the literature for not considering all the gases associated with the fault, the gases selected by Duval in this method provide sufficient information for the detection and identification of the six main IEC faults.
- **Vector 7**: The vector consists of the percentages of the three gas ratios used in the Gouda's triangle method. As with TRT, the gas ratios of this graphical method have been proposed because of their ability to distinguish between faults according to their severity.

2.4 | Datasets used

The study employed 849 labelled DGA data to evaluate and apply the suggested diagnostic techniques. Low energy discharge (D_1), high energy discharge (D_2), Partial discharge (PD), low thermal fault (T_1), medium thermal fault (T_2), and high thermal fault (T_3) are the six main fault types that make up these DGA data. Table 3 displays the quantity of samples categorised by fault kind and reference. As indicated in Table 4, the data collected was randomly divided into two data sets: a test and a training set with a 70:30 training to test ratio.

To verify the efficacy of the suggested fault diagnosis models, a new dataset, not part of the implementation stage, was used. This is the IEC TC10 database containing 117 labelled DGA data [27]. Tables 5 and 6 show the abbreviations of different equipment and data composition of IEC TC10 database respectively.

A total of two databases are used in this article. One for the implementation and evaluation of the individual and combined methods. Another for comparison with several methods of the literature.

3 | RESULTS AND DISCUSSION

3.1 | Metrics for model evaluation

Table 7 presents the metrics used to evaluate the performance of the proposed methods. These metrics include accuracy, sensitivity, precision, specificity, and F1 score. They are calculated based on the number of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) obtained from the confusion matrix as shown in Figure 5. The sensitivity refers to model's ability to discern the true fault. The precision returns the truly positive individual proportion within a population predicted as positive. The specificity refers to

TABLE 3 Distribution of the collected data.

References	Fault types D$_1$	D$_2$	PD	T$_3$	T$_2$	T$_1$	Total
[20]	127	141	55	90	65	114	592
[21]	4	3	2	2	3	3	17
[16]	0	1	0	7	14	1	23
[22]	0	0	0	0	32	27	59
[23]	0	7	1	5	5	0	18
[24]	14	3	19	50	8	0	94
[25]	3	4	3	8	6	4	28
[26]	3	2	3	5	2	3	18
Total	151	161	83	167	135	152	849

TABLE 4 The training and testing dataset's composition.

Dataset	Fault types D$_1$	D$_2$	PD	T$_3$	T$_2$	T$_1$	Total
Training	106	113	58	117	95	106	595
Testing	45	48	25	50	40	46	254

TABLE 5 Composition of validating dataset.

Abbreviations	Equipment	Number of samples
P	Power transformer without communication	OLTC 36
U	Power transformer with communication	OLTC 22
R	Reactor	32
S	/	7
I	Instrument transformer	12
B	Bushing	5
C	Cable	2

TABLE 6 Composition of IEC TC10 database.

	Fault types D$_1$	D$_2$	PD	T$_1$/T$_2$	T$_3$	Total
Number of samples	26	48	9	16	18	117

model's ability to discern the negative population [29]. The macro average metrics are used to evaluate the method on all class while the other metrics to evaluate the method on one class.

3.2 | Implementation results and discussion

The algorithm was designed in .m codes, and MATLAB software was used to carry out the implementation. The MATLAB codes of the proposed methods and datasets used are available online in a repository hosted in Github [30]. Table 8 and Figure 6 present the method and fault diagnostic accuracies obtained for the training and testing datasets respectively.

TABLE 7 Metrics used for model evaluation [28].

Metric	Formula
Accuracy	$\mathrm{Acc} = \frac{\sum_{i=1}^{N} \mathrm{TP}(C_i)}{\sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij}}$
Sensitivity of class	$\mathrm{TPR}(C_i) = \frac{\mathrm{TP}(C_i)}{\mathrm{TP}(C_i)+\mathrm{FN}(C_i)}$
Precision of class	$\mathrm{PPV}(C_i) = \frac{\mathrm{TP}(C_i)}{\mathrm{TP}(C_i)+\mathrm{FP}(C_i)}$
Specificity of class	$\mathrm{TNR}(C_i) = \frac{\mathrm{TN}(C_i)}{\mathrm{TN}(C_i)+\mathrm{FP}(C_i)}$
F1-score of class	$F1(C_i) = 2 \cdot \frac{\mathrm{TPR}(C_i)\cdot \mathrm{PPV}(C_i)}{\mathrm{TPR}(C_i)+\mathrm{PPV}(C_i)}$
Sensitivity macro average	$\mathrm{TPR(macro)} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{TPR}(C_i)$
Precision macro average	$\mathrm{PPV(macro)} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{PPV}(C_i)$
Specificity macro average	$\mathrm{TNR(macro)} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{TNR}(C_i)$
F1-score macro average	$\mathrm{F1(macro)} = 2 \cdot \frac{\mathrm{TPR(macro)}\cdot \mathrm{PPV(macro)}}{\mathrm{TPR(macro)}+\mathrm{PPV(macro)}}$

