Hybrid DGA method for power transformer faults diagnosis based on evolutionary *k*-means clustering and dissolved gas subsets analysis

Arnaud NANFAK [1]*, Samuel EKE [1], Fethi MEGHNEFI [2], Issouf FOFANA [2], Gildas Martial NGALEU [1] Charles Hubert KOM [1]

[1] Laboratory of Energy, Materials, Modelling and Methods, National Higher Polytechnic School of Douala, University of Douala PO Box 2701 Douala – Cameroon
[2] Research Chair on the Aging of Power Network Infrastructure, University of Quebec at Chicoutimi, Chicoutimi, QC G7H 2B1, Canada
*nanfak.arnaud@yahoo.fr

***Abstract*—Considered as the heart of electrical power transmission and distribution networks, power transformers are essential part of the electricity transmission grid. Among the condition monitoring and fault diagnosis tools for these machines, dissolved gas analysis (DGA) has proven its effectiveness in their early detection and classification of faults. Up to date, many methods have been proposed in the literature for the interpretation of DGA data, classified into traditional and intelligent methods. This paper proposes a two-steps hybrid method, which uses the strengths of both methods. The approach uses the evolutionary k-means clustering algorithm based on the genetic algorithm for subset formation and subset analysis by human expertise. In the diagnostic procedure, to determine the condition of a sample, the subset to which it belongs is first identified and then the corresponding diagnostic sub-model is applied. The proposed method has been implemented with 595 DGA data, tested on 254 DGA data and validated on the International Electrotechnical Commission (IEC) TC10 database. Their performances were evaluated and compared with existing traditional, intelligent and hybrid methods. From the results obtained with the IEC TC10 database, the newly proposed approach depicts the best overall diagnosis accuracies. Indeed, the best performance is achieved with the proposed method compared to other models in the literature, with diagnostic accuracy of 98.29 compared to 88.89% of the Gouda triangle method, to 88.03% of the Hyosun Corporation gas ratio method or to 86.32% of the three ratios technique.**

***Index Terms*-Dissolved gas analysis, Evolutionary clustering, Fault diagnosis, Power transformer, Subset analysis.**

I. INTRODUCTION

The early and accurate diagnosis of faults in power transformers is a key factor in ensuring the efficient and safe operation of the power system. Among tools available in literature to achieve this goal, dissolved gas analysis (DGA) is a technique widely used by power transformer' maintenance professionals. DGA is a non-invasive monitoring technique that provides information on the condition of the insulation system in particular and the internal parts in general [1].

Several DGA-based methods are proposed in the literature for power transformers faults diagnosis and can be classified in two main categories: traditional and intelligent methods [2]. Traditional DGA-based methods are methods in which the process of interpreting fault-related gas concentrations depends on the experience of the expert rather than on mathematical tools or formulations. In these methods, experts produce rules relating absolute concentrations, concentration ratios and/or percentages of gases to the various faults. Many traditional methods have been proposed to interpret DGA data such as IEEE key gas method [3], Doernenburg ratios method [3], Rogers Ratios Method [3], Duval Triangle method [3], IEC 60599 ratios method [4], HYOSUN Corporation ratios method [5], three ratios technique [6] or Gouda triangle method [7]. In addition to traditional DGA-based methods, intelligent DGA-based methods rely on artificial intelligence tools to interpret DGA data. Several intelligent DGA-based methods are proposed in literature for this purpose. These methods are based, among others, on artificial neural network (ANN) [8], fuzzy logic [9], supervised [10], unsupervised [11], or ensemble [12] machine learning.

Both traditional and intelligent DGA-based method have strengths and weaknesses. Traditional methods are simple, easy to understand and implement. However, they have some drawbacks in terms of precision and uncertainty. In addition, these methods also have the disadvantages of leading to inconclusive assessments of fault severity, or in the extreme case, misidentification [13]. On the other hand, intelligent methods have relatively high fault diagnosis accuracies and improve the efficiency of DGA. However, these methods are generally complicated and their results depend on the parameters of the artificial intelligence algorithm and feature input vector used. In addition, the research documented in these publications is difficult to replicate [14]. Therefore, intelligent methods are not practically implemented over as wide a range as traditional methods [15]. In order to combine the advantages of both approaches to improve the fault diagnosis of power transformers, this paper proposes a hybrid method based on evolutionary clustering and dissolved gas subset analysis. In this method, evolutionary *k*-means clustering algorithm (k-MCA) based on genetic algorithm

(GA) is used for subsets formation. As a cluster may contain one or more kinds of faults, for each subset, experts produce a sub-model based on rules relating concentration ratios of H_2, CH_4, C_2H_6, C_2H_4 and C_2H_2 to the various faults.

