IJST, Transactions of Electrical Engineering, Vol. 35, No. E1, pp 13-24
Printed in The Islamic Republic of Iran, 2011

NOVEL FUSION APPROACHES FOR THE DISSOLVED GAS ANALYSIS OF INSULATING OIL*

M. ALLAHBAKHSHI AND A. AKBARI**

Dept. of Electrical Engineering, K.N.Toosi University of Technology, Tehran, I. R. of Iran
Email: akbari@kntu.ac.ir

Abstract– Dissolved Gas Analysis (DGA) is the most reliable technique to identify the incipient faults in power transformers. There are several DGA techniques in use such as Doernenburg, Rogers, IEC, etc. On the other side there is an increasing tendency to combine data from multiple sources and models to achieve more reliable results than individuals. This investigation proposes two fusion approaches consisting of fusion architectures and respective combination methods to combine DGA techniques and the gas ratios utilized in these techniques. The proposed approaches in this article apply a modified flexible neuro-fuzzy and a gating network as combination methods. Various gas concentration data were used for training and validating the models. Results showed that the proposed approaches have more advantages compared to the conventional DGA techniques. Finally, the importance degree of each gas-ratio to detect each fault was investigated.

Keywords– Artificial intelligence, data fusion, neuro-fuzzy systems, neural networks, support vector machines (SVM), dissolved gas analysis (DGA), fault diagnosis, power transformers

1. INTRODUCTION

Power transformers are vital and important equipment in power supply utilities. Ageing wave, rising energy consumption and liberation have caused an increase in loading power transformers. It causes thermal, electrical and mechanical stresses of transformers to increase. Therefore, it is very essential to have a suitable and effective evaluation of power transformer conditions.

For many years, dissolved gas-in-oil analysis (DGA) has been one method used to identify the incipient faults in oil-filled power transformers. Insulation oils under thermal and electrical stresses produce combustible gases through chemical reactions. These gases are dissolved in oil and include hydrogen, methane, ethane, ethylene, acetylene, carbon monoxide and carbon dioxide. Gas concentrations in ppm are measured by chromatographic analysis. The proportion of each gas concentration depends on the type and severity of the fault. Partial discharge (with low energy), thermal faults (medium energy up to large-scale), and electrical arcing (with high energy discharge) are the principal faults which can be identified by DGA techniques. Rogers [1], IEC-60599 [2], Doernenburg [1, 3] and Duval [4] are the most commonly used DGA techniques. Each technique uses some gas ratios for fault diagnosis and some techniques compare gas concentrations to the specified levels to evaluate a transformer's condition. Table 1 shows eight gas ratios applied in the mentioned DGA techniques, while Tables 2-4 and Fig. 1 show corresponding criteria for each technique.

Furthermore, several artificial intelligent methods have been applied to obtain accurate results. Wang [5] proposed a model to combine a fault diagnosis neural network and conventional DGA techniques. The model combines each fault probability in decision level through a simple geometric mean and represents

*Received by the editors May 4, 2010; Accepted March 10, 2011.
**Corresponding author

more efficient results than individual techniques. In [6], Chen applied a fuzzy method with soft boundaries and acquired more precise fault diagnosis. Huang [7] handled a special evolving wavelet network for transformer condition monitoring with fast training process. Moreover, some efforts have been made to utilize self organization map networks (SOM) [8], CMAC neural networks [9] and other artificial intelligent methods [10-11] for increasing fault diagnosis performance. Insufficient data samples for testing and applying artificial intelligent systems regardless of the existent knowledge of conventional methods are the main weaknesses of the mentioned researches.

Nowadays hybrid models and combination methods are noteworthy and expose more effective performances. The main objective of these methods is to use existent knowledge (criteria, conditions and gas ratios) in the conventional DGA techniques and the learning ability of the artificial intelligent models simultaneously. In [12], a Dempster-Shafer's method has been used to combine the diagnostic results of three DGA techniques. This is a static method based on the evidence theory [12]. In [13] a computational system, based on a combination of some traditional techniques, a general regression neural network and a fuzzy system, was introduced. Recently, a new multi-agent system was introduced in [14] for detecting incipient faults in power transformers. They employed a fusion approach to combine the results of the Roger, IEC, Duval's-triangle and ANN techniques. The proposed fusion approach is a Dempster-Shafer's method the same as the applied method in [12]. In this method, portions of each technique in the output are fixed, and so it is a static method based on the evidence theory. A significant improvement in the diagnosis results compared to the individual techniques was reported. Using dynamic method with flexible weights is the main difference between our proposed approaches and the applied technique in [12, 14]. Furthermore, we proposed a new fusion approach in feature level fusion which has more accurate results than the fusion in decision level.

