Power Transformer Fault Prediction using Naive Bayes and Decision tree based on Dissolved Gas Analysis

Yassine Mahamdi, Ahmed Boubakeur, Abdelouahab Mekhaldi, and Youcef Benmahamed

Abstract− Power transformers are the basic elements of the power grid, which is directly related to the reliability of the electrical system. Many techniques were used to prevent power transformer failures, but the Dissolved Gas Analysis (DGA) remains the most effective one. Based on the DGA technique, this paper describes the use of two of the most effective machine learning algorithms: Naive Bayes and Decision Tree for the identification of power transformer's faults. In our investigation, 9 different input vectors have been developed from widely known DGA techniques. 481 samples have been used and 6 types of faults have been considered. The evaluation result of the implementation of the proposed methods shows an effectiveness of 86.25% in power transformer's fault recognition.

***Keywords*− Decision Tree, Naive Bayes, DGA, Input vectors, Power transformer faults, Accuracy rate.**

NOMENCLATURE

DGA	Dissolved Gas Analysis
DT	Decision Tree
NB	Naive Bayes
PD	Partial Discharges
D1	Low Energy Discharges
D2	High Energy Discharges
T1	Thermal Faults < 300 °C
T2	Thermal Faults of 300 °C to 700 °C
T3	Thermal Faults > 700 °C

I. INTRODUCTION

The Dissolved Gas Analysis (DGA) is the most common and effective method for detecting transformer faults [1]. It can immediately prevent internal transformer failures, which generally avoids huge economic losses. The DGA uses the values of the concentrations of the various gases released in the transformer oil due to the decomposition of the oil and the insulating paper.

In-service transformer is exposed to two types of stresses: electrical and thermal [2]. Due to these stresses, the transformer oil and paper decompose, releasing a set of gases that reduce their dielectric strength. The nature and quantity of each dissolved gas produced in transformer oil can indicate the internal condition of the transformer.

The most common gases produced by the decomposition of oil are: ethane ($C2H6$), ethylene ($C2H4$), acetylene ($C2H2$), methane ($CH4$) and hydrogen ($H2$) [3-4]. In addition to carbon dioxide ($CO2$) and carbon monoxide (CO) that are formed as a result of the decomposition of the insulating paper [5-6], while, the nitrogen ($N2$) and the Oxygen ($O2$) are the non-fault gases.

Manuscript received October 22, 2021; revised December 29, 2021.

Y. Mahamdi, A. Boubakeur, A. Mekhaldi, and Y.Benmahamed are with Ecole Nationale Polytechnique (e-mail: yassine.mahamdi@g.enp.edu.dz, ahmed.boubakeur@g.enp.edu.dz, abdelouahab.mekhaldi@g.enp.edu.dz, youcef.benmahamed@g.enp.edu.dz).

Digital Object Identifier (DOI): 10.53907/enpesj.v2i1.63

The main three conditions that can cause gas generation in a transformer are:

1) Corona (partial discharge)
2) Thermal heating
3) Arcing

These differ mainly in the intensity of the energy that is dissipated by the default [2], [7].

There are many approaches developed for the analysis of dissolved gases in transformer oil and interpret their meaning including IEC Ratios, DORNENBURG Ratios, Rogers Ratios, Duval Triangle and Pentagon, and, Key Gas method [2], [7-10]. However, these techniques have certain limitations such as the existence of non-decision areas and erroneous results [11-12]. To overcome this situation, several artificial intelligence techniques have been used to improve the diagnostic accuracy of power transformers.

Recently, intelligent methods, such as fuzzy logic inference systems [13], artificial neural networks [4], [14-15], support vector machines (SVM) [16-19], K-nearest neighbors [16], [20], Bayesian neural networks [21], hybrid grey wolf optimization technique [5], and some other machine learning algorithms have been applied to diagnosis the power transformer faults and have impressive performances [22-24].

In this paper, the Naive Bayes and the Decision Tree algorithms were used in faults identification. The originality comes from the introduction of several input vectors. These input vectors have been constructed using widely known DGA techniques, so that the most suitable input data that gives the best performance of each algorithm and achieves the best prediction of fault in power transformers can be identified.