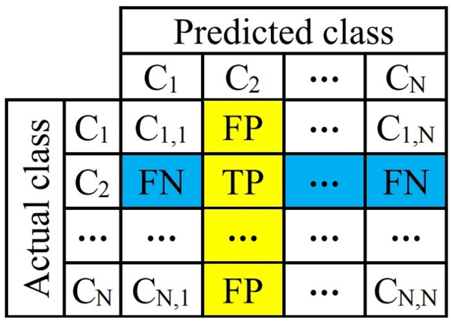

FIGURE 5 TP, FP, TN and FN identification from the confusion matrix of a multiclass classification problem [28]. FN, false negative; FP, false positive; TN, true negative; TP, true positive.

TABLE 8 Method diagnostic accuracies obtained for training and testing datasets.

Methods	Diagnosis accuracies (%) Training dataset	Testing dataset
Individual method 1	89.92	88.98
Individual method 2	87.90	87.01
Individual method 3	87.73	84.25
Individual method 4	90.08	87.40
Individual method 5	88.40	88.19
Individual method 6	88.74	87.80
Individual method 7	85.88	83.07
Combined technique	91.09	90.94

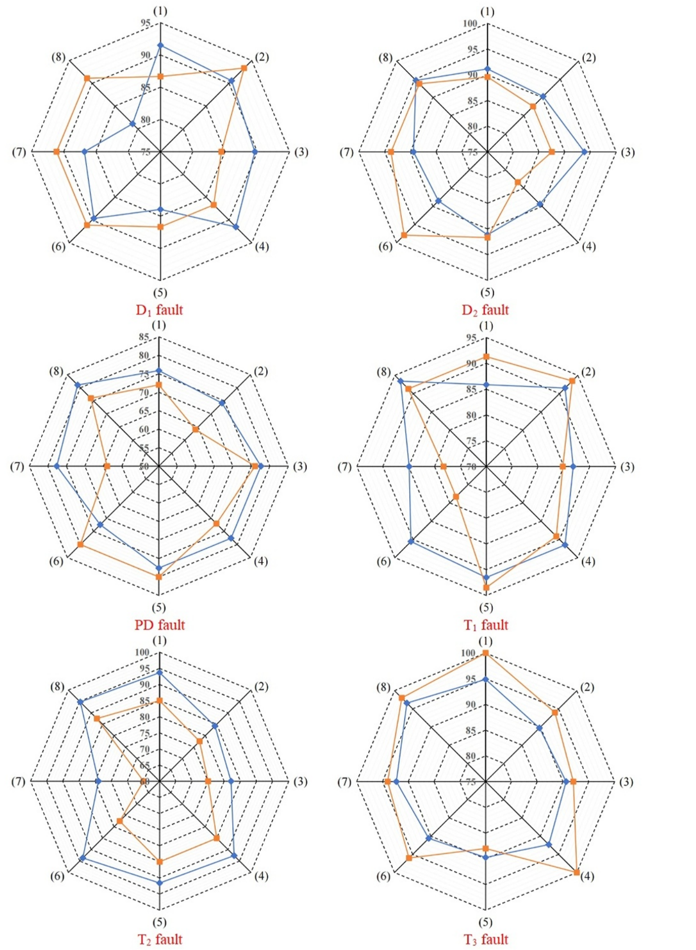

FIGURE 6 The diagnostic accuracies of the proposed methods obtained on the training and testing datasets according to the fault type. (1): Individual method 1. (2): Individual method 2. (3): Individual method 3. (4): Individual method 4. (5): Individual method 5. (6): Individual method 6. (7): Individual method 7. (8): Combine technique.

From the fault diagnostic outcomes shown in Figure 6, for D_1 fault, the individual method 2 has the best performance with a diagnostic accuracy of 93.33%. With a diagnostic accuracy of 97.92%, the individual method 6 performed the best for D_2 fault. For PD fault, the best performance is achieving by the individual methods 5 and 6. For T_1 fault, the individual methods 2 and 5 have the best performance with a diagnostic accuracy of 93.48%. For T_2 fault it's the combined technique that performed the best with a diagnostic accuracy of 87.50%. For T_3 fault, the best performance is achieving by the individual methods 1 and 4. From these results, the feature vector used for cluster formation impacts the performance of the returned diagnostic method. The feature input vector used has an impact on the quality of the clusters formed, and therefore on the performance of the human expert in separating the different fault types associated with the same cluster.

From the fault diagnostic outcomes shown in Table 8, among the individual methods, methods 1 and 4 with the diagnostic accuracies of 89.92% and 90.08% respectively have the best performance on the training dataset. With a diagnostic accuracy of 91.09%, the combined technique performed the best of all the proposed methods. Individual method 7 based on vector 7, which consists of the percentage ratios of the Gouda triangle method, is the least effective method with a diagnostic accuracy of 85.88%. The diagnostic accuracies obtained with the seven individual methods, on the testing dataset, are close to those obtained on the training dataset. Like on the training dataset, the methods 1 and 4 performed the best. With a diagnostic accuracy of 90.94%, the best performance on the testing dataset was achieved by combined technique.