Fault diagnosis methods for power transformers based on subset analysis have already been proposed in the literature [16]–[18]. In these papers, to identify the state of a sample, the subset to which it belongs is first identified and then the corresponding diagnostic sub-model is applied. In [16], for a given sample, the corresponding subset is determined using a set of rules based on combinations of the relative proportions of the different fault-related gases. The subset determined gives an idea of the potential faults of the sample. Subsequently, gas ratios are used to discriminate between the potential faults and the "real" fault. In [17], subsets are created by grouping samples with the same combination of maximum and minimum concentration(s) of the different fault-related gases. The fault prediction of a new sample is performed using the sub-model corresponding to the subset to which it belongs. In [18], subsets are created using *k*-MCA. Then, using the *k*-nearest-neighbor algorithm, the three closest clusters to an unknown sample are identified. And based on the characteristics of the clusters (percentages of the different faults that constitute the clusters) and the distance weighting factors, the fault is determined by voting between the faults of the 3 subsets. The diagnostic method proposed in this paper is based on these works. It was carried on using 595 samples dataset, tested on 254 samples dataset. The IEC TC10 database will be used for validation and the results obtained are compared with those of the following diagnostic methods:

- Traditional DGA-methods: modified Rogers' four ratios method [19], modified IEC ratios method [19], IEC 60599 method, clustering method [16], three ratios technique, Gouda triangle method, Duval triangle method, HYOSUN Corporation ratios method, and combined technique N°1 [20];
- Intelligent DGA-methods: Self-organizing map clusters method [21], conditional probability method [22], CSUS ANN method [8];
- Hybrid DGA-methods: combined techniques 2, 3 and 4 [20] and combined technique [23].

The remaining of this paper is organized as follows: The transformer fault types detectable by DGA and the analysis of dissolved gas are given in section 2. The principle and the flow chart of proposed method and the evolutionary *k*- MCA used in this paper are presented in section 3. In section 4, the performance and effectiveness of the proposed hybrid method are evaluated and compared with others methods in the literature using IEC TC10 database. The section 5 concludes the paper.

II. BACKGROUND AND PRINCIPLE OF DISSOLVED GAS ANALYSIS

Faults in power transformers due to deterioration of their insulation system (oil and paper) are grouped into two main categories, namely electrical faults and thermal faults. Based on IEC 60599, the two main types of faults can, according to their severity, be divided into 6 types of faults, as summarized in Table I.

TABLE I
FAULT CLASSIFICATION ACCORDING TO IEC 60599 AND IEEE C57.104 STANDARD

acronyms	Faults
PD	Partial discharge
D_1	Low-energy discharge
D_2	High-energy discharge
T_1	Thermal fault, $T < 300°C$
T_2	Thermal fault, $300°C < T < 700°C$
T_3	Thermal fault, $T > 700°C$

TABLE II
GAS GENERATED ACCORDING TO THE TYPE OF TRANSFORMER FAULT [24]

Fault type	**Major gas (es)**	**Minor gas (es)**
PD	H_2, CH_4, CO	C_2H_6, C_2H_2, CO_2
D_1	H_2, C_2H_2	/
D_2	H_2, C_2H_2, CO, CO_2	CH_4, C_2H_4, C_2H_6
T_1	CH_4, C_2H_6, CO, CO_2	H_2, C_2H_4
T_2	C_2H_4, CH_4	H_2
T_3	C_2H_4	H_2, C_2H_6

Depending on the type of fault and its location, different fault-related gases can be produced. Hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), propane (C_3H_8) and propylene (C_3H_6) result from faults (electrical and thermal) occurring in the transformer oil [18], [24]. Through oxidation or hydrolysis, the oil molecules degrade generating these combustible gases. When cellulose insulation is involved in the occurrence of faults, carbon monoxide (CO) and carbon dioxide (CO2) are generated. These gases indicate a thermal fault. Other gases such as oxygen (O2) and nitrogen (N2) are also produced [24]. Table II summarizes the main gases produced according to the type of transformer faults.

The nature of the gases formed and their relative proportions provide information on the incipient fault, its intensity and the type of materials affected [3], [4]. Each fault has a distinctive signature in terms of the quantity and combination of different gases associated with the fault. In addition, the particular combination of gases generated depends on the temperature level and/or the energy produced by the fault [3], [4]. Figure 1 shows the influence of temperature on the production of fault-related gases. The acceptable limits of the concentrations of the various fault-related gases make it possible to distinguish between normal and abnormal operating conditions and constitute an alarm signal that should trigger an in-depth analysis by the DGA's diagnostic methods.

III. METHODOLOGY

In this section, the principle and the flow chart of hybrid DGA-method for transformers fault diagnosis are presented. In addition, *k*-MCA and evolutionary algorithms used in this paper are presented.

A. Principle of the method

In this paper, a hybrid DGA-method based on evolutionary clustering and dissolved gas subset analysis is proposed. There are two steps in this hybrid method: subsets formation and subsets analysis. In subset formation step, a number of clusters