This article is arranged as follows: in the Second Section, the principle of our methodology for identifying power transformers faults has been described. The process and the results of implementing the two algorithms using our proposed

2716-912X © 2022 Ecole Nationale Polytechnique

input vectors are discussed in the third section, where, the best-input vector for each technique has been identified. Finally, the conclusions from this work were summarized and potential future work was mentioned.

II. Methodology

A. Data collection:

The construction of our proposed input space needs gas concentration values, for this purpose, transformer oil samples are periodically taken for the DGA test using gas chromatography analysis [25-26]. Generally, mixtures of all gases are present in an oil sample, where the relative amount of each, could be an indicator of the existing faults, such as, partial discharges (PD), thermal faults > 700 °C (T3), thermal faults of 300 °C to 700 °C (T2), thermal faults < 300 °C (T1), high energy discharges (D2) and low energy discharges (D1) [7].

In this work, a database of 481 samples has been used in training and testing the proposed methods. This database has been extracted from the literature [27]. The distribution of the training and the testing samples according to their fault type is shown in Table 1.

Table. I
DISTRIBUTION OF TRAINING AND TESTING DATASET SAMPLES

Fault Types	Samples For Training	Samples For Testing
Partial Discharge	32	16
Thermal Faults > 700 °C	57	28
Thermal Faults of 300 °C to 700 °C	32	16
Thermal Faults < 300 °C	63	32
High Energy Discharges	84	42
Low Energy Discharges	53	26
TOTAL	321	160

B. Proposed Input vectors:

The following attributes have been considered in the construction of our proposed input vectors:

1) Using the concentration of the usual five key gases in ppm:

$$X = [C_2H_6, C_2H_4, C_2H_2, CH_4, H_2] \quad (1)$$

2) Using the ratios between key gases (The IEC Ratios):

$$X = [\frac{C_2H_4}{C_2H_6}, \frac{C_2H_2}{C_2H_4}, \frac{CH_4}{H_2}] \quad (2)$$

3) Using the relative percentages of gases:

$$X = [\%C_2H_6, \%C_2H_4, \%C_2H_2, \%CH_4, \%H_2] \quad (3)$$

4) Using ROGER's four-ratio:

$$X = [\frac{C_2H_6}{CH_4}, \frac{C_2H_4}{C_2H_6}, \frac{C_2H_2}{C_2H_4}, \frac{CH_4}{H_2}] \quad (4)$$

5) Using DORNENBURG's four-ratios:

$$X = [\frac{C_2H_2}{CH_4}, \frac{C_2H_4}{C_2H_6}, \frac{C_2H_2}{C_2H_4}, \frac{CH_4}{H_2}] \quad (5)$$

6) Using Duval's triangle coordinates:

$$X = [C_a, C_b] \quad (6)$$

Where

$$C_a = \frac{1}{3}\frac{\sum_{i=0}^{k-1}(a_i+a_{i+1})(a_ib_{i+1}-a_{i+1}b_i)}{\sum_{i=0}^{k-1}(a_ib_{i+1}-a_{i+1}b_i)} \quad (7)$$

And

$$C_b = \frac{1}{3}\frac{\sum_{i=0}^{k-1}(b_i+b_{i+1})(b_ia_{i+1}-b_{i+1}a_i)}{\sum_{i=0}^{k-1}(b_ia_{i+1}-b_{i+1}a_i)} \quad (8)$$

The a_i are calculated by the equations:

$$a_0 = \%CH_4 \cos\left(\frac{\pi}{2}\right)$$
$$a_1 = \%C_2H_4 \cos\left(\frac{\pi}{2}+\varphi\right) \quad (9)$$
$$a_2 = \%C_2H_2 \cos\left(\frac{\pi}{2}+2\varphi\right)$$

And the b_i could be obtained by replacing "*cos*" with "*sin*" in the last equations with $\alpha = 2\pi/3$

7) Using Duval's pentagon coordinates:

$$X = [C_a, C_b] \quad (10)$$

Where

$$C_a = \frac{1}{6}\frac{\sum_{i=0}^{k-1}(a_i+a_{i+1})(a_ib_{i+1}-a_{i+1}b_i)}{\sum_{i=0}^{k-1}(a_ib_{i+1}-a_{i+1}b_i)} \quad (11)$$