Table 1. Key gas ratios

R_1	R_2	R_3	R_4
$\frac{CH_4}{H_2}$	$\frac{C_2H_2}{C_2H_4}$	$\frac{C_2H_2}{CH_4}$	$\frac{C_2H_6}{C_2H_2}$
R_5	R_6	R_7	R_8
$\frac{C_2H_4}{C_2H_6}$	$\%CH_4$	$\%C_2H_2$	$\%C_2H_4$

$$*: \%CH_4 = \frac{100x}{x+y+z}, \%C_2H_2 = \frac{100y}{x+y+z},$$

$$\%C_2H_4 = \frac{100z}{x+y+z} \quad Where:$$

$$x = (CH_4),\ y = (C_2H_2),\ z = (C_2H_4)\ in\ ppm$$

Table 2. Rogers ratios technique [1]

R_2	R_1	R_5	Diagnosis
<0.1	>0.1 <1.0	<1.0	Unit normal
<0.1	<0.1	<1.0	PD
0.1-3.0	0.1-1.0	>3.0	Arcing
<0.1	>0.1 <1.0	1.0 3.0	Low temperature thermal
<0.1	>1.0	1.0-3.0	Thermal<700 °C
<0.1	>1.0	>3.0	Thermal>700 °C

Table 3. Doernenburg ratios technique [1]

R_1	R_2	R_3	R_4	Diagnosis
>1.0	<0.75	<0.3	>0.4	Thermal
<0.1	NS	<0.3	>0.4	PD
>0.1 <1.0	>0.75	>0.3	<0.4	Arcing

Table 4. IEC ratios technique [2]

Diagnosis	R_5	R_1	R_2
Partial Discharge(PD)	<0.2	<0.1	NS
Discharge of low energy(D1)	>1	0.1-.5	>1
Discharge of high energy(D2)	>2	0.1-1	0.6-2.5
Thermal fault T1 (T<300ºc)	<1	NS	NS
Thermal fault T2 (300 ºc< T<700ºc)	1-4	>1	<0.1
Thermal fault T3 (T>700ºc)	>4	>1	<0.2

The proposed tool uses a rules table for combining methods output (approximately a voting- table). Another research has been done to combine some traditional DGA ratios and obtain more accurate performances by using grey clustering analysis [15]. These methods need less time and data for training.

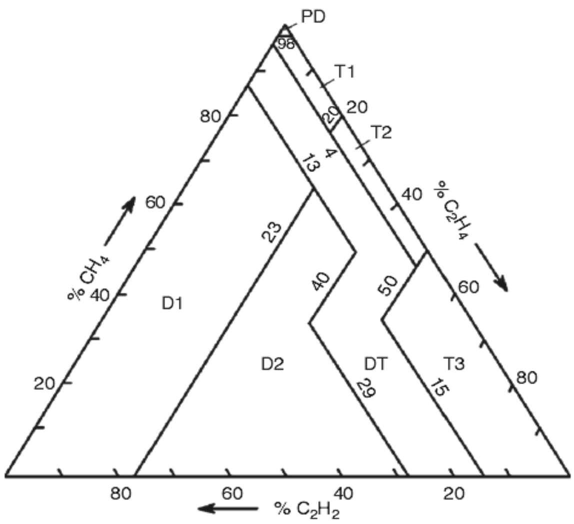

Fig. 1. Coordinates and fault zones of Duval’s Triangle [4]

This paper proposes two new architectures and relevant methods to combine data in the feature and decision level of the DGA techniques. In other words, using existent knowledge of the conventional methods and a combination of their features (gas ratios) to increase the diagnostic accuracy is the most important aim of the investigation. In section II two proposed combination architectures are described and respective combination methods, including gating network and flexible neuro-fuzzy method, are introduced in section 3. Finally, section 4 demonstrates the diagnosis results of the two proposed models.