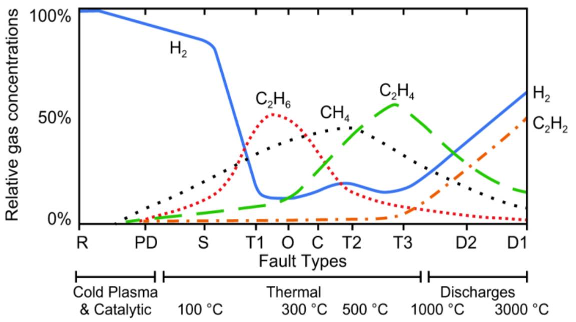

Fig. 1. Comparative proportion of dissolved gas concentrations in mineral oil as a function of temperature and fault type [3].
R: Catalytic reaction; O: Overheating, T <250◦C without carbonization of paper; S: Stray gassing, T <200◦C; C: Possible paper carbonization

is generated using evolutionary *k*-MCA. After clustering, as a cluster may contain one or more kinds of faults, in subsets analysis, a traditional diagnosis sub-models is proposed by human experts to separate the different faults-related to the subset. These sub-models are based on gas ratios approach. Fifteen gas ratios involving the five main hydrocarbon gases namely H_2, CH_4, C_2H_6, C_2H_4, and C_2H_2 are used. The final diagnostic model is obtained by combining the different sub-models. Figure 2 presents the schematic view of the approach used to implement the proposed method. The ratios used in subset analysis step to discriminate between faults in the same cluster are given in Table III. These include Roger's ratios (R_6, R_{12} and R_{13}), Gouda's ratios (R_8 and R_{14}), Duval's ratios (R_9 to R_{11}), Nanfak's ratios (R_1 to R_5) and others.

B. Formation of subsets

Cluster formation is the first step in the implementation of the proposed method. It is done using *k*-MCA.

1) Clustering problem and data pre-processing: Clustering is an unsupervised learning process that aims to partition an unlabeled dataset into groups called clusters, based on similarities between the data. The problem of clustering can be summarized as follows [25]:

Let $X = \{x_1,...,x_n\}$ be a set of n data samples, where each sample x_i , $i = 1,...,n$ is an *m*-dimensional feature vector. A clustering of X is a collection $C = \{C_1,...,C_k\}$ of k non-overlapping subsets of X such that:

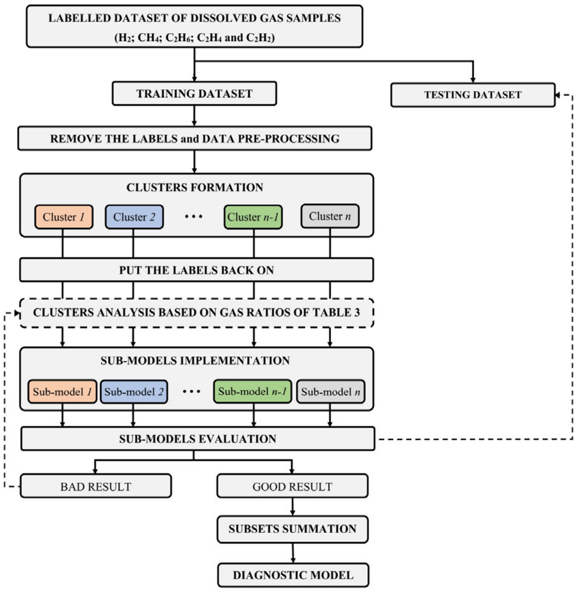

Fig. 2. Schematic view of the approach used for the implementation of the proposed hybrid method

$$\begin{cases} \{C_1 \cup ... \cup C_k\} = X, \quad \{C_i \cap C_j\} = \varnothing \quad \forall i, j \in \{1,2,...,k\}, \quad i \neq j \\ C_i \neq \varnothing \quad \forall i, j \in \{1,2,...,k\} \end{cases} \tag{1}$$

For subsets formation, the training data are pre-processed. The values of the gas concentrations are replaced by their relative proportions in the sample, obtained using (2):

$$p_i = \frac{C_i}{\sum_{j=1}^{5} C_j} \tag{2}$$

Where C_1 to C_5 are the concentrations (in ppm) of H_2, CH_4, C_2H_6, C_2H_4 and C_2H_6 respectively.

2) K-means clustering algorithm: The *k*-MCA is a partitioning-based clustering technique that groups a data set into k clusters by optimizing a criteria function. For a dataset $X = \{x_1,...,x_n\}$, the principle of this algorithm is to find the collection $C = \{C_1,...,C_k\}$ of k non-overlapping subsets of X which minimizes the total intra-cluster variance also known as the sum of squared errors (SSE), defined as follows [26]:

$$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n} \omega_{ij} \left(d\left(x_i, m_j\right)\right)^2 \tag{3}$$

TABLE III
GAS RATIOS USED IN THE SUBSET ANALYSIS STEP

Ratio	Ref.	Expression	Ratio	Ref.	Expression	Ratio	Ref.	Expression
R_1	[17]	$(CH_4 + C_2H_6)$/THHG	R_6	[3]	C_2H_2/C_2H_4	R_{11}	[3]	$C_2H_2/(CH_4 + C_2H_4 + C_2H_2)$
R_2	[17]	$(CH_4 + C_2H_4)$/THHG	R_7	/	$(C_2H_6 + C_2H_4)/(H_2 + CH_4 + C_2H_2)$	R_{12}	[3]	CH_4/H_2
R_3	[17]	$C_2H_6/(CH_4 + C_2H_4)$	R_8	[6]	$(CH_4 + C_2H_2)/C_2H_4$	R_{13}	[3]	C_2H_4/C_2H_6
R_4	[17]	$(H_2 + CH_4)$/THHG	R_9	[3]	$CH_4/(CH_4 + C_2H_4 + C_2H_2)$	R_{14}	[6]	$(C_2H_6 + C_2H_4)/(H_2 + C_2H_2)$
R_5	[17]	$(C_2H_4 + C_2H_2)$/THHG	R_{10}	[3]	$C_2H_4/(CH_4 + C_2H_4 + C_2H_2)$	R_{15}	/	$(C_2H_6 + C_2H_2)/C_2H_4$

