



Investigation on machine learning algorithms to support transformer dissolved gas analysis fault identification

Ekojono¹ · Rahman Azis Prasajo² · Meyti Eka Apriyani¹ · Anugrah Nur Rahmanto¹

Received: 23 August 2021 / Accepted: 19 February 2022 / Published online: 21 March 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Dissolved gas analysis (DGA) is a powerful tool to monitor the condition of a power transformer. Several interpretation methods have been proposed, one of the most reliable of which is the graphical Duval triangle method (DTM). The method consists of several triangles, which still requires expertise for fault identification. The use of computer-based technology has been implemented in recent years to support transformer fault identification. However, no study has done thorough investigation on the use of suitable machine learning algorithm for the ML-based implementation of this matter. This study examines six commonly used machine learning algorithms to support DGA fault identification of power transformer: decision tree, support vector machine, random forest (RF), neural network, Naïve Bayes, and AdaBoost. Three DGA fault identification methods for mineral oil insulated transformer were studied, namely DTM1, DTM4, and DTM5. The training and testing datasets were generated for each DGA method, and trained to each ML algorithm. The tenfold cross validation was used to evaluate the results using five criteria, namely classification accuracy, area under curve, F1, Precision, and Recall. RF models demonstrated the best performance in classifying fault codes of most DGA methods. A validation was carried out using the validation dataset, comparing the selected RF-based models to the graphical DGA fault identification. The combination method was also implemented in the developed model. The results show that the proposed model is reliable, and especially useful to be used for fault identification of a large number of transformer populations.

Keywords Dissolved gas analysis · Power transformers · Machine learning · Fault identification

1 Introduction

Fault identification is necessary for fault diagnosis of a power transformer as it plays an important role in the electrical power distribution. Dissolved gas analysis is a powerful tool to diagnose a power transformer through oil measurement. It is considered as the most reliable technique in detecting

incipient fault within power transformers [1]. There are several DGA interpretation methods that have been proposed and commonly used, such as the Key Gas Method, Rogers Ratio, Doenenburg Ratio, IEC Ratio, Duval Triangle Method, and Duval Pentagon Method. Among the existing interpretation methods, Duval Triangle Method provides the most accurate and consistent analysis [2]. Like previous studies, research [3] confirmed that Duval Triangle resulted in the most similar results from the transformer observation. Duval Triangle Method, along with the pentagon version, is the best methods for more detailed faults diagnosis in transformers [4].

The use of computer-based technology in recent years has led to significant improvement in transformer condition assessment and monitoring. A fuzzy logic approach has been proposed for consistent interpretation of dissolved gas-in-oil analysis [5]. Support Vector Machine algorithm was developed and proposed to diagnose fault in power transformer in [6, 7]. Random Forest model was proposed to predict missing interfacial tension parameter in [8]. Neural

✉ Rahman Azis Prasajo
rahmanazisp@polinema.ac.id

Ekojono
ekojono@polinema.ac.id

Meyti Eka Apriyani
meytieka@polinema.ac.id

Anugrah Nur Rahmanto
anugrahnur@polinema.ac.id

¹ Information Technology Department, Politeknik Negeri Malang, Malang 65141, Indonesia

² Electrical Engineering Department, Politeknik Negeri Malang, Malang 65141, Indonesia

network was proposed to diagnose fault in power transformer based on IEC code in [9]. Multiple machine learning classifier was implemented and compared to predict Health Index of power transformer in [10]. In recent years, the graphical or chart-based interpretation has been improved by computer-based technology [11]. A study of [12] proposed a new fault diagnostic based on the Duval Triangle which employs the combination of fuzzy and evidential reasoning techniques. This study aims to improve the Duval Triangle Method so that it can produce easy to understand approach and is able to identify simultaneous faults. A study of [13] developed Java language based software implementation of the Duval Triangle 1. A study of [14] proposed the ANFIS-based Duval Triangle model to for fault diagnosis of power transformer. The proposed ANFIS model produced higher accuracy compared to the ANN algorithm.

The Machine Learning is useful on power transformer assessment as it reduces the dependency on personnel expertise, and improve consistency. Moreover, it is useful to assess a large number of transformers, which consists of hundreds to thousands of transformers data. However there has been no thorough investigation on the use of the suitable machine learning algorithm for the ML-based implementation of Duval Triangle. Previous studies has also done mostly on Duval Triangle 1, while based on the result in [4], other Duval Triangle in combination can improve the consistency of the result. Therefore, this study aims to investigate and compare various popular ML algorithms, and propose the best performing model to support graphical DGA interpretation. The ML-based Duval Triangle 1, 4, and 5 were proposed, and the combination method were implemented. The evaluation of the proposed method was done using the validation dataset.