Based on the diagnostic results obtained on the testing dataset, the statistical performance metrics of the proposed methods were calculated and are presented in Table 9. The results of the macro average metrics for evaluating the proposed methods across all classes are presented in Figure 7. Regarding the results in Figure 7, the combined technique outperforms the other methods. In fact, it presents the best results in terms of sensitivity, precision, specificity and F1 score. This can be explained by the complementary nature of the latter, which incorporates all the advantages of each individual method. In fact, the analysis of the results in Table 9 shows that no individual method performs well on all classes (faults) at once. For example, individual method 2 tends to identify and classify the D_2 fault well, while it performs less well for the PD fault.

3.3 | Sensitivity and significance difference analysis of the proposed methods

Sensitivity analysis is a valuable technique for assessing the impact of variations in input parameters or model configurations on diagnostic outcomes. In other words, it quantifies how the uncertainty in a model's results is related to the uncertainty in its input data. One-at-A-Time analysis, where one input parameter is changed while the others are held constant, is used in this study to perform the sensitivity analysis. In this experiment, the concentration of one of the gases is changed before the DGA sample is submitted to the diagnostic method for interpretation. This experiment is performed on all test data and the results obtained are summarised in Table 10. Table 11 shows the descriptive statistical results of sensitivity analysis. From these results, it can be concluded that individual methods 1, 4, and 5 are the least sensitive to variations in gas concentration values. However, this is not true for methods 2, 4, 6, and 7, as they are highly sensitive to these variations. This could be explained by the nature of the input feature vectors.

To carry out an analysis of the significant differences between the proposed methods, the Friedman test is applied to the diagnostic accuracies presented in Table 10. This test is a non-parametric test used to identify and analyse statistically significant differences between the proposed methods. The Friedman test was performed using two hypothetical statements, the null hypothesis (H0) and the alternative hypothesis (H1). The null hypothesis states that there is no difference between the proposed methods, implying that the feature input

TABLE 9 Proposed method performances.

		TP	FP	TN	FN	TPR (C_i) (%)	PPV(C_i) (%)	TNR (C_i) (%)	F1(C_i) (%)	Acc (%)
Individual method 1	PD	18	2	227	7	72.00	90.00	99.13	80.00	88.98
	D_1	39	5	204	6	86.67	88.64	97.61	87.64	
	D_2	43	5	201	5	89.58	89.58	97.57	89.58	
	T_1	42	7	201	4	91.30	85.71	96.63	88.42	
	T_2	34	1	213	6	85.00	97.14	99.53	90.67	
	T_3	50	8	196	0	100.00	86.21	96.08	92.59	
Individual method 2	PD	16	5	224	9	64.00	76.19	97.82	69.57	87.01
	D_1	42	7	202	3	93.33	85.71	96.65	89.36	
	D_2	42	2	204	6	87.50	95.45	99.03	91.30	
	T_1	43	11	197	3	93.48	79.63	94.71	86.00	
	T_2	31	1	213	9	77.50	96.88	99.53	86.11	
	T_3	47	7	197	3	94.00	87.04	96.57	90.38	
Individual method 3	PD	19	4	225	6	76.00	82.61	98.25	79.17	84.25
	D_1	38	7	202	7	84.44	84.44	96.65	84.44	
	D_2	42	5	201	6	87.50	89.36	97.57	88.42	
	T_1	39	7	201	7	84.78	84.78	96.63	84.78	
	T_2	30	4	210	10	75.00	88.24	98.13	81.08	
	T_3	46	13	191	4	92.00	77.97	93.63	84.40	
Individual method 4	PD	18	1	228	7	72.00	94.74	99.56	81.82	87.40
	D_1	39	12	197	6	86.67	76.47	94.26	81.25	
	D_2	40	5	201	8	83.33	88.89	97.57	86.02	
	T_1	41	5	203	5	89.13	89.13	97.60	89.13	
	T_2	34	5	209	6	85.00	87.18	97.66	86.08	
	T_3	50	4	200	0	100.00	92.59	98.04	96.15	
Individual method 5	PD	20	1	228	5	80.00	95.24	99.56	86.96	88.19
	D_1	39	5	204	6	86.67	88.64	97.61	87.64	
	D_2	44	10	196	4	91.67	81.48	95.15	86.27	
	T_1	43	4	204	3	93.48	91.49	98.08	92.47	
	T_2	34	2	212	6	85.00	94.44	99.07	89.47	
	T_3	44	8	196	6	88.00	84.62	96.08	86.27	
Individual method 6	PD	20	2	227	5	80.00	90.91	99.13	85.11	87.80
	D_1	41	3	206	4	91.11	93.18	98.56	92.13	
	D_2	47	3	203	1	97.92	94.00	98.54	95.92	
	T_1	36	7	201	10	78.26	83.72	96.63	80.90	
	T_2	31	7	207	9	77.50	81.58	96.73	79.49	
	T_3	48	9	195	2	96.00	84.21	95.59	89.72	
Individual method 7	PD	16	3	226	9	64.00	84.21	98.69	72.73	83.07
	D_1	41	5	204	4	91.11	89.13	97.61	90.11	
	D_2	45	4	202	3	93.75	91.84	98.06	92.78	
	T_1	36	9	199	10	78.26	80.00	95.67	79.12	
	T_2	26	8	206	14	65.00	76.47	96.26	70.27	
	T_3	47	14	190	3	94.00	77.05	93.14	84.68	