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

Where m_j is the centroid of the class C_j, $d\left(x_i, m_j\right)$ the Euclidean distance between m_j and x_i, and:

$$\omega_{ij} = \begin{cases} 1 & \text{if } x_i \in C_j \\ 0 & else \end{cases} \tag{4}$$

To do this, the *k*-MCA has 3 main steps [25]:

- Step 1: Define the value of *k* and initialize the clusters by randomly assigning *k* points as centroids of the *k* clusters;
- Step 2: Assign each point to the cluster that is closest to the centroid;
- Step 3: Update the centroids based on the assigned data using equation (3) given by (5):

$$m_j = \frac{\sum_{i=1}^{n} \omega_{ij} x_i}{\sum_{i=1}^{n} \omega_{ij}} \tag{5}$$

Steps 2 and 3 are repeated until no data changes its cluster membership, or the criterion function does not improve during a number of iterations. Algorithm 1 below presents a pseudo-code of the *k*-means algorithm [27].

```
Algorithm 1 : k-means algorithm
Input: k, number of clusters; $X$, a data set of $n$ samples
Output: A set of k clusters
Random selection of k initial cluster centers $m_j$ with $j = 1,...,k$
Repeat
    For each sample $x_i$ in $X$ Do
        Determine the distance between $x_i$ and the different centroids $m_j$
        Assign $x_i$ to the nearest cluster
    End For
    Update the centroids based on assigned data
Until cluster centers are stable (Stop-iteration criteria satisfied)
Return clustering result
```

3) GA-based evolutionary (k-MCA): GA is an evolutionary algorithm inspired by Darwinian evolution and genetics [26]. It is based on genetic operators such as natural selection, crossover and mutation, which at each iteration produce a new population from the current population. Based on their fitness, the selection operator selects a part of the current population for the next iteration. The pseudo-code of the GA- based evolutionary *k*-MCA is presented in Algorithm 2 [27]. The parameters of this algorithm are given in Table IV. In this GA-based evolutionary k-MCA, a population of individuals containing the candidate centroids is initially created and each individual is evaluated by calculating its SSE. It is the initialization step. For the initialization, feature vectors are randomly selected from the dataset to constitute the initial population. After evaluating the individuals in the population, the population is ranked to determine and mark the *nPop*/2 best individuals. The population is iteratively refined by selecting the parents from the *nPop*/2 best individuals, applying the crossover operator for the generation of offspring, applying the mutation operator to the offspring obtained for the generation of mutants, re-evaluating the merged population consisting of the current population, the offspring and the mutants, and updating the population by natural selection by selecting the *nPop* best individuals from the merged population for the next iteration. The optimal cluster collection is obtained by the candidate centroids of the best individual of the last iteration.

TABLE IV
PARAMETERS OF GA-BASED *k*-MCA

Parameter	Symbol	Value
Number of clusters	k	120
Population size	*nPop*	70
Crossover percentage	*pc*	0.92
Mutation percentage	*pm*	0.30
Mutation rate	μ	0.02
Selection pressure	β	8

IV. RESULTS AND DISCUSSION

A. Data collection

To implement and test the diagnostic methods proposed in this paper, 849 dissolved gas samples with known faults were collected from credible sources in the literature. These DGA data are composed of 6 main types of faults. Table V shows the number of samples by fault type and by reference. The collected data were randomly divided, with a training-test ratio of 70:30, into two datasets, one training and one testing, as shown in Table VI. The training dataset is used to implement of the diagnostic models. The testing dataset is used to evaluate the observations made on the training dataset.

B. Implementation results and discussion

For the implementation 120 clusters were formed. The implementation was performed using MATLAB software and the algorithm was programmed in .m codes. The MATLAB codes are available online in a repository hosted in Github [34] and the pseudo-code of the proposed method is presented in the appendix. The diagnostic accuracies obtained are presented

TABLE V
DISTRIBUTION OF COLLECTED DATA

	Fault types						
Ref.	PD	D_1	D_2	T_1	T_2	T_3	Total
[13]	0	0	1	1	14	7	23
[17]	55	127	141	114	65	90	592
[28]	1	0	7	0	5	5	18
[29]	2	4	3	3	3	2	17
[30]	0	0	0	27	32	0	59
[31]	19	14	3	0	8	50	94
[32]	3	3	4	4	6	8	28
[33]	3	3	2	3	2	5	18
Total	83	151	161	152	135	167	849