2 Methodology

The development of ML-based DTM was initially done by generating training and testing datasets for DTM1, DTM4, and DTM5. Subsequently, six ML algorithms were trained, namely decision tree (DT), support vector machine (SVM), random decision tree (RF), neural network (NN), Naïve Bayes (NB), and AdaBoost (AB). To evaluate the performance of the models, several parameters were used, such as Classification Accuracy, Area Under Curve, F1, Precision, and Recall. Confusion matrices were used to visualize the results. The developed model with the highest performance would be selected. In addition, validation dataset, which was obtained from the previous published articles, was used to evaluate the proposed ML-based DGA Interpretation to the graphical one. Finally, a combination method was implemented in the ML-based DTM1, DTM4, and DTM5. Figure 1 shows the flowchart of the methodology in this study.

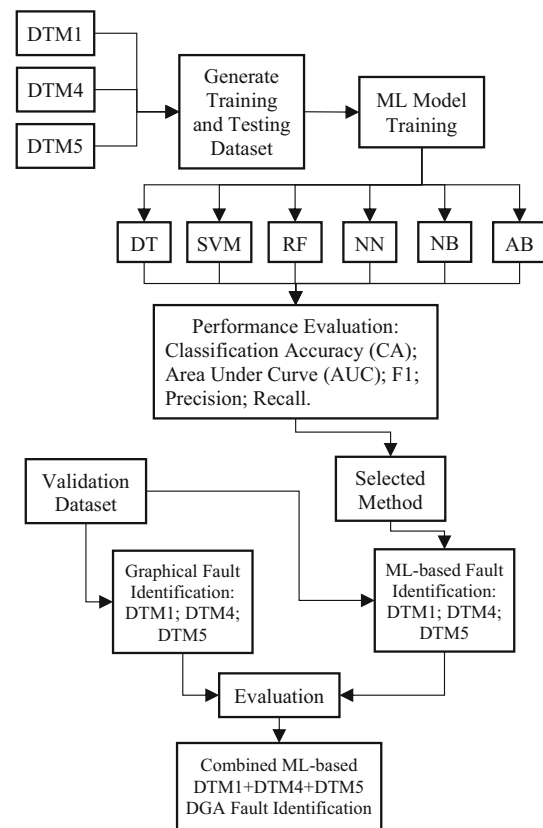


Fig. 1 Flowchart of the methodology

2.1 Graphical dissolved gas analysis interpretation

Duval triangle method (DTM) is presented in [15] consists of several triangles. It is still considered the most reliable method to identify fault within power transformers [1–4, 6, 7]. Most in-service power transformers are still using mineral type oil insulation, and the DTM that is suitable for those transformers are DTM1, DTM4, and DTM5. Therefore, those three DGA fault identification methods were investigated in this study.

A. Duval Triangle Method 1

DTM1 uses the percentage of three gasses: CH₄, C₂H₄, and C₂H₆. Table 1 shows the seven fault codes of DTM1, which are PD, D1, D2, T1, T2, T3, and DT. Figure 2 shows the graphical representation of DTM1.

B. Duval Triangle Method 4

The DTM4 uses the percentage of three gasses: H₂, CH₄, and C₂H₆. Table 2 shows the five fault codes of DTM4. DTM4 is an addition to the DTM1, and was used to obtain more information related to faults within low temperature (PD, T1, or T2) after DTM1 [15]. If the results of DTM1 is D1, D2, or T3, DTM4 should not be used. Mixture of faults

Table 1 Fault codes and the description of DTM1

Code	Description
PD	Corona partial discharges
D1	Electrical discharges of low energy
D2	Electrical discharges of high energy
T1	Thermal faults of temperature $T < 300$ C
T2	Thermal faults, $300\text{ C} < T < 700$ C
T3	Thermal faults, $T > 700$ C
DT	Mixtures of electrical and thermal faults