And

$$C_b = \frac{1}{6}\frac{\sum_{i=0}^{k-1}(b_i+b_{i+1})(b_ia_{i+1}-b_{i+1}a_i)}{\sum_{i=0}^{k-1}(b_ia_{i+1}-b_{i+1}a_i)} \quad (12)$$

The a_i are calculated using the following equations:

$$a_0 = \%H_2 \cos\left(\frac{\pi}{2}\right)$$
$$a_1 = \%C_2H_6 \cos\left(\frac{\pi}{2}+\varphi\right)$$
$$a_2 = \%CH_4 \cos\left(\frac{\pi}{2}+2\varphi\right) \quad (13)$$
$$a_3 = \%C_2H_4 \cos\left(\frac{\pi}{2}+3\varphi\right)$$
$$a_4 = C_2H_4 \cos\left(\frac{\pi}{2}+4\varphi\right)$$

Also, the b_i could be obtained by replacing "*cos*" with "*sin*" in the last equations with $\alpha = 2\pi/5$

8) In this case, a combination of two of the previously mentioned input vectors has been done, Roger's and DORNENBURG's ratios:

$$X = [\frac{C_2H_6}{CH_4}, \frac{C_2H_2}{CH_4}, \frac{C_2H_4}{C_2H_6}, \frac{C_2H_2}{C_2H_4}, \frac{CH_4}{H_2}] \quad (14)$$

9) To further improve fault recognition by expanding the proposed input space , another combination was made in the case of this input vector, Duval's triangle-pentagon coordinate's combination:

$$X = [C_{a1}, C_{b1}, C_{a2}, C_{b2}] \quad (15)$$

Where *{Ca1, Cb1}* are calculated using the triangle method, while *{Ca2, Cb2}* are calculated according to the pentagon one.

C. AI techniques:

1) Naive Bayes

The Naive Bayes algorithm is a simple probabilistic classifier based on Bayes theorem that calculates a set of probabilities by counting the frequency and combinations of values in a given data set. The algorithm assumes that all variables are independent considering the value of the class variable [28], which is rarely existent in real-world applications, so it is characterized as Naive, but the algorithm tends to learn quickly in a variety of controlled classification problems [29]. Bayes theorem is a mathematical formula used to determine the posterior probability $P(x|y)$ from $P(x)$, $P(y)$, and $P(y|x)$:

$$P(x|y) = \frac{P(y|x) \times P(x)}{P(y)} \quad (16)$$

Where $P(x|y)$ refers to the subsequent possibility of the hypothesis x conditioned by some evidence y and $P(x)$ is the prior probability of x; $P(y|x)$ is the likelihood for y given x, and $P(y)$ is the prior or marginal probability of y. Therefore, Bayes theorem can be clearly interpreted as an alternative form in (17) with respect to each item in (16):

$$Posterior\ Probability = \frac{Likelihood \times Prior\ Probability}{Probability\ of\ Evidence} \quad (17)$$

2) Decision tree

The decision tree algorithm is a non-parametric supervised machine learning's classifier used to split data into a set of branches. The construction of the tree is conducted from top to bottom in a recursive divide-and-conquer manner [30]. Decision tree classifiers are easier to interpret than other classification methods because decision tree is able to break down complex decision-making process into multi union of simpler decisions [31]. Generally, decision tree classifier training is based on finding the best split at each node as long as the complete data set is not analyzed [32]. The said principle leads to the idea of partitioning the feature space until the interrupt criterion is satisfied in each list, or until all points in a given leaf belong to one class. In order to meet the criteria, it is necessary to select the most common class among the data in the list or the one with the highest information gain. Figure 1 illustrates the basic structure of a decision tree.