Algorithm 2: GA-based evolutionary *k*-means clustering algorithm

```
Input: k, number of clusters; $X$ , a data set of $n$ samples; nPop, population size; crossover percentage; mutation percentage; mutation rate; selection pressure
Output: A set of k clusters
Random selection of k initial cluster centers $m_j$ with $j = 1,...,k$
Initialization of the population pop
Evaluation of each individual in the population pop
Repeat
    Select the nPop/2 best individuals
    Select parents from the nPop/2 best individuals: roulette wheel selection
    Apply crossover: Arithmetic crossover
    Evaluate the offspring
    Select an offspring: roulette wheel selection
    Mutate the genes of offspring (Mutation)
    Evaluate the mutants
    Create a merged population of the current population with the generated offspring and mutants
    Rank the individuals of the merged population by fitness
    Select the nPop best individuals for the next iteration
Until Stop-iteration criteria satisfied
Return clustering result
```

TABLE VI
COMPOSITION OF TRAINING AND TESTING DATASET

	Fault types						
Dataset	PD	D_1	D_2	T_1	T_2	T_3	Total
Training	58	106	113	106	95	117	595
Testing	25	45	48	46	40	50	254

in Table VII and Table VIII according to fault type and fault severity respectively.

The results presented in Table VII and Table VIII show that diagnostic accuracies of 96.91% and 90.44% were obtained in terms of fault type and fault severity, respectively, on all samples in the training dataset. This means that 614 samples out of the 680 in the database were correctly diagnosed. The observations made on the training dataset to build the different sub-models on the one hand and the final model on the other hand were evaluated on the testing dataset. On the latter, diagnostic accuracies of 96.45% and 88.17% were obtained in terms of fault type and fault severity, respectively.

TABLE VII
THE DIAGNOSIS ACCURACIES OF THE PROPOSED METHOD ACCORDING TO FAULT TYPE

	Fault diagnostic accuracy (%)			
	P	D	T	Total
Training dataset	76.56	97.59	98.91	96.91
Testing dataset	76.56	97.59	98.91	96.91
Total	88.17	96.83	98.89	96.45

TABLE VIII
THE DIAGNOSIS ACCURACIES OF THE PROPOSED METHOD ACCORDING TO FAULT SEVERITY

	Fault diagnostic accuracy (%)						
Dataset	PD	D_1	D_2	T_1	T_2	T_3	Total
Training	76.56	90.76	91.54	88.62	91.74	96.30	90.44
Testing	76.56	90.76	91.54	88.62	91.74	96.30	90.44
Total	75.00	93.55	87.50	80.00	88.89	96.97	88.17

C. Validation and comparison with other methods using IEC TC10 database

In order to validate the effectiveness of fault diagnosis models proposed, the IEC TC10 database is used. In this database, 117 DGA labelled of various equipment in service are provided [35]. The faults of this database were identified by visual inspection on several equipment including power transformer without communication on-load tap changers (P), power transformers with communication on-load tap changers (U), reactors (R), instrument transformers (I), bushings (B) and cables (C). To evaluate the performance of proposed methods, existing DGA-based methods of literature including Traditional, intelligent and hybrid methods are used for comparison. The diagnostic accuracies obtained with 117 cases of IEC TC10 databases are presented in Table IX. Diagnostic accuracies of 98.29%, was achieved by the proposed hybrid method. These results are higher than the 88.89% of the Gouda tringle method, 88.03% of the Hyosun Corporation ratios method or 86.32% of the three ratios techniques.

Table X shows the diagnostic accuracies per equipment obtained by the different methods. Based on these results, for power transformers without on-load tap changers, the proposed method has the best performance with diagnostic accuracy of 94.44% following to three ratios techniques and Hyosun corporation ratios method with diagnostic accuracies of 91.67% and 88.89% respectively. For power transformers with communication on-load tap changers, the proposed method and Hyosun corporation ratios method have the best performance with diagnostic accuracy of 100.00% following to three ratios techniques and Gouda tringle method with diagnostic accuracies of 95.45%. The same performance was achieved on the other equipment. These results highlight the impact of the subset approach in the data mining and fault signature identification. This approach allows a microscopic study of the labeled database. The improved performance of the performance of the proposed diagnostic methods. Moreover, the use of unsupervised machine learning for the creation of subsets improves their quality, evaluated from the expert's ability to distinguish faults within the same group.

V. Conclusion

This paper proposes a new hybrid method based on evolutionary clustering and dissolved gas subset analysis. The proposed method operates in two steps and performs to diagnose the 6 main IEC faults. In the first step, the DGA data are grouped into cluster using evolutionary *k*-means clustering algorithm using genetic algorithm. Then, in second step, after clustering, a traditional diagnosis sub-models are proposed by human experts to separate the different faults-related to the subsets. The gas ratios of fault-related gases including H_2, CH_4, C_2H_6, C_2H_4, and C_2H_2 are used to implement the sub-models. A total of 966 labelled samples covering six fault types were used in this paper. The first group of 849 samples were used to implement and evaluate the proposed diagnostic model. The validate results show that the best performance was achieved with the proposed hybrid method compared to existing methods in the literature. The diagnostic accuracies of 98.29% was obtained by the proposed hybrid method. These accuracies are higher than 88.89% of the Gouda tringle method, 88.03% of the Hyosun Corporation ratios method or 86.32% of the three ratios techniques. In future research, several input vector features can be used for clusters formation in order to improve the quality of subset formed.