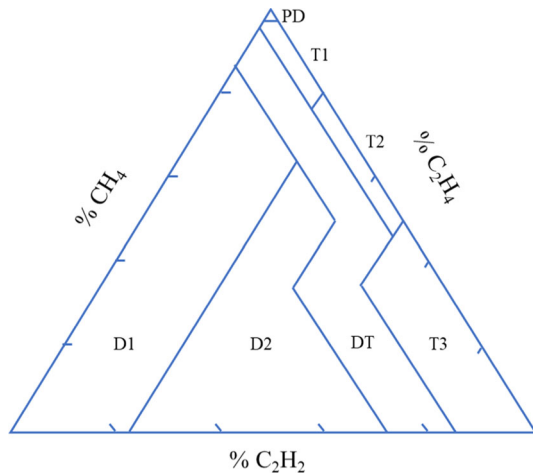


Fig. 2 Duval triangle 1 dissolved gas fault identification

Table 2 Fault code and the description of DTM4

Code	Description
PD	Corona partial discharges
S	Stray gassing of mineral oil ($T < 200$ C)
C	Hot spots with carbonization of paper ($T > 300$ C)
O	Overheating ($T < 250$ C)
N/D	Not determined

can be indicated by the DTM4 and DTM 5 that do not agree. Figure 3 shows the graphical representation of DTM4.

C. Duval Triangle Method 5

The DTM5 uses the percentage of three gasses: CH4, C2H4, and C2H6. Table 3 shows the seven fault codes of DTM5. DTM5 was employed to obtain more information after DTM1 identified T2 or T3. DTM5 should not be used when the DTM1 indicated D1 or D2 faults. DTM5 could be used to confirm the uncertainty of the results of DTM1 and DTM4 [15, 16]. Mixture of faults can be indicated by DTM4 and

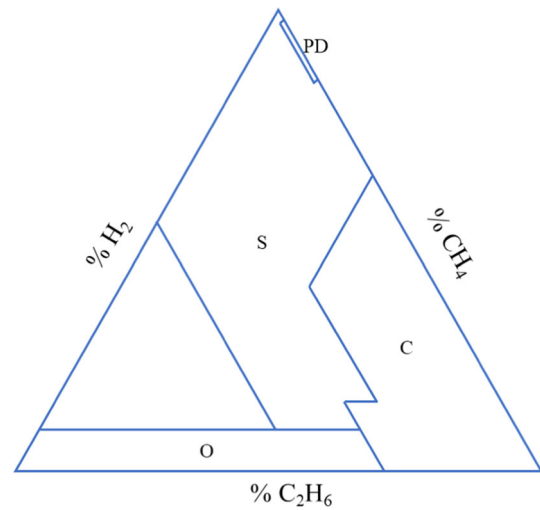


Fig. 3 Duval triangle 4 dissolved gas fault identification

Table 3 Fault codes and the description of DTM5

Code	Description
PD	Corona partial discharges
S	Stray gassing of mineral oil ($T < 200$ C)
C	Hot spots with carbonization of paper ($T > 300$ C)
O	Overheating ($T < 250$ C)
T2	Thermal faults, $300\text{ C} < T < 700$ C
T3	Thermal faults, $T > 700$ C
N/D	Not determined

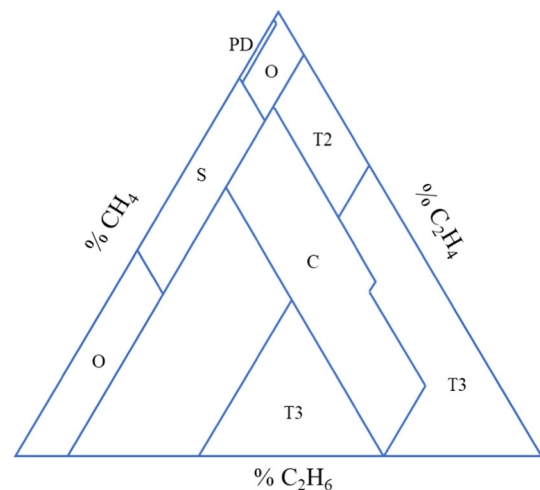


Fig. 4 Duval triangle 5 dissolved gas fault identification

DTM 5 that do not agree. Figure 4 shows the graphical representation of DTM5.

2.2 Machine learning algorithms

The use of ML to assist power transformer condition assessment has been reported in several literatures [17–19]. This study investigates six different machine learning algorithms to support the graphical dissolved gas analysis fault identification method. Those algorithms are Decision Tree, Support Vector Machine, Random Forest, Neural Network, Naïve Bayes, and AdaBoost.