2. FUSION ARCHITECTURES

New researches have a growing tendency toward using data fusion approaches, especially in signal processing, diagnostics and image processing systems. These approaches combine received information from various sources and models in different levels to achieve improved accuracies and more specific inferences than could be achieved by the use of a single source or model alone [16]. Various data fusion structures could be used in the different applications, but the most suitable architecture for diagnostic systems is the parametric data fusion architecture [17]. In such system data combination could be implemented in three ways as follows:

a) Data level fusion

In this architecture, raw data received from sensors or measurement devices are combined in an appropriate method to achieve more reliable and accurate data. The fused data are used for feature extraction. Subsequently, a fault identity process is performed to detect the type of fault from the extracted feature vector. The identity process in fault diagnostic systems creates a fault probability vector. This vector consists of probabilities for all types of faults. Raw data fusion gives the most accurate and also time-consuming results, assuming appropriate data association.

b) Feature level fusion

In this architecture, individual data sources produce respective features. These features are combined to give a new feature which is applied in the fault identification process to produce a fault probability vector. Since the feature vector has a smaller size than the raw data it needs less time to combine.

c) Decision level fusion

In this architecture, each data source performs an identity process based on its own data and features. Different identity processes could use the same data source or features. The fault probability vector provided individual identity processes are combined using decision level fusion techniques such as Bayesian inference, weighted decision methods, or Dempster-Shafer's method [17].

The selection among the mentioned architectures depends on the type of data, number of data sources, features and identity processes. In DGA, data sources and sensors are usually limited, but there are several features, especially gas ratios and diagnostic techniques, as mentioned before. For applying data fusion advantages in DGA techniques, two architectures are investigated and proposed as presented in Fig. 2.

Fig. 2a shows the proposed architecture for decision level fusion (DLF). In this architecture, gas concentrations are input data and the R1-R8 ratios act as features in four conventional DGA techniques. Each DGA technique uses its own ratios and predicts the fault probability vector individually. Four achieved fault probability vectors are combined in association module. A gating network module calculates a portion of each technique in the outputs. The gating network topology will be described in the next section.

Fig. 2b illustrates the proposed architecture for feature level fusion (FLF). This architecture applies and combines all ratios to calculate a fault probability vector, whereas in an ideal feature level fusion all ratios are combined to produce a new feature. So it is essential to realize that the architecture has some differences with the mentioned feature level fusion. A flexible neuro-fuzzy system defines the criteria and weight of each ratio through a learning process. The neuro-fuzzy system will be explained in the next part as a fusion method.

Detectable faults by DGA techniques vary in type and number as shown in Tables 2-4 and Fig. 1. All detected faults in the DGA techniques are divided to Partial Discharge Fault (PDF), Thermal Fault (THF), and Discharge Fault (DF).

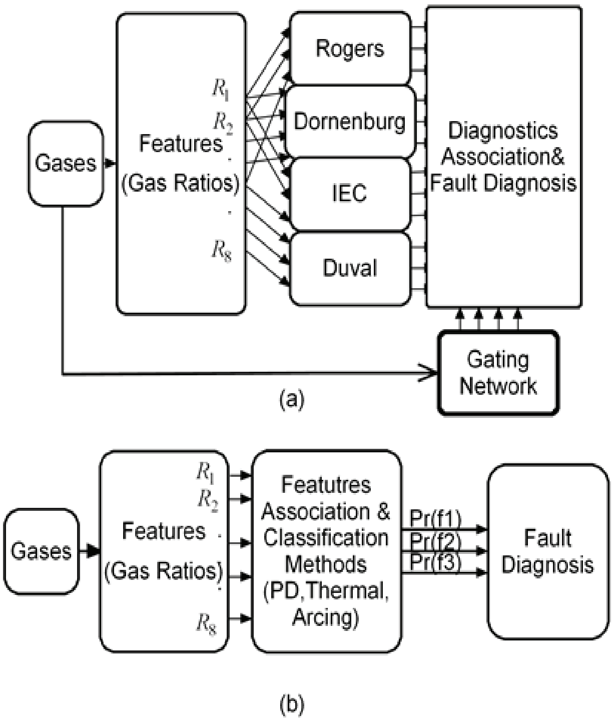

Fig. 2. The proposed fusion architectures, a): decision level fusion structure, b): feature (ratio) level fusion structure

3. FUSION METHODS

a) Gating network

This method is used for decision level fusion architecture. It basically combines available knowledge in the expert DGA systems, assuming the achieved results from the fused techniques are more accurate than individuals. The proposed block diagram with one layer neural network and softmax function for gating network is shown in Fig. 3. The weights of each DGA technique in the output vector depends on the gas concentrations and can change, hence it is called a dynamic method [18].