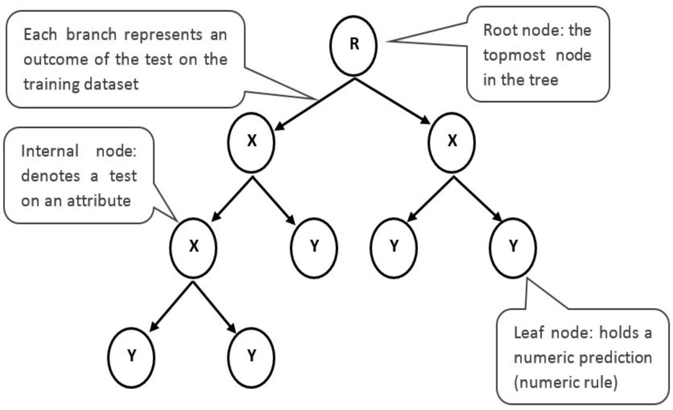

Fig. 1: Decision Tree general structure

Among other classification algorithms, Decision Tree have the following advantages:

- Good performance with large data sets
- Requires little data preparation
- Able to handle both numerical and categorical data
- Easy to display graphically
- Easy to understand and interpret

<u>Construction of decision tree:</u>

In order to select the best variable to split, the Decision Tree uses the information gain. The equation for calculating information gain is as follows:

$$Gain(T, A) = Entropy(T) - \sum_{i=1}^{n} \frac{T_i}{T} Entropy(T_i) \quad (18)$$

Where $Gain(T, A)$ is the information gain of set T (training data) on an attribute A, and T_i is a subgroup of T for which: A has value i.

The Entropy of node T is defined as:

$$Entropy(T) = -\sum_{i=1}^{n} p(i) \log p(i) \quad (19)$$

Where $p(i)$ is the proportion of T belonging to a class i.

III. Results and Discussion

To evaluate the performance of Naïve Bayes and Decision tree algorithms using our proposed input vectors according to six types of transformer faults, namely, they are categorized into partial discharges (PD), low energy discharge (D1), high energy discharge (D2), low thermal fault (T1), medium thermal fault (T2), and high thermal fault (T3), a set of 481 samples has been used to train and test the two methods; 67% of the dataset were used for the training and 33% for the testing, using the MATLAB software. Figure 2 illustrates a brief description of the proposed method.

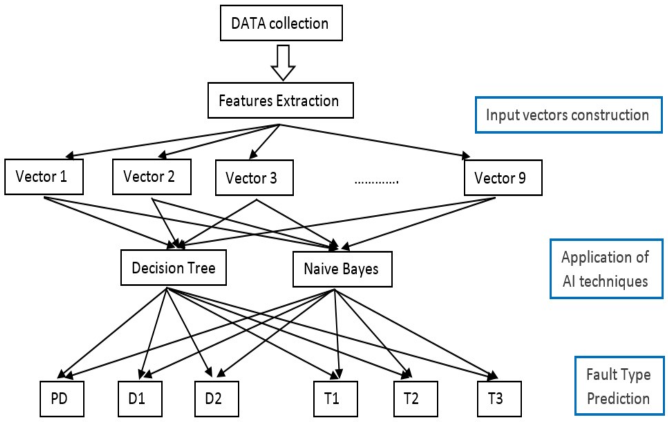

Fig. 2: The general structure of transformer fault recognition using the Naïve Bayes and the Decision Tree Algorithms

Table 2 shows the results of the implementation of the two classifiers using the proposed input vectors.

Table. II

Faults diagnosis accuracies using the Naïve Bayes and the Decision Tree algorithms with all the proposed input vectors

Input vector	Naïve Bayes	Decision tree
Vector 1	25.62	75.62
Vector 2	81.87	80.62
Vector 3	13.75	83.12
Vector 4	11.25	83.75
Vector 5	28.75	77.50
Vector 6	58.25	45.00
Vector 7	42.50	78.75
Vector 8	28.75	76.25
Vector 9	86.25	78.75

From Table 2, it is easy to see that the highest prediction accuracy is obtained using Vector 9 (combined Duval's pentagon and triangle) with the Naïve Bayes algorithm (86.25%). Whereas, in the case of the Decision Tree, the input Vector 4 (Roger's four-ratio method) gives the highest prediction accuracy, up to 83.75%.

In order to deepen the study, the performance of each algorithm with its appropriate input vector was evaluated based on the accuracy of each fault type diagnosis (Figure 3).