TABLE 9 (Continued)

		TP	FP	TN	FN	TPR (C_i) (%)	PPV(C_i) (%)	TNR (C_i) (%)	F1(C_i) (%)	Acc (%)
Combined technique	PD	19	2	227	6	76.00	90.48	99.13	82.61	90.94
	D_1	41	5	204	4	91.11	89.13	97.61	90.11	
	D_2	45	3	203	3	93.75	93.75	98.54	93.75	
	T_1	42	4	204	4	91.30	91.30	98.08	91.30	
	T_2	35	4	210	5	87.50	89.74	98.13	88.61	
	T_3	49	5	199	1	98.00	90.74	97.55	94.23	

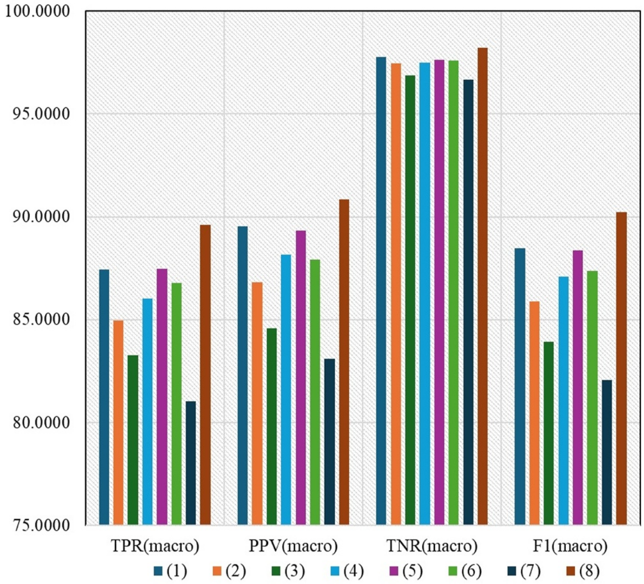

FIGURE 7 Proposed method performances across all classes. (1): Individual method 1. (2): Individual method 2. (3): Individual method 3. (4): Individual method 4. (5): Individual method 5. (6): Individual method 6. (7): Individual method 7. (8): Combine technique.

vectors used in the subset formation stage had no effect on the diagnostic model obtained. The alternative hypothesis suggests that there is a difference between the methods. The Friedman test results are shown in Tables 12 and 13. Table 12 shows the differences in the mean ranks of the proposed individual methods. Table 13 confirms this by rejecting the null hypothesis with a p-value of less than 0.05 and a chi-square value of greater than 3.841. This indicates that there was a significant difference between the proposed individual methods. This means that the different feature input vectors used provides crucial information that significantly impacts the quality of the clusters formed during subset formation stage. And consequently, on the human expert's ability to distinguish between faults in the same cluster.

3.4 | IEC TC10 database validation and comparison with current techniques

A comparison is made between nine methods from the literature [5, 7, 9, 10, 13, 31–34] and those proposed in this paper. The aim of this comparison is to assess the effectiveness of the proposed methods in comparison with those already in existence. Tables 14 and 15 show the diagnostic accuracies using 117 instances from the IEC TC10 databases. Table 14 presents the diagnostic accuracies by type of fault. The diagnostic accuracies of the methods based on vectors 1 and 4 were 97.44% and 98.29%, respectively, representing the highest accuracies obtained by the individual methods. The methods based on vectors 2, 3 and 7 are next in line, with a diagnostic accuracy of 96.58%. The method based on vector 5, with a diagnostic accuracy of 94.87%, is the least effective individual approach. These findings indicate that the individual methods proposed are more accurate than the Gouda triangle method (88.89%) [9], the Hyosun gas ratio method (88.03%) [31] and the three ratios technique (86.32%) [7]. The proposed combined technique exhibits the highest diagnostic accuracy compared to the individual techniques employed in its implementation and to the literature methods utilised for comparison. The diagnostic accuracy of this combined technique is 100.00%.