A. Decision Tree

Decision tree is one of the most commonly used algorithms in model building. It has a strong generalization ability and convenient pruning, and it is fast fitting [20]. In this algorithm, classification is carried out by splitting the data into nodes by class purity using information gain. Several studies have implemented Decision Tree into power transformer condition monitoring and diagnostics [20–22]. A study of [21] presented the use of C4.5 decision tree to predict transformer fault from gas values of online condition monitoring. A study of [22] implemented decision tree to classify the frequency response analysis to diagnose fault within transformer windings. A study of [20] developed the rules of the fuzzy logic using decision tree algorithm to estimate the degree of polymerization in oil-paper insulation system.

B. Support Vector Machine

Many studies have successfully implemented Support Vector Machine in power transformer condition monitoring and diagnostics [23–26]. This algorithm transforms the samples in the original input space to a higher dimensional space, then searches for the optimal separation hyperplane [23]. SVM is important because of its linearity and flexibility for large data setting. It has good generalization properties because it minimizes the structural misclassification risk in the training process [25, 26].

C. Random Forest

Random forest classifier is an ensemble learning method which consists of a set of decision tree that is developed from a bootstrap sample from the training data. This algorithm consists of a collection of tree-structured classifiers with independent identically distributed random vectors. Each tree casts a unit vote for choosing the most popular class at input [27]. An ensemble of the classifiers is given, and the training data are drawn randomly from the distribution of the random vector and the margin function is defined. Random forests do not overfit as more trees are added and produce a limiting value for the generalization error [27]. Several studies has

proposed the successful use of Random Forest classifier [10, 28, 29].

D. Neural Network

Neural Network is one of the most widely used algorithms to build predictive models. This algorithm was inspired by the functionality of human brain which consists of neurons to process information in parallel [30]. This approach can reveal highly nonlinear input–output relationships and acquire knowledge directly from the training data through a learning process [31].

E. Naïve Bayes

This algorithm uses Bayes' rule with assumption that the attributes are conditionally independent given the class. Naïve Bayes uses the information in the sample to estimate the posterior probability $P(y|x)$ for each class y given an object x . Classification is done once such estimates are obtained [32].

F. AdaBoost

Adaptive Boosting (AdaBoost) is an approach to improve the prediction accuracy in machine learning by combining many relatively weak and inaccurate rules [33]. Several studies implementing AdaBoost in power transformer condition monitoring and diagnostics are [34, 35].

The performance evaluation of each model is described in the following section.

2.3 Evaluation

To evaluate the performance of the proposed method, several parameters were compared.

- *Classification accuracy (CA)* This CA value is the accuracy of the classification compared to the target category. The highest CA is 1, and the lowest is 0. CA of 1 means that all of the classification results correspond to the actual category. CA was calculated using (1).

$$CA = \frac{tp + tn}{(tp + tn + fp + fn)} \quad (1)$$

- *Precision* This value measures the ratio of correctly predicted positive observations (tp) to the total predicted positive observations (tp + fp). Precision was calculated using (2).

$$Precision = \frac{tp}{(tp + fp)} \quad (2)$$

Table 4 Number of samples for DTM1 training data

Fault type	<i>n</i>	Fault type	<i>n</i>
PD	26	T2	48
D1	520	T3	108
D2	400	DT	270
T1	30		
		Total	1402

- *Recall* This measures the ratio of correctly predicted positive observations (tp) to the entire observations in the actual class (tp + fn). Recall was calculated using (3).

$$\text{Recall} = \frac{\text{tp}}{(\text{tp} + \text{fn})} \tag{3}$$

- *F1* This value shows the weighted average of Recall and Precision. The best value of F1 is 1, and the worst is 0. Typically, the highest F1 possible is desired. F1 was calculated using (4).

$$F1 = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{4}$$

- *Area under curve (AUC)* AUC value is the measure of classifier ability to differentiate between classes. The higher the value, the better the model performance in classifying different classes.
- *Confusion Matrix* This table visualizes the performance of the proposed model to do classification. In this paper, each row represents the actual class from graphical DTM, and each column represents the predicted class from ML-based DTM.

2.4 Datasets

Two types of datasets were used in this study. The first one was the DGA fault identification training datasets, which were generated to train and test the ML-based DTM. The second was the validation datasets, which was used to evaluate the proposed model and compare the fault identification results to the graphical DTM.