TABLE IX
THE DIAGNOSIS ACCURACIES OF THE PROPOSED METHOD AND SOME EXISTING METHODS OF LITERATURE ACCORDING TO FAULT SEVERITY

		Fault diagnostic accuracy (%)					
	Diagnostic methods	PD	D_1	D_2	T_1/T_2	T_3	Total
Traditional method	Modified Rogers' four ratios method [19]	53.85	97.92	100.00	62.50	77.78	80.34
	Modified IEC ratios method [19]	46.15	95.83	88.89	62.50	83.33	77.78
	IEC 60599 method [4]	76.92	33.33	44.44	62.50	61.11	52.14
	Three ratios technique [6]	73.08	97.92	88.89	68.75	88.89	86.32
	Clustering method [16]	57.69	77.08	88.89	68.75	66.67	70.94
	Gouda triangle method [7]	88.46	97.92	100.00	56.25	88.89	88.89
	Duval triangle method [3]	80.77	97.92	100.00	43.75	88.89	85.47
	HYOSUN Corporation ratios method [5]	77.78	80.77	97.92	75.00	88.89	88.03
	Combined technique N°1 [20]	53.85	97.92	100.00	56.25	83.33	80.34
Hybrid methods	Combined technique N°2 [20]	57.69	87.50	100.00	50.00	77.78	75.21
	Combined technique N°3 [20]	53.85	91.67	77.78	37.50	77.78	72.65
	Combined technique N°4 [20]	57.69	87.50	100.00	62.50	77.78	76.92
	Combined technique [23]	53.85	77.08	88.89	18.75	55.56	61.54
Intelligent methods	Conditional probability method [22]	50.00	89.58	100.00	50.00	61.11	71.79
	CSUS ANN method [8]	53.85	91.67	77.78	37.50	77.78	72.65
	Self-organizing map clusters method [21]	53.85	77.08	88.89	18.75	55.56	61.54
Proposed methods	GA-based *k*-MCA method	100.00	96.15	97.92	100.00	100.00	98.29

TABLE X
THE DIAGNOSIS ACCURACIES OF THE PROPOSED METHOD AND SOME EXISTING METHODS OF LITERATURE ACCORDING TO EQUIPMENT

		Diagnostic accuracy of equipment (%)								
	Diagnostic methods	B	C	I	P	R	S	U	Empty	Total
Traditional method	Modified Rogers' four ratios method [19]	20.00	50.00	100.00	80.56	81.25	71.43	86.36	100.00	80.34
	Modified IEC ratios method [19]	20.00	50.00	91.67	77.78	78.13	71.43	90.91	0.00	77.78
	IEC 60599 method [4]	40.00	0.00	58.33	47.22	56.25	42.86	63.64	0.00	52.14
	Three ratios technique [6]	20.00	100.00	91.67	91.67	84.38	71.43	95.45	100.00	86.32
	Clustering method [16]	20.00	100.00	91.67	63.89	84.38	42.86	68.18	100.00	70.94
	Gouda triangle method [7]	60.00	100.00	91.67	83.33	93.75	85.71	95.45	100.00	88.89
	Duval triangle method [3]	40.00	100.00	91.67	83.33	90.63	71.43	90.91	100.00	85.47
	HYOSUN Corporation ratios method [5]	40.00	100.00	83.33	88.89	90.63	71.43	100.00	100.00	88.03
	Combined technique N°1 [20]	20.00	50.00	100.00	77.78	81.25	71.43	90.91	100.00	80.34
Hybrid methods	Combined technique N°2 [20]	20.00	100.00	83.33	75.00	87.50	42.86	72.73	100.00	75.21
	Combined technique N°3 [20]	0.00	100.00	75.00	69.44	90.63	57.14	68.18	100.00	72.65
	Combined technique N°4 [20]	20.00	100.00	91.67	77.78	87.50	42.86	72.73	100.00	76.92
	Combined technique [23]	20.00	100.00	66.67	55.56	71.88	57.14	59.09	100.00	61.54
Intelligent methods	Conditional probability method [22]	20.00	100.00	83.33	75.00	78.13	42.86	68.18	100.00	71.79
	CSUS ANN method [8]	0.00	100.00	75.00	69.44	90.63	57.14	68.18	100.00	72.65
	Self-organizing map clusters method [21]	60.00	50.00	91.67	77.78	84.36	/	72.73	/	61.54
Proposed methods	GA-based *k*-MCA method	100.00	100.00	100.00	94.44	96.88	100.00	100.00	100.00	98.29