Diagnostic accuracies by equipment are presented in Table 15. According to these findings, the combined technique is the most effective, with a diagnostic accuracy of 100.00% for power transformers without communication on-load tap-changers. The individual methods follow, with diagnostic accuracies ranging from 91.67% to 97.22%. With diagnostic accuracies of 91.67% and 88.89%, respectively, the three ratios approach [7] and the Hyosun gas ratios method [31] come next. In the case of power transformers with communication on-load tap-changers, the Hyosun gas ratio method [31], the combined technique and individual methods 1, 2, 4 and 6 are the most effective, with a diagnostic accuracy of 100.00%. These diagnostic methods are followed by the individual methods based on vectors 3 and 7, the three-ratio method [7] and the Gouda triangle method [9], all of which have a diagnostic accuracy of 95.45%.

3.5 | Computational analysis

Implementation and execution times are further indicators of diagnostic methods. The implementation time considers the creation of the 120 subsets and the analysis of the subsets by the experts to implement the 120 sub-models. The execution time considers the time between the input of the gas

TABLE 10 Sensitivity analysis results.

Input variations		Model diagnostic accuracies (%) (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Without uncertainty		88.98	87.01	84.25	87.40	88.19	87.80	83.07	90.94
With uncertainty	−10% de H_2	85.43	72.05	72.05	85.43	83.46	86.61	79.13	86.22
	−5% de H_2	87.40	79.92	78.35	87.40	86.61	87.80	81.10	88.58
	+5% de H_2	87.80	79.13	80.31	87.01	87.01	87.40	81.50	87.40
	+10% de H_2	87.01	75.98	75.20	84.65	83.07	87.40	82.28	85.43
	−10% de CH_4	86.61	74.02	71.26	85.04	83.86	73.23	68.50	81.89
	−5% de CH_4	87.01	78.35	77.17	86.22	86.61	82.68	76.77	87.01
	+5% de CH_4	87.40	79.92	78.35	87.01	86.61	79.53	79.13	85.83
	+10% de CH_4	86.61	74.02	73.23	85.83	85.83	75.59	75.20	83.07
	−10% de C_2H_6	86.61	72.44	76.77	85.43	84.65	87.40	80.71	87.80
	−5% de C_2H_6	88.19	78.74	82.28	86.61	86.61	87.80	82.68	88.58
	+5% de C_2H_6	87.01	77.95	81.50	86.22	87.01	87.80	83.46	88.97
	+10% de C_2H_6	84.25	70.47	76.38	85.43	83.86	86.61	81.50	87.01
	−10% de C_2H_4	84.65	64.57	70.08	83.86	85.04	75.59	75.98	81.89
	−5% de C_2H_4	87.80	75.59	77.95	86.22	87.01	80.31	80.71	85.83
	+5% de C_2H_4	87.80	76.77	74.80	85.43	87.40	80.71	74.02	85.04
	+10% de C_2H_4	84.65	70.87	67.72	83.46	86.61	71.65	68.11	80.71
	−10% de C_2H_2	87.01	83.86	81.89	85.04	87.01	83.46	79.13	88.19
	−5% de C_2H_2	88.58	85.43	83.46	86.61	87.40	86.22	80.71	90.16
	+5% de C_2H_2	89.76	86.22	83.86	87.40	88.19	87.40	81.50	90.94
	+10% de C_2H_2	88.98	84.65	83.07	86.22	86.61	85.43	80.71	90.55

Note: (1): Individual method 1. (2): Individual method 2. (3): Individual method 3. (4): Individual method 4. (5): Individual method 5. (6): Individual method 6. (7): Individual method 7. (8): Combine technique.

TABLE 11 Descriptive statistical results of sensitivity analysis.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean	87.12	77.52	77.62	85.90	86.13	83.26	78.85	86.76
Std. deviation	1.46	5.85	4.86	1.11	1.52	5.34	4.34	3.00
Minimum	84.25	64.57	67.72	83.46	83.07	71.65	68.11	80.71
Maximum	89.76	87.01	84.25	87.40	88.19	87.80	83.46	90.94

Note: (1): Individual method 1. (2): Individual method 2. (3): Individual method 3. (4): Individual method 4. (5): Individual method 5. (6): Individual method 6. (7): Individual method 7. (8): Combine technique.

concentration values into the computer and the return of the diagnostic result. Table 16 shows the implementation and execution times for the proposed methods. These results were obtained with MATLAB 2021a on a personal computer with an Intel(R) Core(TM) i5-5200U CPU @ 2.20 GHz 2.20 GHz with 8 Gb of RAM. The observed implementation times can be attributed to the number of subsets created and the fact that the analysis of these subsets is carried out by a human expert. However, once implemented, the execution time is of the order of the second. Much of this execution time is spent reading the DGA data in the Excel file. In practice, this execution time has no negative impact on the maintenance process.

TABLE 12 Friedman's test mean rank results.

Method	Mean rank
Individual method 1	6.48
Individual method 2	2.26
Individual method 3	1.74
Individual method 4	5.05
Individual method 5	5.36
Individual method 6	4.88
Individual method 7	2.24

TABLE 13 Friedman's test Statistical results.