A. DGA Fault Identification Training Datasets

To train the ML-based DTM model, three training datasets were developed. Table 4 shows a total of 1402 data for DTM1, and the number of samples for each fault type. There were seven fault types as the target class in this dataset. Three features were used as input parameters, which are %CH4 (percentage of CH4), %C2H2, and %C2H4. The calculations

Table 5 Number of samples for DTM4 training data

Fault type	<i>n</i>	Fault type	<i>n</i>
PD	26	O	215
S	610	N/D	301
C	403		
		Total	1555

Table 6 Number of samples for DTM5 training data

Fault type	<i>n</i>	Fault type	<i>n</i>
PD	26	T2	91
S	126	T3	444
C	268	N/D	325
O	168		
		Total	1448

of the gas percentage are shown in Eqs. 1–3. These calculations were also carried out with DTM4 and DTM5 dataset, using different gasses combinations.

$$\%CH4 = \frac{CH4_{(ppm)} \times 100\%}{CH4_{(ppm)} + C2H2_{(ppm)} + C2H4_{(ppm)}} \tag{5}$$

$$\%C2H2 = \frac{C2H2_{(ppm)} \times 100\%}{CH4_{(ppm)} + C2H2_{(ppm)} + C2H4_{(ppm)}} \tag{6}$$

$$\%C2H4 = \frac{C2H4_{(ppm)} \times 100\%}{CH4_{(ppm)} + C2H2_{(ppm)} + C2H4_{(ppm)}} \tag{7}$$

Table 5 shows a total of 1555 data for DTM1, and the number of samples for each fault type. There were five classes of fault types as the target in this dataset. Three features were used as input parameters, which are %CH4, %C2H4, and %C2H6. The calculations of the gas percentage were the same with the previously shown in Eqs. 5–7, but using different gasses combinations.

Table 6 shows a total of 1448 data for DTM5, and the number of samples for each fault type. There were seven classes of fault types as the target in this dataset. Three features that were used as input parameters are %CH4, %H2, and %C2H6.

B. Validation Dataset

As many as 1017 transformer dissolved gas data from previous studies [2, 4–7, 9, 19, 25, 36–72] were gathered and used to evaluate the proposed methods. All of the data contained the concentration of dissolved gas in mineral oil insulation of transformer main tank, with five gasses in ppm (parts per million). These data were used to validate the proposed model and compared to the result of graphical DTM.

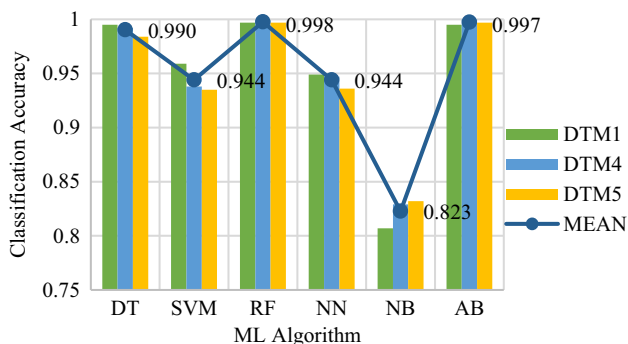


Fig. 5 Classification accuracy comparisons of seven ML algorithms

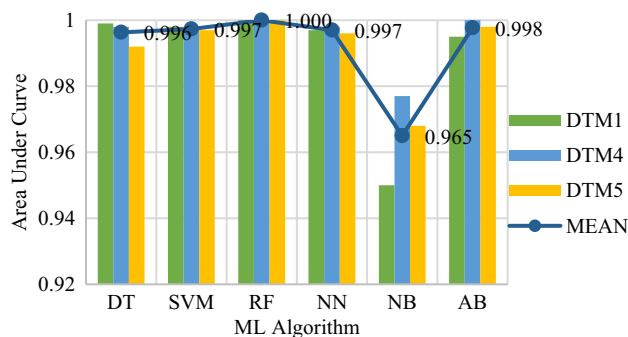


Fig. 6 Area under curve comparisons of seven ML algorithms

3 Results and analysis

This section discusses the results of the analysis. First, eighteen ML-based fault identifications were developed and compared. The best performing ML algorithm was selected. Subsequently, an evaluation was conducted to the proposed model using the validation dataset. After that, the combined ML-based DTM was implemented to prove the applicability of the proposed method to identify fault within the transformer using dissolved gas-in-oil data.