APPENDIX

PSEUDO CODE

```
1. Load the centroid matrix: M
2. Input the dissolved gas sample concentrations
3. Compute the gas ratios $R_1$ to $R_{15}$ (Tableau III)
4. Compute the feature input vector:
   $X = [\%H_2 \quad \%CH_4 \quad \%C_2H_6 \quad \%C_2H_4 \quad \%C_2H_2]$ (Eq. (5))
5. Compute the distances between the sample and the centroids
   $d = pdist2(X, M)$        (6)

6. Identify the subset of the sample

  if $d_{min} == d_1$ then
        N = $Cluster_1$;
  elseif $d_{min} == d_2$ then
        N = $Cluster_2$
  elseif $d_{min} == d_3$ then
        N = $Cluster_3$

        ⋮

  elseif $d_{min} == d_{119}$ then
        N = $Cluster_{119}$
  else $d_{min} == d_{120}$
        N = $Cluster_{120}$
  end if
7. Identify the fault type of the sample

   Switch N
      Case $Cluster_1$
         disp ('Low energy discharge: $D_1$')
      Case $Cluster_2$
         if $R_6 \geq 25$
            if $R_{15} \geq 250$
               disp ('Low energy discharge: $D_1$')
            else
               disp ('Partial discharge: PD')
            end
         else
            if $R_3 \geq 0.1$
               if $R_{13} < 3.5$
                  disp ('Low energy discharge: $D_1$')
               else
                  disp ('High energy discharge: $D_2$')
               end
            else
               disp ('Partial discharge: PD')
            end
      Case $Cluster_3$
         if $R_3 < 0.1$
            disp ('High temp. thermal fault: $T_3$')
         else
            if $R_6 < 0.1$
               if $R_{15} > 0.25$
                  disp ('Medium temp. thermal fault: $T_2$')
               else
                  disp ('High temp. thermal fault: $T_3$')
               end
            else
               disp ('High temp. thermal fault: $T_3$')
            end
         end

         ⋮

      Otherwise
         disp('Combination thermal and discharges: DT')

   end
```

REFERENCES

[1] Y. Zhang, Y. Tang, Y. Liu, and Z. Liang, "Fault diagnosis of transformer using artificial intelligence: A review," Front. Energy Res., vol. 10, no. 2, pp. 1–10, 2022.

[2] H. Zheng and R. Shioya, "A Comparison between Artificial Intelligence Method and Standard Diagnosis Methods for Power Transformer Dis- solved Gas Analysis Using Two Public Databases," IEEJ Trans. Electr. Electron. Eng., vol. 15, no. 9, pp. 1305–1311, 2020.

[3] "IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers," IEEE Std C57104-2019 Revis. IEEE Std C57104-2008, pp. 1–98, 2019.

[4] IEC 60599," Mineral Oil-Impregnated Electrical Equipment in Ser- vice–Guide to the Interpretation of Dissolved and Free Gases Analysis," International Electrotechnical Commission: Geneva, Switzerland, 2015.

[5] S. Kim, S. Kim, H. Seo, J. Jung, H. Yang, and M. Duval, "New methods of DGA diagnosis using IEC TC 10 and related databases Part 1: application of gas-ratio combinations," IEEE Trans. Dielectr. Electr. Insul., vol. 20, no. 2, pp. 685–690, 2013.

[6] O. E. Gouda, S. H. El-Hoshy, and H. H. E.L.-Tamaly, "Proposed three ratios technique for the interpretation of mineral oil transformers based dissolved gas analysis," IET Gener. Transm. Distrib., vol. 12, no. 11, pp. 2650–2661, 2018.

[7] O. E. Gouda, S. H. El-Hoshy, and H. H. E.L.-Tamaly, "Condition assessment of power transformers based on dissolved gas analysis," IET Gener. Transm. Distrib., vol. 13, no. 12, pp. 2299–2310, 2019.

[8] S. S. M. Ghoneim, I. B. Taha, and N. I. Elkalashy, "Integrated ANN-based proactive fault diagnostic scheme for power transformers using dissolved gas analysis," IEEE Trans. Dielectr. Electr. Insul., vol. 23, no. 3, pp. 1838–1845, 2016.

[9] J. C. Ferna´ndez, L. B. Corrales, F. H. Herna´ndez, I. F. Ben´ıtez, and J. R. Nu´n˜ez, "A Fuzzy Logic Proposal for Diagnosis Multiple Incipient Faults in a Power Transformer," in Progress in Artificial Intelligence and Pattern Recognition, Cham, pp. 187–198, 2021.

[10] K. N. V. P. S. Rajesh, U. M. Rao, I. Fofana, P. Rozga and A. Paramane, "Influence of Data Balancing on Transformer DGA Fault Classification With Machine Learning Algorithms," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 30, no. 1, pp. 385-392, 2023.

[11] S. Eke, G. Clerc, T. Aka-Ngnui, and I. Fofana, "Transformer condition assessment using fuzzy C-means clustering techniques," IEEE Electr. Insul. Mag., vol. 35, no. 2, 2019.

[12] A. Nanfak, C. Kom, and S. Eke, "Hybrid Method for Power Transform- ers Faults Diagnosis Based on Ensemble Bagged Tree Classification and Training Subsets Using Rogers and Gouda Ratios," Int. J. Intell. Eng. Syst., vol. 15, no. 5, pp. 12–24, 2022.

[13] G. Odongo, R. Musabe, and D. Hanyurwimfura, "A Multinomial DGA Classifier for Incipient Fault Detection in Oil-Impregnated Power Transformers," *Algorithms*, vol. 14, no. 4, p. 128, 2021.

[14] O. E. Gouda, S. H. El-Hoshy, and S. S. M. Ghoneim, "Enhancing the Diagnostic Accuracy of DGA Techniques Based on IEC-TC10 and Related Databases," IEEE Access, vol. 9, pp. 118031–118041, 2021.