The output of gating networks g_i^j shows a portion of the *i*-th DGA technique in the *j*-th fault probability and can be calculated as follows by softmax function:

$$g_i^j = \frac{\exp(s_i^j)}{\sum_{k=1}^{4} \exp(s_i^j)} \quad , \; i = 1,2,3,4 \; and \; j = 1,2,3 \tag{1}$$

where s_i^j is the inner product of the gas concentrations vector and gating networks weights as follows:

$$s_i^j = \sum_{k=1}^{7} w_{ik}^j . x_k \tag{2}$$

In which x_k is the percentage of *k*-th gas concentration to all gases. w_{ik}^j is the connecting weight between *k*-th gas and g_i^j. On the other side, *i*-th DGA technique gives fault probability vector $\overline{o}_i$ using its criteria and ratios. Since these techniques use crisp criteria, $\overline{o}_i$ elements belong to $\{0,1\}$.

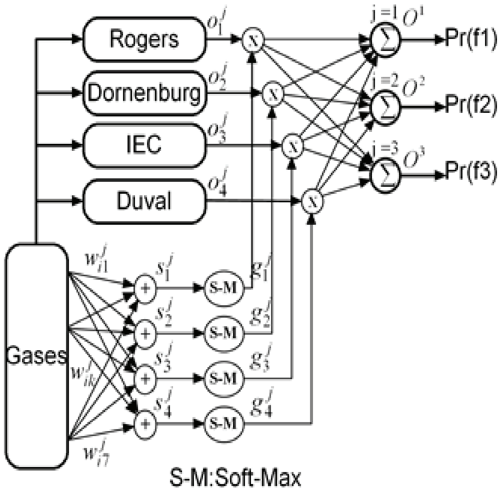

Fig. 3. Topology of the gating network

Now, the overall probability for each fault can be calculated as follows using gating network output and obtained fault probability vectors:

$$\Pr(f_j) = O^j = \sum_{i=1}^{4} g_i^j . o_i^j \ , \ j = 1,2,3 \tag{3}$$

where $\Pr(f_j)$ is j-th overall fault probability and j=1,2,3 for PDF,THF, DF, respectively.
Assuming j-th desired output O^j for input vector $\bar{x}$ is d^j, and applying mean squared error (MSE), the performance index can be expressed as follows:

$$E_j = \frac{1}{2} e_j^2 = \frac{1}{2}(d^j - O^j)^2 \tag{4}$$

According to gradient descent method and chain rule:

$$\frac{\partial E_j}{\partial w_{ik}^j} = \frac{\partial E_j}{\partial O^j} . \frac{\partial O^j}{\partial g_i^j} . \frac{\partial g_i^j}{\partial s_i^j} . \frac{\partial s_i^j}{\partial w_{ik}^j} \tag{5}$$

Then incremental variation Δw_{ik}^j is:

$$\Delta w_{ik}^j = -\eta . \frac{\partial E_j}{\partial w_{ik}^j} = \eta . e_j . o_i^j . g_i^j . (1 - g_i^j) . x_k \tag{6}$$

where η is the backpropagation learning rate. So during training w_{ik}^j are updated as follows:

$$w_{ik}^j(n+1) = w_{ik}^j(n) + \Delta w_{ik}^j(n) \tag{7}$$

b) Flexible neuro-fuzzy system

Neuro-Fuzzy systems exploit the learning capability of neural networks for enhancing the performances of fuzzy systems. A neuro-fuzzy system is a neural network with antecedent and consequents of fuzzy rules as nodes. The aim is optimizing the weights and parameters of the model through minimization of mean squared error [19]. There are several neuro-fuzzy system structures such as Adaptive Network-Based Fuzzy Inference System (ANFIS), Self Adaptive Neuro-Fuzzy Inference System (SANFIS), Flexible Neuro-Fuzzy System and others.

Referring to Tables 2-4 shows each DGA technique applies a set of predefined criteria for some ratios to detect the type of fault. To survey the importance and adjust the new boundary conditions of each of the

DGA ratios a flexible neuro-fuzzy system has been designed and is depicted in Fig. 4. Flexible neuro-fuzzy systems use a variable structure. Connectives and everything else are flexible to improve performance compared to the usual neuro-fuzzy systems [19].