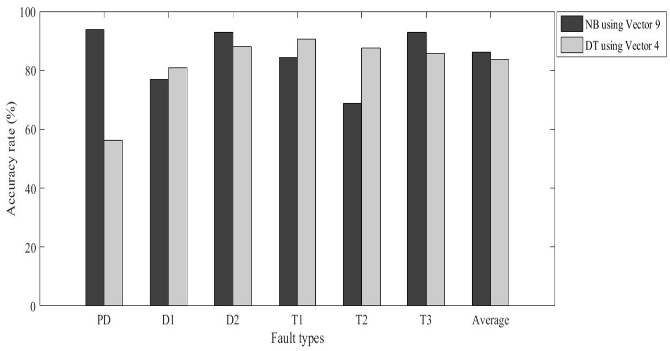

Fig. 3: Histogram of accuracy rate

From Figure 3, it is clear that the performance of each algorithm differs depending on the type of fault. For example, in the case

Table. IV

TANSFORMER FAULTS DIAGNOSIS USING THE TRADITIONAL DGA METHODS AND OUR PROPOSED METHODS

Sample	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	Ref.	Duval	Rogers Ratios	IEC Ratios	NB - Vector 9	DT - Vector 4	Act.
01	120	140	30	0	120	[33]	D1	UD*	UD	T2	T1	T1
02	3700	6400	2400	10	7690	[33]	D1	UD	UD	T3	T1	T2
03	125	680	290	20	900	[33]	D1	UD	UD	T3	T1	T3
04	120	10	30	25	5	[33]	T3	UD	PD	D1	PD	D1
05	140	95	10	80	60	[33]	D2	D2	D2	D2	D2	D2
06	240	17	0	5	40	[33]	D1	UD	UD	PD	PD	PD
07	650	53	20	0	34	[33]	D1	PD	UD	PD	PD	PD
08	1076	95	71	231	4	[33]	T3	UD	UD	PD	PD	PD
09	6454	2313	121	6432	2159	[34]	D2	UD	D2	D2	D2	D2
10	305	100	33	541	161	[34]	D2	UD	D2	D2	D2	D1
11	1230	163	27	692	233	[34]	D2	UD	D2	D1	D2	D2
12	33046	619	58	0	2	[34]	PD	PD	UD	PD	PD	PD
13	796	999	234	31	1599	[34]	D1	UD	UD	T3	T1	T3
14	34	21	4	56	49	[34]	D2	D2	D2	D2	D2	D2
15	960	4000	1290	6	1560	[34]	D1	UD	UD	T2	T1	T2
16	6	2990	29990	67	26076	[34]	D1	UD	UD	T1	T1	T1
17	2500	10500	4790	6	13500	[34]	D1	UD	UD	T3	T1	T2
18	300	700	280	36	1700	[34]	D1	UD	UD	T3	T1	T3
19	37800	1740	249	8	8	[34]	PD	PD	PD	PD	PD	PD
20	1450	940	211	61	322	[34]	D1	UD	UD	PD	T1	T1

UD* (Undefined): the method used is unable to determine the type of the fault

of the partial discharges (PD), the Naïve Bayes has the best performance, while, in the case of medium thermal fault (T2), the Decision Tree has the superiority in such fault recognition. Overall, the Naïve Bayes algorithm remains the one with the greatest precision. Table 3 shows the overall result of transformer fault diagnosis using our proposed input vectors with the two used classification algorithms.

Table. III

THE OVERALL RESULTS OF DIAGNOSTIC ACCURACIES USING NAÏVE BAYES AND DECISION TREE CLASSIFIERS

Fault	Vector 1		Vector 2		Vector 3	
Type	NB	DT	NB	DT	NB	DT
PD	25	37.5	87.5	75	93.75	62.5
D1	57.69	73.07	80.76	84.61	3.84	84.61
D2	11.9	69.04	80.95	73.80	2.38	83.33
T1	40.62	100	81.25	84.37	3.12	93.75
T2	0	62.50	68.75	75	25	87.5
T3	14.28	89.28	89.28	89.28	0	78.57