3.1 ML-based dissolved gas analysis interpretation

The datasets generated that are summarized in Tables 4, 5 and 6 are trained to six machine learning algorithms, namely DT (Decision Tree), SVM (Support Vector Machine), RF (Random Forest), NN (Neural Network, NB (Naïve Bayes), and AB (AdaBoost). The tenfold cross validation was employed to evaluate each of the trained model. Figure 5 shows the performance of each ML algorithm in classifying each DTM in terms of classification accuracy. The green bar shows the accuracy of ML-Based DTM1, the blue bar is for DTM4, and yellow bar is for DTM5. The dark blue dot shows the mean of accuracy for each ML algorithm in predicting fault types of DTM1, DTM4, and DTM5. RF resulted in the highest average accuracy of 0.998, followed by AdaBoost for 0.997. Decision Tree also obtain relatively high accuracy, 0.990, while Support Vector Machine and Neural Network obtain similar results of 0.944. In this case, Naïve Bayes got the lowest average of accuracy of 0.823.

Figures 6, 7, 8 and 9 show the comparison of area under curve (AUC), F1, Precision, and Recall of seven ML algorithms in classifying fault types of transformer dissolved gas analysis. Table 7 shows the performance of seven ML algorithms. Random Forest and AdaBoost performed better than other algorithms compared in this study, with Random Forest models slightly better. This result is consistent with all of the evaluation parameters. Therefore, RF models were proposed to be implemented further.

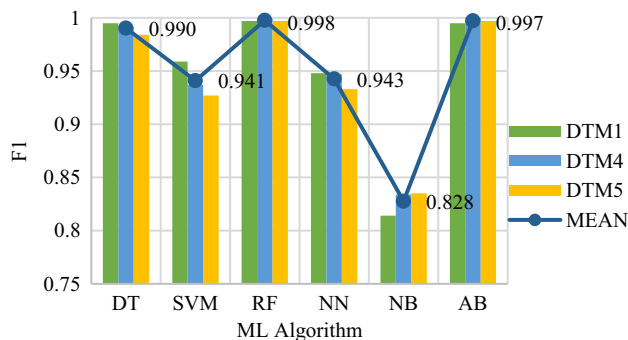


Fig. 7 F1 comparisons of seven ML algorithms

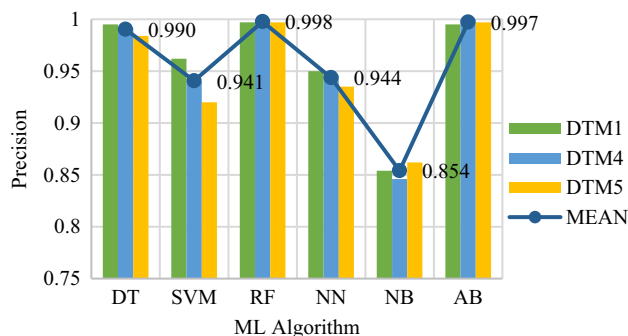


Fig. 8 Precision comparisons of seven ML algorithms

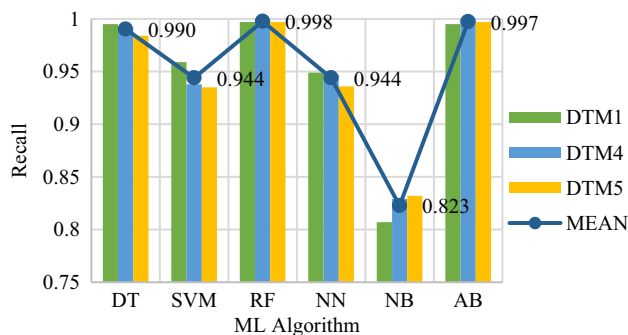


Fig. 9 Recall comparisons of seven ML algorithms

Table 7 Performance of seven ML algorithms

ML algorithm	AUC	CA	F1	PR	RC
DT	0.996	0.990	0.990	0.990	0.990
SVM	0.997	0.944	0.941	0.941	0.944
RF	1.000	0.998	0.998	0.998	0.998
NN	0.997	0.944	0.943	0.944	0.944
NB	0.965	0.823	0.828	0.854	0.823
AB	0.998	0.997	0.997	0.997	0.997