The proposed system has an independent network for each fault, which acts individually. Furthermore, each network uses other networks data as input for training. After training, the weights of the ratios in the outputs and the new appropriate criteria (boundary conditions) would be determined.

In Fig. 4, Ω_i^j is the output of sigmoid membership function for R_i in j-th fault and can be calculated as follows:

$$\Omega_i^j = \frac{1}{1+\exp(-a_i.(R_i - cp_i^j))}, \quad i = 1,2,...,8 \;\; and \;\; j = 1,2,3 \tag{8}$$

where R_i is i-th ratio as shown in Table 1, cp_i^j is crossover point and a_i is responsible for the slope of function at the crossover point. For simplicity we utilize unique and fixed a_i for each R_i. Crossover points are computed as:

$$cp_i^j = \frac{\varepsilon^i}{1+\exp(-c_i^j)} \tag{9}$$

where ε^i, c_i^j are calculated through network learning.

In (9), ε^i determines the variation domain of crossover point cp_i^j, and for simplicity we utilize fixed values using conventional DGA techniques as presented in the next section.

To calculate the weights of i-th rule in j-th fault detection network, Fw_i^j is used as follows:

$$Fw_i^j = \frac{1}{1+\exp(-w_i^j)} \tag{10}$$

where w_i^j is calculated through network learning. Small and near to zero Fw_i^j represents a negligible portion of R_i in the j-th fault detection network and vice versa. Now each neuron output is given by:

$$o_i^j = 1 - Fw_i^j.(1-\Omega_i^j) \tag{11}$$

Therefore, the fault probability is calculated as:

$$\Pr(f_j) = \mathop{T}_{i=1}^{8}(o_i^j) / \sum_{k=1}^{3} \mathop{T}_{i=1}^{8}(o_i^k) \tag{12}$$

where T is a fuzzy t-norm such as *min* or *product*. Backpropagation algorithm and MSE as the performance index were applied for learning procedure and parameter adjustment.

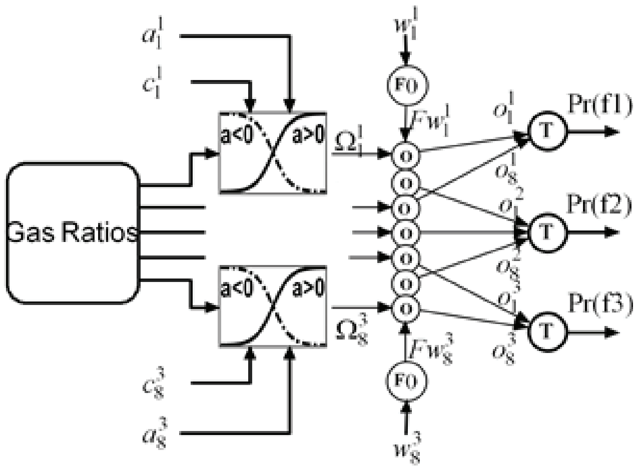

Fig. 4. Flexible neuro-fuzzy system structure

4. RESULTS AND DISCUSSION

To investigate the performances of the conventional DGA techniques and compare their ability with the proposed systems based on soft computing methods, an effort was made to collect sufficient data from several sources. About 200 cases of various faults and corresponding gas concentrations were obtained from five published articles [5-6], [20-22]. The mentioned data consist of 13 PDF, 70 THF and 101 DF cases. Since the number of PDF was not enough, 15 PDF cases simulated in the laboratory from [4] were used. Subsequently, 117 collected data from [22] and 5 PDF cases from [4] were used as the training data and the remaining data as the test data in the proposed systems for the performance evaluation process.

Total Diagnostic accuracy was calculated based on the successful diagnosis for all faults to the total data. For each fault, diagnostic accuracy was calculated based on the successful fault detection to the related data for each type of fault.

For comparison with an intelligent based technique, SVM technique was selected. The main idea of a support vector machine is to construct a hyperplane as the decision surface in such a way that the margin of separation between two classes of data is maximized. SVM can provide a good generalization performance for classification problems [18]. Three individual SVM classifiers as PD, Thermal and Discharge classifier were constructed. Each classifier was trained to detect one special fault, but data from all types of faults were used for training.

Despite the other methods, each classifier has two performances. The first is the accuracy to detect the type of fault with the related SVM, and the second performance is the accuracy to reject the other type of faults with the classifier. Since these classifiers are independent, there is a probability of a conflict with the results of the classifiers. For the mentioned conflict, we assumed the result as a wrong diagnosis. The input data for the SVM were normalized gas concentrations. Quadratic kernel function and quadratic programming were used to map the training data into kernel space and to find the separating hyperplane respectively.