Fault	Vector 4		Vector 5		Vector 6	
Type	NB	DT	NB	DT	NB	DT
PD	0	56.25	81.25	87.5	0	18.75
D1	50	80.76	92.30	80.76	0	50
D2	2.38	88.09	2.38	85.71	97.61	59.52
T1	3.12	90.62	3.12	81.25	84.37	56.25
T2	18.75	87.50	43.75	50	0	31.25
T3	0	85.71	3.57	67.85	89.28	28.57

Fault	Vector 7		Vector 8		Vector 9	
Type	NB	DT	NB	DT	NB	DT
PD	93.75	68.75	81.25	87.50	93.75	87.50
D1	0	69.23	92.30	76.92	76.92	73.07
D2	80.95	85.71	0	85.71	92.85	83.33
T1	59.37	81.25	6.25	87.50	84.37	90.62
T2	0	62.50	37.5	31.25	68.75	68.75
T3	14.28	89.28	89.28	89.28	0	78.57

For the validation stage, the results of Table 3 confirm that the optimal choice of classification algorithm with appropriate input data is critical in the diagnosis of transformer faults.

In order to assess the improvement in fault prediction using our proposed methods, Table 4 presents the diagnosis accuracies of the most common traditional DGA methods (Duval's triangle, Rogers's ratios and IEC ratios) and our proposed methods using a random dataset. The results in Table 4 shows the superiority of the NAÏVE BAYES algorithm with the 9th input vector and the DECISION TREE algorithm with the 4th input vector in transformer fault diagnosis (70% and 60% respectively), while, the other three methods developed poor diagnostic accuracies, Duval triangle (30%), Rogers's ratios (25%), and IEC Ratios (30%).

IV. CONCLUSION

The Naïve Bayes and the Decision Tree classification algorithms were used to identify power transformer faults. A dataset of 481 samples was employed and 9 different input vectors were considered. The Naive Bayes algorithm achieved a diagnostic accuracy of 86.25% when using the 9th input vector (Duval's triangle-pentagon coordinates combination), compared to 83.75% in the case of the Decision Tree using the 4th input vector (ROGER's four-ratio). These diagnostic results show an improvement in the identification of transformer faults over other traditional DGA methods. Significant differences in diagnostic accuracy were obtained when using the same classification algorithm with different input vectors, this investigation shows the appropriate input vector for the diagnosis of power transformers using the Naive Bayes and the Decision Tree algorithms. In a future work, we will extend the proposed input space using other input vectors with an improved machine learning algorithm.

REFERENCES

[1] S. S. M. Ghoneim, I. B. M. Taha, and N. I. Elkalashy, “Integrated ANN-based proactive fault diagnostic scheme for power transformers using dissolved gas analysis”, IEEE Transactions on Dielectrics and Electrical Insulation, vol. 23, no. 3, pp. 1838–1845, Jun. 2016.

[2] ‘’Mineral Oil-Filled Electrical Equipment in Service - Guidance on the Interpretation of Dissolved and Free Gases Analysis’’, IEC Standard IEC 60599, IEC, Geneva, Switzerland, Edition 2.1, May 2007.

[3] F. Jakob and J. J. Dukarm “Thermodynamic estimation of transformer fault severity”, IEEE Trans. on Power Delivery, vol. 30, no. 4, pp. 1941–1948, 2015. DOI: 10.1109/TPWRD.2015.2415767

[4] I. B. M. Taha, S. S. Dessouky and S. S. M. Ghoneim “Transformer fault types and severity class prediction based on neural pattern-recognition techniques”, Electric Power Systems Research, vol. 191, no. 7, 9, pp. 106899, 2021.

[5] A. Hoballah, D. A. Mansour and I. B. M. Taha “Hybrid grey wolf optimizer for transformer fault diagnosis using dissolved gases considering uncertainty in measurements”, IEEE Access, vol. 8, pp. 139176–139187, 2020.

[6] M. A. Izzularab, G. E. M. Aly and D. A. Mansour “On-line diagnosis of incipient faults and cellulose degradation based on artificial intelligence methods”, IEEE Int. Conf. on Solid Dielectrics, Toulouse, France, 2004.

[7] M. Duval and A. dePabla “Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases”, IEEE Electrical Insulation Magazine, vol. 17, no. 2, pp. 31–41, Mar. 2001.