Method	Mean rank
N	7
Chi-square	94.9
df	6
p	<0.001

TABLE 14 Diagnostic accuracies by type of fault obtained by the proposed and literature methods, on the IEC TC10 database.

Diagnostic methods		Fault diagnostic accuracy (%) D_1	D_2	PD	T_1/T_2	T_3	Total
Existing methods	[7]	97.92	88.89	73.08	68.75	88.89	86.32
	[9]	97.92	100.00	88.46	56.25	88.89	88.89
	[5]	97.92	100.00	80.77	43.75	88.89	85.47
	[31]	80.77	97.92	77.78	75.00	88.89	88.03
	[32]	87.50	100.00	57.69	62.50	77.78	76.92
	[33]	77.08	88.89	53.85	18.75	55.56	61.54
	[34]	89.58	100.00	50.00	50.00	61.11	71.79
	[13]	91.67	77.78	53.85	37.50	77.78	72.65
	[10]	77.08	88.89	53.85	18.75	55.56	61.54
Proposed methods	Individual method 1	96.15	97.92	100.00	100.00	94.44	97.44
	Individual method 2	92.31	100.00	100.00	93.75	94.44	96.58
	Individual method 3	96.15	97.92	88.89	93.75	100.00	96.58
	Individual method 4	96.15	100.00	100.00	100.00	94.44	98.29
	Individual method 5	92.31	95.83	100.00	100.00	88.89	94.87
	Individual method 6	96.15	100.00	88.89	100.00	83.33	95.73
	Individual method 7	96.15	97.92	100.00	93.75	94.44	96.58
	Combined technique	100.00	100.00	100.00	100.00	100.00	100.00

TABLE 15 Diagnostic accuracies by equipment obtained by the proposed and literature methods, on the IEC TC10 database.

Diagnostic methods		Fault diagnostic accuracy (%) B	C	I	P	R	S	U	Empty	Total
Existing methods	[7]	20.00	100.00	91.67	91.67	84.38	71.43	95.45	100.00	86.32
	[9]	60.00	100.00	91.67	83.33	93.75	85.71	95.45	100.00	88.89
	[5]	40.00	100.00	91.67	83.33	90.63	71.43	90.91	100.00	85.47
	[31]	40.00	100.00	83.33	88.89	90.63	71.43	100.00	100.00	88.03
	[32]	20.00	100.00	91.67	77.78	87.50	42.86	72.73	100.00	76.92
	[33]	20.00	100.00	66.67	55.56	71.88	57.14	59.09	100.00	61.54
	[34]	20.00	100.00	83.33	75.00	78.13	42.86	68.18	100.00	71.79
	[13]	0.00	100.00	75.00	69.44	90.63	57.14	68.18	100.00	72.65
	[10]	60.00	50.00	91.67	77.78	84.36	/	72.73	/	61.54
Proposed methods	Individual method 1	100.00	100.00	100.00	91.67	96.88	100.00	100.00	100.00	97.44
	Individual method 2	60.00	100.00	100.00	97.22	100.00	85.71	100.00	100.00	96.58
	Individual method 3	80.00	100.00	91.67	97.22	100.00	100.00	95.45	100.00	96.58
	Individual method 4	80.00	100.00	100.00	97.22	100.00	100.00	100.00	100.00	98.29
	Individual method 5	80.00	100.00	100.00	97.22	100.00	100.00	81.82	100.00	94.87
	Individual method 6	80.00	100.00	91.67	91.67	100.00	100.00	100.00	100.00	95.73
	Individual method 7	100.00	100.00	100.00	94.44	100.00	85.71	95.45	100.00	96.58
	Combined technique	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

TABLE 16 Implementing' and execution's times.

Method	Implementing's time	Execution's time (s)
Individual method 1	1 day 16 h 55 min	3.60
Individual method 2	2 days 14 h 05 min	3.61
Individual method 3	2 days 01 h 15 min	3.61
Individual method 4	1 day 10 h 12 min	3.60
Individual method 5	1 day 07 h 12 min	3.62
Individual method 6	2 days 20 h 23 min	3.62
Individual method 7	2 days 04 h 45 min	3.61