Matlab toolbox was used for training the SVM classifiers with the same applied training data for FLF and DLF methods. The diagnostic accuracies of all classifiers for training data were 100%, but for the test data performance of the PD, the classifier reduced to 77.78 for detecting the PD faults and 83.33% for total cases (to detect the PD faults and reject the data for other types of fault). Also, the performances of the other two classifiers reduced for the test data as shown in Table 5.

a) DLF results

Presented criteria in Tables 2-4 and Fig. 1 were exactly applied (crisp boundaries). Different T1, T2 and T3 thermal faults were assumed as THF, also D1,D2 were assumed as DF. So each DGA technique has three individual crisp outputs which represent probability of faults. In the case of no-decision condition, where input values do not fall in the specified regions, the respective technique is not involved in the combination process.

Diagnostic performances of the four mentioned conventional DGA techniques and the proposed systems are given in Table 5 based on the collected data. Duval's performance for training data is the most accurate technique with about 92% accuracy. Other techniques accuracies are 55% for Rogers, 74% for Doernenburg, 100% for SVM and 77% for IEC. Combination of these techniques, using gating network method and decision level fusion, increased the accuracy to 94% for training data. These results demonstrate appropriate fusion because in an improper fusion, the performance of the overall model would reduce compared to individuals.

Accuracy of the proposed DLF model was 89% for test data, while calculated accuracies for Duval, IEC, Rogers, Doernenburg and SVM were 86, 69, 61, 71 and 85%, respectively. Accuracy reductions for Duval, SVM and IEC techniques are more noticeable.

b) FLF settings and results

1. Settings and limitations: The proposed flexible neuro-fuzzy system uses one rule and eight antecedents for each fault detection network as shown in Fig. 4. In this system, Fw describes the significance of particular antecedents in all rules.

Referring to the Table, $R_1 - R_5$ can vary in $[0, \infty)$ theoretically, but the variation domain for $R_6 - R_8$ is $[0,1]$. So ε^i is set as follows:

$$\varepsilon^i = \begin{cases} 5\ , & if\ \ i \le 5 \\ 1\ , & if\ \ i > 5 \end{cases} \tag{13}$$

To achieve soft boundaries and parameter reduction, the absolute value of a_i^j is set as:

$$\left| a_i^j \right| = 100 \tag{14}$$

the sign of a_i^j should be defined according to Tables 2-4 and Fig.1. For instance, R_1 is small in the PD fault and so in the respective membership function, *sign(*a_1^1*)* is defined as -1 and so on. The outcome is:

$$Sign(a_i^j) = \begin{bmatrix} -1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & 1 & -1 & 1 \end{bmatrix}^T \tag{15}$$

It should be noticed that the sigmoid function can cover just "smaller than" and "larger than" descriptions for ratios in the fuzzy domain. Some expressions such as "between a and b" need Gaussian membership function. For the sake of simplicity, we neglected these terms. The remaining parameters for adjusting in the learning process are $\{c, w\}$ as shown in Fig. 4.

2. FLF results: The proposed feature (ratio) level fusion method (FLF) was applied using a PC Pentium-4 3 GHz with 2 GB RAM and MATLAB software. Backpropagation algorithm was applied for the training process and mean squared error as performance index. Calculated Fw matrix after the training was:

$$Fw = \begin{array}{c} \begin{array}{ccc} \text{PDF} & \text{THF} & \text{DF} \end{array} \\ \begin{bmatrix} 0.99 & 0.02 & 0.01 \\ 0.05 & 0.96 & 0.99 \\ 0.12 & 0.02 & 0.01 \\ 0.07 & 0.02 & 0.01 \\ 0.09 & 0.01 & 0.90 \\ 0.99 & 0.97 & 0.99 \\ 0.31 & 0.96 & 0.07 \\ 0.67 & 0.11 & 0.13 \end{bmatrix} \end{array} \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ R_6 \\ R_7 \\ R_8 \end{array} \tag{16}$$

Each column has eight weights for relative fault as shown in (16). As mentioned before, a larger amplitude of Fw_i^j shows the more significant role of the *i*-th ratios in the *j*-th fault detection.