[8] “Ieee guide for the interpretation of gases generated in oil-immersed transformers”, IEEE Std C57.104-2008 (Revision of IEEE Std C57.104-1991), pp. 1–36, 2009.

[9] M. Duval “A review of faults detectable by gas-in-oil analysis in transformers”, IEEE Electrical Insulation Magazine, vol. 18, no. 3, pp. 8–17, May 2002. DOI: 10.1109/MEI.2002.1014963

[10] N. A. Bakar and A. Abu-Siada “Fuzzy logic approach for transformer remnant life prediction and asset management decision”, IEEE Transactions on Dielectrics and Electrical Insulation, vol. 23, no. 5, pp. 3199–3208, Oct. 2016.

[11] 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, Oct. 2016.

[12] I. B. M. Taha, S. S. M. Ghoneim, and A. S. A. Duaywah ‘‘Refining DGA methods of IEC code and Rogers four ratios for transformer fault diagnosis’’, IEEE Power Energy Soc. Gen. Meeting (PESGM), pp. 1–5, Jul. 2016

[13] S.M. Islam, T. Wu, G. Ledwich “A novel fuzzy logic approach to transformer fault diagnosis’’. IEEE Trans. Dielectr. Electr. Insul, 7, 177–186, 2000

[14] S. Souahlia, K. Bacha and A. Chaari “MLP neural network-based decision for power transformers fault diagnosis using an improved combination of Rogers and Doernenburg ratios DGA”, Int. Journal of Electrical Power & Energy Systems, vol. 43, no. 1, pp. 1346–1353, 2012

[15] V. Miranda, A.R.G. Castro ‘’Improving the IEC table for transformer failure diagnosis with knowledge extraction from neural networks’’. IEEE Trans. Power Deliv., 20, 2509–2516, 2005

[16] Y. Benmahamed, M. Teguar, A. Boubakeur ‘’Application of SVM and KNN to Duval Pentagon 1 Transformer Oil Diagnosis’’. IEEE Trans. Dielect. Electr. Inst., 24, 3443–3451, 2017

[17] K. Bacha, S. Souahlia and M. Gossa “Power transformer fault diagnosis based on dissolved gas analysis by support vector machine”, Electrical Power Systems Research, vol. 83, no. 1, pp. 73–79, 2012.

[18] H. Zheng, Y. Zhang, J. Liu, H. Wei, J. Zhao, R. Liao ‘’A novel model based on wavelet LS-SVM integrated improved PSO algorithm for forecasting of dissolved gas contents in power transformers’’. Electr. Power Syst. Res., 155, 196–205, 2018

[19] Y. Zhang, H. Wei, R. Liao, Y. Wang, L. Yang, 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. 2017, 12, 830–839.

[20] O. Kherif, Y. Benmahamed, M. Teguar, A. Boubakeur and S. S. M. Ghoneim “Accuracy Improvement of Power Transformer Faults Diagnostic Using KNN Classifier With Decision Tree Principle” in IEEE Access, vol. 9, pp. 81693-81701, 2021.

[21] J. I. Aizpurua, V. M. Catterson, B. G. Stewart, S. D. J. McArthur, B. Lambert, B. Ampofo, G. Pereira, and J. G. Cross ‘’Power transformer dissolved gas analysis through Bayesian networks and hypothesis testing’’, IEEE Trans. Dielectrics Electr. Insul., vol. 25, no. 2, pp. 494–506, Apr. 2018. DOI: 10.1109/TDEI.2018.006766

[22] M. T. Yang, L. S. Hu ‘’Intelligent fault types diagnostic system for dissolved gas analysis of oil-immersed power transformer”. IEEE Trans. Dielectr. Electr. Insul., 20, 2317–2324, 2013

[23] J. Dai, H. Song, G. Sheng, X. Jiang ‘’Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network”. IEEE Trans. Dielectr. Electr. Insul., 24, 2828–2835, 2017

[24] P. Mirowski, Y. LeCun ‘’Statistical Machine Learning and Dissolved Gas Analysis: A Review’’. IEEE Trans. Power Deliv., 27, 1791–1799, 2012. DOI: 10.1109/TPWRD.2012.2197868

[25] P. Mirowski and Y. LeCun “Statistical Machine Learning and Dissolved Gas Analysis: A Review”, IEEE Trans. Power Del., vol. 27, no. 4, pp.1791–1799, Oct. 2012.