4 | CONCLUSION

In this paper, eight DGA-based diagnostic methods have been proposed for faults diagnostic of power transformers. Among the proposed methods, seven are two-step individual methods based on clustering and subset analysis, and the last one is a combined technique that integrates the diagnostic results of individual methods into a unified diagnostic model. In the individual methods, the *k*-means algorithm is used in the first step for the formation of the clusters. In the second step, traditional gas ratios sub-models were proposed to separate the different fault types associated with the same cluster. Seven feature input vectors were used in the subset formation step corresponding to seven individual methods proposed. In the combined technique, to integrate the individual methods into the unified diagnostic model, an SVM classifier is trained with the diagnostic results of the individual methods as input features and the fault types as labels. In this paper, 966 DGA samples including the six primary IEC defects were utilised. The suggested procedures were applied and assessed using the first set of data, which consisted of 849 samples. Based on the training and testing dataset results, the combined technique performs best of all the proposed methods. Using the second batch of 117 samples, the performance and efficacy of the suggested methods were assessed and compared with those of the current methods. The validation and comparison results showed that, the combined technique depicts the best overall diagnostic accuracies. These findings indicate that the two-stage diagnostic approach has the potential to enhance the performance of conventional methods. Furthermore, the influence of the feature input vector on the efficacy of diagnostic models employing artificial intelligence tools was demonstrated. The proposed combined approach, which integrates multiple individual methods into a unified diagnostic system, is an attractive alternative for enhancing subsequent diagnostic methods. In future research, a hybrid approach based on multisource information fusion can be explored to improve power transformers' fault diagnostic.

AUTHOR CONTRIBUTIONS

Arnaud Nanfak: Conceptualization; formal analysis; methodology; software; writing – original draft; writing – review & editing. **Abdelmoumene Hechifa**: Formal analysis; methodology; writing – original draft; writing – review & editing. **Samuel Eke**: formal analysis; supervision. **Abdelaziz Lakehal**: Formal analysis; supervision; writing – review & editing. **Charles Hubert Kom**: Formal analysis; supervision. **Sherif S. M. Ghoneim**: Formal analysis; supervision; writing – review & editing.

ACKNOWLEDGEMENTS

None.

CONFLICT OF INTEREST STATEMENT

None.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in reference number [30].

ORCID

Arnaud Nanfak https://orcid.org/0000-0002-6432-3365

REFERENCES

1. Taha, I.B.M., Mansour, D.E.A., Ibrahim, S.: Power transformer fault diagnosis based on dga using a convolutional neural network with noise in measurements. IEEE Access 9, 111162–111170 (2021). https://doi.org/10.1109/access.2021.3102415
2. Demirci, M., Gözde, H., Taplamacioglu, M.C.: Improvement of power transformer fault diagnostic by using sequential Kalman filter sensor fusion. Int. J. Electr. Power Energy Syst. 149, 109038 (2023). https://doi.org/10.1016/j.ijepes.2023.109038
3. Ekojono, et al.: Investigation on machine learning algorithms to support transformer dissolved gas analysis fault identification. Electr. Eng. 104(5), 3037–3047 (2022). https://doi.org/10.1007/s00202-022-01532-5
4. Zheng, H., Shioya, R.: A comparison between artificial intelligence method and standard diagnostic methods for power transformer dissolved gas analysis using two public databases. IEEJ Trans. Electr. Electron. Eng. 15(9), 1305–1311 (2020). https://doi.org/10.1002/tee.23197
5. IEEE Std C.57.104-2019: IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers (2019). https://doi.org/10.1109/IEEESTD.2019.8890040
6. IEC Publication 60599: Mineral Oil-Filled Electrical Equipment in Service - Guidance on the Interpretation of Dissolved and Free Gases Analysis (2015)
7. Gouda, O.E., El-Hoshy, S.H., E.L.-Tamaly, H.H.: Proposed three ratios technique for the interpretation of mineral oil transformers based dissolved gas analysis. IET Gener. Transm. Distrib. 12(11), 2650–2661 (2018). https://doi.org/10.1049/iet-gtd.2017.1927
8. Mansour, D.-E.A.: Development of a new graphical technique for dissolved gas analysis in power transformers based on the five combustible gases. IEEE Trans. Dielectr. Electr. Insul. 22(5), 2507–2512 (2015). https://doi.org/10.1109/tdei.2015.004999
9. Gouda, O.E., El-Hoshy, S.H., E.L.-Tamaly, H.H.: Condition assessment of power transformers based on dissolved gas analysis. IET Gener. Transm. Distrib. 13(12), 2299–2310 (2019). https://doi.org/10.1049/iet-gtd.2018.6168
10. Misbahulmunir, S., Ramachandaramurthy, V.K., Thayoob, Y.H.M.D.: Improved self-organizing map clustering of power transformer dissolved gas analysis using inputs pre-processing. IEEE Access 8, 71798–71811 (2020). https://doi.org/10.1109/access.2020.2986726
11. Eke, S., et al.: Transformer condition assessment using fuzzy C-means clustering techniques. IEEE Electr. Insul. Mag. 35(2), 47–55 (2019). https://doi.org/10.1109/mei.2019.8636165
12. Nanfak, A., Kom, C.H., Eke, S.: Hybrid method for power transformers faults diagnostic based on ensemble bagged tree classification and training subsets using Rogers and Gouda ratios. Int. J. Intell. Eng. Syst. 15(5) (2022)