Table 5. Diagnostic accuracy (%) of DGA techniques and the proposed systems

	PD	Thermal	Discharge	Total
		Training Data		
Rogers	6.25	58.82	63.51	54.84
Doernenburg	31.25	76.47	82.43	74.19
IEC	25.00	82.35	86.49	77.42
Duval	56.25	91.18	100	91.94
DLF	56.25	100	100	94.35
FLF	56.25	100	100	94.35
SVM	100	100	100	100
		Test Data		
Rogers	22.22	86.11	40.74	61.11
Doernenburg	55.56	94.44	44.44	70.83
IEC	22.22	94.44	51.85	69.44
Duval	33.33	94.44	92.59	86.11
DLF	44.44	97.22	92.59	88.89
FLF	55.56	100	92.59	91.67
SVM*	77.78(83.33)	91.67(88.89)	77.78(88.89)	84.72(87.03)

* In this method two diagnostic accuracies are calculated, the first is the accuracy to detect the right fault and the other is the accuracy to detect the right fault and reject the data for other types of faults.

According to (16), the most important ratio is R_6. It has considerable weights for all three faults. The next important ratio is R_2 with two important weights for detecting THF and DF. In the next stage there are R_1, R_5, R_7 and R_8 with just one significant weight and the remaining ratios are negligible in fault diagnosis.

Table 5 shows the obtained results of the FLF method. The proposed method demonstrates satisfactory performance for detecting THF and DF. The resultant accuracy is superior to the mentioned conventional techniques, DLF and SVM method. Performance of the DGA techniques in PDF detection is not desirable, and consequently their combination results in unsatisfactory performance. Nevertheless, the proposed systems performances in PDF detection are at least equal to or better than the individual techniques, except for the SVM method as shown in Table 5.

New studies have shown that just PDs of the corona-type are detectable by DGA techniques [4]. These types of PDs occur in the gas phase of voids or gas bubbles and are very different from PDs of the sparking type occurring in the oil phase. To investigate the performance of the proposed methods, the collected PD cases were divided into two mentioned types. The performances of the proposed systems and conventional DGA techniques for detecting corona-type PDs are illustrated in Table 6. This confirms the appropriate ability of the proposed systems for detecting PDs of the corona type.

Table 6. Successful diagnostics of various DGA techniques and the proposed systems for detecting corona-type PDs

	Training data	Test data
Rogers	1/9	0/5
Doernenburg	4/9	2/5
IEC	4/9	2/5
Duval	9/9	3/5
DLF	9/9	3/5
FLF	9/9	5/5
SVM	9/9	3/5

To combine the knowledge and advantages of the conventional DGA techniques, two combination methods were proposed and evaluated. In the first step a combination methodology in decision level was introduced. This method combines the diagnoses of conventional DGA methods with a simple gating network. The ability of such combination can be improved as shown in Tables 5, 6. The portion of each diagnostic technique in the output is calculated through the gating network. Although the diagnostic accuracy can be improved in a proper combination, its performance is basically dependent on the performances of combined DGA techniques. This means that the combination method has a wrong diagnosis when all techniques fail to identify the type of fault. The second proposed method (FLF) combines all introduced ratios in the conventional DGA techniques as features in a neuro-fuzzy system. Due to the size of the data, some rules and parameters were eliminated or used as fixed weights. Nevertheless, this method showed the most accurate performance as shown in Table 5.

Assuming the IEC as a reference method, the proposed FLF method showed considerable improvement in the diagnostic accuracy compared to several of the published investigations for test data. The diagnostic accuracy for the proposed FLF method was 91.67% and 69.44% for IEC, and consequently 22.23% improvement in the diagnostic accuracy compared to IEC, while this improvement was 14%, 10%, 10.8% and 6% in [9], [10], [7] and [23], respectively.

5. CONCLUSION

Several combination methods, machine learning and rules mining techniques were introduced in electrical and computer systems [24-25]. To implement these methods in fault diagnostic systems, new fusion approaches were proposed for the conventional DGA techniques and related features. DGA techniques with several ratios and criteria have long been used for power transformer fault diagnosis. They have relative advantages and disadvantages. Many artificial techniques were applied to obtain a more accurate result. Nevertheless, there are limited investigations that have used DGA knowledge and the learning ability of artificial intelligent systems simultaneously. In this investigation, two combination systems were applied in the decision level and feature level for conventional DGA techniques. Gas concentrations of 196 faulty cases were collected for training and validating the methods. Three types of faults including partial discharge, thermal faults and discharge fault were selected as the main faults. The results showed:

- Combination in the feature level, using the flexible neuro-fuzzy system, gives the most reliable results. It also determines the weight of each ratio for individual fault detection networks.
- Combination in the decision level gives more reliable performance than individual techniques and is less accurate than feature level fusion. More parameters for adjusting in the learning process and the basic limitation of each DGA technique are the possible reasons.
- Both combination approaches showed dominant performances compared to conventional DGA techniques, while Duval's technique had the best performance in the studied DGA techniques.
- By using FLF, it has been demonstrated that R_2, R_6 have important capability for detecting main faults. In other words, concentrations of methane, ethylene and acetylene have a key role in fault diagnosis. This explains the superiority of Duval's technique because it uses these three gases as inputs.
- The ability of the DLF method in fault identification is considerably dependent on the performances of the conventional methods. So the undefined conditions for some criteria in the conventional ratio techniques will decrease the performances of these methods. Also, the overlapped region in Duval's technique decreases the performance of the DLF method, while the FLF method basically solves the problem through using the fuzzy boundaries.

REFERENCES

1. IEEE Guide for the interpretation of gases in oil-immersed transformers, (1991). ANSI/IEEE Standard C57.104-1991.

2. IEC Publication 60599, (1999). *Mineral oil-impregnated electrical equipment in service—guide to the interpretation of dissolved and free gases analysis*.
3. Doernenburg, E. & Strittmater, W. (1974). Monitoring oil-cooling transformers by gas analysis. *Brown Boveri Rev*, Vol. 61, pp. 238-274.
4. Duval, M. (2002). A review of faults detectable by gas-in-oil analysis in transformers. *IEEE Elect. Insul. Mag.,* Vol. 18, No. 3, pp. 8-17.
5. Wang, Z., Liu, Y. & Griffin, P. (1998). A combined ANN and expert system tool for transformer fault diagnosis. *IEEE Trans. Power Del*., Vol. 13, No. 4, pp.1224-1229.
6. Chen, A. P. & Lin, C. C. (2002). Fuzzy approaches for fault diagnosis of transformers. *Fuzzy Sets and Systems*, Vol. 118, pp. 139-151, (2001).
7. Huang, Y. C. & Huang, C. M. (2002). Evolving wavelet networks for power transformer condition monitoring. *IEEE Trans. On Power Deliv.,* Vol. 17, No. 2, pp. 412-416.
8. Thang, K. F., Aggarwal, R. K., McGrail, A. J. & Esp, D. G. (2003). Analysis of power transformer dissolved gas data using the self-organizing map. *IEEE Trans. On Power Deliv.*, Vol. 18, No. 4, pp. 1241-1248.
9. Hung, C. P. & Wang, M. H. (2004). Diagnosis of incipient faults in power transformers using CMAC neural network approach. *Elec. Power Sys. Res*., Vol. 71, pp. 235-244.
10. Kuo, H. C., Chang, H. K. & Wang, Y. Z. (2004). Symbiotic evolution-based design of fuzzy-neural transformer diagnostic system. *Elect. Power Sys. Res*., Vol. 72, pp. 235-244.
11. Duraisamy, V., Devarajan, N., Samsundareswari, D., Vasanth, A. A. M. & Sivanandam, S. N. (2007). Neuro fuzzy schemes for fault detection in power transformer. *Applied Soft Computing*, Vol. 7, pp. 534-539.
12. Allahbakhsi, M. & Akbari, A. (2007). Multiple-model data fusion for DGA techniques. *Proc. 2007, International Symposium on High Voltage Engineering Conf*., Slovenia.
13. Morais, D. R. & Rolium, J. G. (2006). A hybrid tool for detection of incipient faults in transformers based on the dissolved gas analysis of insulating oil. IEEE Trans. On Power Deliv., Vol. 21, No. 2, pp. 673-680.
14. Akbari, A., Setayeshmehr, A., Fofana, I., Borsi, H. & Gockenbach, E. (2010). Intelligent agent-based system using dissolved gas analysis to detect incipient faults in power transformers. *IEEE Electrical Insulation Magazine*, Vol. 26, No. 6, pp. 28-41.
15. Lin, G. H., Wu, C. H. & Huang, P. Z. (2009). Grey clustering analysis for incipient fault diagnosis in oil-immersed transformers. *Expert Sys. With Appl*., Vol. 36, pp. 1371-1379.