[26] N. Abu Bakar, A. Abu-Siada, and S. Islam “A Review of Dissolved Gas Analysis Measurement and Interpretation Techniques”, IEEE Electr. Insul. Mag., vol. 30, no. 3, pp. 39–49, May-June 2014.

[27] I.B.M. Taha, A. Hoballah, S.S.M. Ghoneim ‘’Optimal ratio limits of Roger’s four-ratios and IEC 60599 code methods using particle swarm optimization fuzzy-logic approach’’. IEEE Trans. Dielect. Electr. Inst., 27, 222–230, 2020. DOI: 10.1109/TDEI.2019.008395

[28] M. S. Mücahid, Y. Ali ‘’Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification’’. Intern. J. of Intelligent Systems and Applications in Engineering, IJISAE, 7(2), 88–91, 2019

[29] G. Dimitoglou, J. A. Adams and C. M. Jim ‘’Comparison of the C4. 5 and a Naïve Bayes classifier for the prediction of lung cancer survivability’’. arXiv preprint arXiv:1206.1121, 2012

[30] J. R. Quinlan ‘’Induction of decision trees’’. Mach Learn 1(1):119–127, 1986.

[31] A. Alqudsi and A. El-Hag “Application of machine learning in transformer health index prediction”, Energies, vol. 12, no. 14, pp. 2694, 2019. DOI: 10.3390/en12142694

[32] J. R. Quinlan “C4.5: Programs for Machine Learning”, Morgan Kaufmann Publishers, 1993

[33] J. Soni and D. Suthar “An Experimental Analysis to Check Accuracy of DGA Using Duval Pentagonal Method in Power Transformer”, ICRISET2017 the Int. Conf. on Research and Innovations in Science, Engineering and Technology, vol 1, pp. 394--401, Aug. 2017.

[34] J. Dukarm and F. Jakob “Thermodynamic estimation of transformer fault severity” IEEE/PES Transmission and Distribution Conference and Exposition (T&D), Dallas, TX, USA, pp. 1-1, 2016

Yassine Mahamdi received the Master’s and Engineer's degrees in Electrical Engineering from Ecole Nationale Polytechnique (ENP) of Algiers in 2018. He is currently a Ph.D. student at the Electrical Engineering Department in the same School. His field of research is based on transformer diagnostics and artificial intelligence.

Ahmed Boubakeur received in 1975 the degree of Engineer in electrical engineering from Ecole Nationale Polytechnique (ENP) of Algiers, and in 1979 he obtained the Doctorate in Technical Sciences from the Institute of High Voltage Engineering of the Technical University of Warsaw of in Poland. He is currently a professor at ENP of Algiers where he has been giving lectures and supervising research in the field of High Voltage Engineering since 1982. His principal research areas are discharge phenomena, insulators pollution, lightning, polymeric cables insulation, transformer oil ageing, neural network and fuzzy logic application in HV insulation diagnosis, and electric field calculation and measurement. He is an IEEE senior member, member of IEEE/DEIS and a member of the Algerian HV Power Systems Association ARELEC (National Algerian Comity of CIGRE and ENP Elders Association ADEP). He has been member of the Editorial Board and Associate Editor of IET/SMT.

Abdelouahab Mekhaldi received the degree of Engineer in 1984 in electrical engineering, the MSc degree in 1990 and a Ph.D. in high voltage engineering in 1999 from Ecole Nationale Polytechnique (ENP) of Algiers. He is currently a Professor at ENP. His main research areas are in discharge phenomena, outdoor insulators pollution, polymeric cables insulation, lightning, artificial intelligence application in high voltage insulation diagnosis and electric field calculation.

Youcef Benmahamed received the degree of Engineer and Master’s degree in power electronics engineering in 2014 and Ph.D in High Voltage techniques in 2019 from Ecole Nationale Polytechnique (ENP) of Algiers. His research interests are in diagnosis, artificial intelligence and optimization techniques.