13. Ghoneim, S.S., Taha, I.B., Elkalashy, N.I.: Integrated ANN-based pro-active fault diagnostic scheme for power transformers using dissolved gas analysis. IEEE Trans. Dielectr. Electr. Insul. 23(3), 1838–1845 (2016). https://doi.org/10.1109/tdei.2016.005301
14. Wu, X., He, Y., Duan, J.: A deep parallel diagnostic method for transformer dissolved gas analysis. Appl. Sci. 10(4), 1329 (2020). https://doi.org/10.3390/app10041329
15. Islam, M.M., Lee, G., Hettiwatte, S.N.: A nearest neighbour clustering approach for incipient fault diagnostic of power transformers. Electr. Eng. 99(3), 1109–1119 (2017). https://doi.org/10.1007/s00202-016-0481-3
16. Odongo, G., Musabe, R., Hanyurwimfura, D.: A multinomial DGA classifier for incipient fault detection in oil-impregnated power transformers. Algorithms 14(4), 128 (2021). https://doi.org/10.3390/a14040128
17. Hechifa, A., et al.: Improved intelligent methods for power transformer fault diagnosis based on tree ensemble learning and multiple feature vector analysis. Electr. Eng. 106(3), 2575–2594 (2023) (In Press). https://doi.org/10.1007/s00202-023-02084-y
18. Davidenko, I.V., Ovchinnikov, K.V.: Identification of transformer defects via analyzing gases dissolved in oil. Russ. Electr. Eng. 90(4), 338–343 (2019). https://doi.org/10.3103/s1068371219040035
19. Conservation and control of oil-insulated components by diagnostic of gas in oil. Electr. Coop. Res. Assoc. N. 36(1) (1980)
20. Nanfak, A., et al.: Interpreting dissolved gases in transformer oil: a new method based on the analysis of labelled fault data. IET Gener. Transm. Distrib. 15(21), 3032–3047 (2021)
21. Islam, M.M., Lee, G., Hettiwatte, S.N.: Application of Parzen Window estimation for incipient fault diagnostic in power Transformers. High Volt. 3(4), 303–309 (2018)
22. Shang, H., et al.: A novel fault diagnostic method for power transformer based on dissolved gas analysis using hypersphere multiclass support vector machine and improved D–S evidence theory. Energies 12(20), 4017 (2019)
23. Mharakurwa, E.T., Nyakoe, G.N., Akumu, A.O.: Power transformer fault severity estimation based on dissolved gas analysis and energy of fault formation technique. J. Electr. Comp. Eng., 10 (2019). Article ID 9674054
24. Zhang, X., et al.: Research on transformer fault diagnostic: based on improved firefly algorithm optimized LPboost–classification and regression tree. IET Gener. Transm. Distrib. 15(20), 2926–2942 (2021)
25. Zhang, Y., et al.: A new support vector machine model based on improved imperialist competitive algorithm for fault diagnostic of oil-immersed transformers. J. Electr. Eng. Tech. 12(2), 830–839 (2017)
26. Ganyun, L.V., et al.: Fault diagnostic of power transformer based on multi-layer SVM classifier. Elec. Power Syst. Res. 74(1), 9–15 (2005)
27. Duval, M., de Pabla, A.: Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC10 databases. IEEE Electr. Insul. Mag. 17(2), 31–41 (2001)
28. Markoulidakis, I., et al.: Multiclass confusion matrix reduction method and its application on net promoter score classification problem. Technologies 9(4), 81 (2021)
29. Hendel, M., et al.: Using genetic direct M-SVM model improved by Kohonen map and dempster-shafer theory to enhance power transformers diagnostic. Sustainability 15(21), 15453 (2023)
30. Nanfak, A.: DGA MATLAB Codes GitHub Repository. https://github.com/NanaudKmer/IET_SVM_Combined_Technique_DGA_Method
31. Kim, S.W., et al.: New methods of DGA diagnostic using IEC TC 10 and related databases Part 1: application of gas-ratio combinations. IEEE Trans. Dielectr. Electr. Insul. 20(2), 685–690 (2013). https://doi.org/10.1109/tdei.2013.6508773
32. Ward, S.A., et al.: Towards precise interpretation of oil transformers via novel combined techniques based on DGA and partial discharge sensors. Sensors 21(6), 2223 (2021). https://doi.org/10.3390/s21062223
33. Badawi, M., et al.: Reliable estimation for health index of transformer oil based on novel combined predictive maintenance techniques. IEEE Access 10, 25954–25972 (2022). https://doi.org/10.1109/access.2022.3156102
34. Taha, I.B., et al.: Conditional probability-based interpretation of dissolved gas analysis for transformer incipient faults. IET Gener. Transm. Distrib. 11(4), 943–951 (2017). https://doi.org/10.1049/iet-gtd.2016.0886

How to cite this article: Nanfak, A., et al.: A combined technique for power transformer fault diagnosis based on k-means clustering and support vector machine. IET Nanodielectr. 1–13 (2024). https://doi.org/10.1049/nde2.12088