Table 8 Accuracy of RF-Based DTM on Validation Dataset

Model	No. of sample	Accuracy (%)
DTM1	1071	95.5
DTM4	1071	96.2
DTM5	1071	95.1

Table 9 Confusion matrix of RF-based DTM1 on validation dataset

		PREDICTED							True Positive	False Negative
		PD	D1	D2	T1	T2	T3	DT		
ACTUAL	PD	109	0	0	0	0	0	0	100%	0%
	D1	0	113	0	0	0	1	0	99%	1%
	D2	0	2	177	0	0	0	0	99%	1%
	T1	18	0	0	201	1	2	0	91%	9%
	T2	0	0	0	1	142	0	0	99%	1%
	T3	0	0	0	2	4	225	12	93%	7%
	DT	0	0	1	7	4	0	60	83%	17%

3.2 Evaluation

After developing the RF-based DTM1, DTM4, and DTM5, an evaluation was conducted using the validation dataset. As many as 1071 dissolved gas data were obtained and subsequently classified using the developed RF-based models. The results of the fault identification were then compared to the graphical DTM and analyzed. The resulting accuracy of RF-based DTM on the validation dataset is shown in Table 8. All of the models performed well and DTM4 obtained the highest accuracy of 96.2%.

Tables 9, 10 and 11 show the confusion matrix of RF-based DTM on the validation dataset. The rows of the confusion matrix shows the actual fault identifications by graphical DTM, while the columns shows the results of RF-based DTM in classifying the dissolved gasses into fault. True positive is a ratio of correctly classified sample per class, while false negative is the ratio of incorrectly classified sample per class.

Table 10 Confusion matrix of RF-based DTM4 on validation dataset

		PREDICTED					True Positive	False Negative
		PD	S	C	O	N/D		
ACTUAL	PD	42	2	0	0	0	95%	5%
	S	16	440	6	1	2	95%	5%
	C	1	10	292	0	0	96%	4%
	O	0	1	0	204	1	99%	1%
	N/D	1	1	0	1	61	95%	5%

Table 11 Confusion matrix of RF-based DTM5 on validation dataset

		PREDICTED							True Positive	False Negative
		PD	S	C	O	T2	T3	N/D		
ACTUAL	PD	24	0	0	0	0	0	0	100%	0%
	S	0	73	0	1	0	0	0	99%	1%
	C	0	0	194	0	3	4	1	96%	4%
	O	30	5	0	227	0	0	0	87%	13%
	T2	0	1	0	3	29	1	0	85%	15%
	T3	0	0	0	0	0	372	0	100%	0%
	N/D	0	4	2	1	0	0	107	94%	6%

The results shows that RF-based DTM1 is excellent in identifying partial discharge fault type, with 100% true positive rate. Satisfactory results were also shown in D1, D2, and T2 fault types with 99% true positive rate. Meanwhile, DT fault type had 17% inaccuracy due to the fact that DT is a mixture of electrical and thermal faults.

Table 10 shows the confusion matrix of RF-based DTM4 on the validation dataset. This model was used to obtain more information related to faults within low temperature (PD, T1, or T2). The results show that RF-based DTM4 is excellent in identifying overheating fault under 250C with 99% accuracy. Other fault codes for this model were also good, as the accuracy ranges from 95 to 96%. Table 11 shows the confusion matrix of RF-based DTM5. This model was employed after

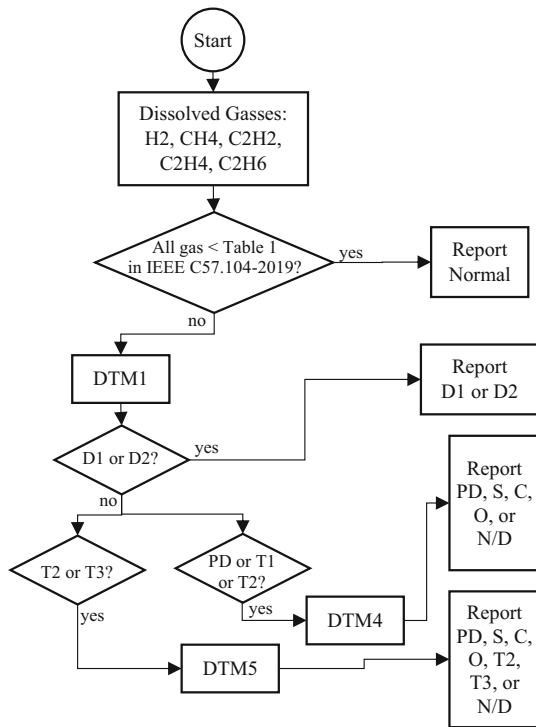


Fig. 10 Flowchart of DGA Fault identification using combination of DTM1, DTM4, and DTM5

DTM1 identified T2 or T3. The results show that DTM5 was excellent in identifying PD and T3. However, the accuracy of 87% and 85% were obtained when identifying O and T2, respectively.