[15] E. Li, L. Wang, and B. Song, "Fault Diagnosis of Power Transformers With Membership Degree," IEEE Access, vol. 7, pp. 28791–28798, 2019.

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

[17] A. Nanfak, S. Eke, C. H. Kom, R. Mouangue, and I. Fofana, "Inter- preting dissolved gases in transformer oil: A new method based on the analysis of labelled fault data," IET Gener. Transm. Distrib., vol. 15, no. 21, pp. 3032–3047, 2021.

[18] M. M. Islam, G. Lee, and S. N. Hettiwatte, "A nearest neighbour clustering approach for incipient fault diagnosis of power transformers," Electr. Eng., vol. 99, no. 3, pp. 1109–1119, 2017.

[19] I. B. M. Taha, A. Hoballah, and S. S. M. Ghoneim, "Optimal ratio limits of rogers' four-ratios and IEC 60599 code methods using particle swarm optimization fuzzy-logic approach," IEEE Trans. Dielectr. Electr. Insul., vol. 27, no. 1, pp. 222–230, 2020.

[20] S. A. Ward et al., "Towards Precise Interpretation of Oil Transformers via Novel Combined Techniques Based on DGA and Partial Discharge Sensors," Sensors, vol. 21, no. 6, Art. no. 6, 2021.

[21] S. Misbahulmunir, V. K. Ramachandaramurthy, and Y. H. M. Thayoob, "Improved self-organizing map clustering of power transformer dissolved gas analysis using inputs pre-processing," IEEE Access, vol. 8, pp. 71798–71811, 2020.

[22] I. B. Taha, D.-E. A. Mansour, S. S. Ghoneim, and N. I. Elkalashy, "Conditional probability-based interpretation of dissolved gas analysis for transformer incipient faults," IET Gener. Transm. Distrib., vol. 11, no. 4, pp. 943–951, 2017.

[23] M. Badawi et al., "Reliable Estimation for Health Index of Transformer Oil Based on Novel Combined Predictive Maintenance Techniques," IEEE Access, vol. 10, pp. 25954–25972, 2022.

[24] L. Cheng and T. Yu, "Dissolved Gas Analysis Principle-Based Intelligent Approaches to Fault Diagnosis and Decision Making for Large Oil- Immersed Power Transformers: A Survey," Energies, vol. 11, no. 4, 2018.

[25] C. Pizzuti and N. Procopio, "A K-means Based Genetic Algorithm for Data Clustering," in International Joint Conference SOCO'16-CISIS'16- ICEUTE'16, vol. 527, M. Gran˜a, J. M. Lo´pez-Guede, O. Etxaniz, A´ . Herrero, H. Quintia´n, and E. Corchado, Eds. Cham: Springer International Publishing, pp. 211–222, 2017.

[26] A. M. Ikotun, M. S. Almutari, and A. E. Ezugwu, "K-Means-Based Nature-Inspired Metaheuristic Algorithms for Automatic Data Clustering Problems: Recent Advances and Future Directions," Appl. Sci., vol. 11, no. 23, Art. no. 23, 2021.

[27] I. Aljarah, H. Faris, and S. Mirjalili, Eds., Evolutionary Data Clustering: Algorithms and Applications. Singapore: Springer, 2021.

[28] H. Shang, J. Xu, Z. Zheng, B. Qi, and L. Zhang, "A novel fault diagnosis method for power transformer based on dissolved gas analysis using hypersphere multiclass support vector machine and improved D–S evidence theory," Energies, vol. 12, no. 20, pp. 4017, 2019.

[29] E. T. Mharakurwa, G. N. Nyakoe, and A. O. Akumu, "Power Trans- former Fault Severity Estimation Based on Dissolved Gas Analysis and Energy of Fault Formation Technique," J. Electr. Comput. Eng., vol. 2019, pp. 1–10, Feb. 2019.

[30] M. M. Islam, G. Lee, and S. N. Hettiwatte, "Application of Parzen Window estimation for incipient fault diagnosis in power Transformers," High Voltage, vol. 3, no. 4, pp. 303–309, Dec. 2018.

[31] X. Zhang et al., "Research on transformer fault diagnosis: Based on im- proved firefly algorithm optimized LPboost–classification and regression tree," IET Gener. Transm. Distrib., vol. 15, no. 20, pp. 2926–2942, 2021.

[32] Y. Zhang, H. Wei, R. Liao, Y. Wang,L. Yang, and C. Yan, "A new support vector machine model based on improved imperialist competitive algorithm for fault diagnosis of oil-immersed Transformers," J. Electr. Eng. Technol., vol. 12, no. 2, pp. 830–839, 2017.

[33] H. Shang, J. Xu, Z. Zheng, B. Qi, and L. Zhang, "A novel fault diagnosis method for power transformer based on dissolved gas analysis using hypersphere multiclass support vector machine and improved D–S evidence theory," Energies, vol. 12, no. 20, pp. 4017, 2019.

[34] A. Nanfak, "DGA MATLAB codes GitHub repository "https://github.com/NanaudKmer/IEEE_TDEI Hybrid DGA Method.

[35] M. Duval and A. DePabla, "Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases," IEEE Electr. Insul. Mag., vol. 17, no. 2, Art. no. 2, 2001.