3.3 Combined method

The combination of DTM1, DTM4, and DTM5 is useful to distinguish further the faults inside the transformer besides electrical faults of D1 and D2. When low energy or low temperature faults were identified using the DTM1 (PD, T1 or T2), DTM4 was used to obtain more information. When high, or very high, temperature faults were identified with DTM1 (T2 or T3), DTM5 was used to obtain more information. DTM4 distinguished between relatively minor faults such as S, O, PD, and potentially more dangerous faults C, which involved possible carbonization of paper. The DTM5 method could be used to distinguish between high temperature faults T3/T2 in mineral oil and potentially more dangerous faults C involving possible carbonization of paper [11].

The combination of RF-based DTM1, DTM4, and DTM5 in fault identification was implemented in this study. The use of combined Duval Triangle could improve the consistency of the results [4]. Figure 10 shows the flowchart of DGA interpretation using the combination of DTM1, DTM4, and DTM5. These steps were applied to the 1071 dissolved gas

Table 12 Confusion matrix of combined RF-based DTM1,4, and 5 on validation dataset

		PREDICTED									
		Normal	PD	S	C	O	T2	T3	D1	D2	N/D
ACTUAL	Normal	207	0	0	0	0	0	0	0	0	0
	PD	0	25	0	0	0	0	0	0	0	0
	S	0	5	97	0	1	0	0	0	0	0
	C	0	0	5	139	0	0	2	0	0	0
	O	0	0	0	1	125	0	0	0	0	0
	T2	0	0	0	1	0	4	0	0	0	0
	T3	0	0	0	0	0	0	149	0	0	0
	D1	0	0	0	0	0	0	1	111	0	0
	D2	0	0	0	0	0	0	0	2	176	0
	N/D	0	1	1	0	1	0	0	0	1	27

Table 13 Total accuracy of the proposed RF-based combination method in DGA Fault Identification

DGA interpretation	True positive (%)	False negative (%)
Normal	100	0
PD	100	0
S	94	6
C	95	5
O	99	1
T2	80	20
T3	100	0
D1	99	1
D2	99	1
N/D	87	13
Total accuracy	98.7	

dataset, using graphical DTM and RF-based DTM. The comparison was carried out to evaluate the use of the proposed Random Forest model.

Table 12 shows the confusion matrix of the proposed RF-based combined DTM on the validation dataset. The rows represent the actual DGA interpretation using graphical DTM and the columns represent the predicted category using combined RF-based DTM. The results show that the proposed method has high performance when evaluated with the validation dataset. As many as ten categories were observed, with the total accuracy of 98.7% as shown in Table 13. The resulting method performed well and useful to support graphical DGA fault identification, especially on large number of transformers.

4 Conclusion

This study investigates various machine learning algorithms to support dissolved gas analysis fault identification of power transformer. Duval Triangle Method 1, 4, and 5 were employed as they were the most commonly used and reliable DGA methods. Three datasets were developed to be trained to six different machine learning algorithms. The results show that Random Forest performed the best compared to the others. Subsequently, the resulting RF-models were evaluated using the validation dataset. As many as 1071 data of dissolved gas in transformers were sampled from previous researches and observed in this study. The resulting accuracy of RF-based model for DTM1 is 95.5%, DTM4 is 96.2%, and DTM5 is 95.1%. Then, the combination of DTM1, DTM4, and DTM5 was implemented. The results show that the developed RF-based method performed satisfactorily with 98.7% accuracy. The proposed model is reliable and especially useful to help asset managers in assessing a large number of transformers data.

Funding Funding was provided by Politeknik Negeri Malang (Grant No. SP DIPA 023.18.2.677606/2021).

References

- Abu-Siada A (2019) Improved consistent interpretation approach of fault type within power transformers using dissolved gas analysis and gene expression programming. *Energies* 12:730. <https://doi.org/10.3390/en12040730>
- Bakar NA, Abu-Siada A, Cui H, Li S (2017) Improvement of DGA interpretation using scoring index method. In: ICEMPE 2017—1st international conference on electrical materials and power equipment. <https://doi.org/10.1109/ICEMPE.2017.7982139>
- Pattanadech N, Sasomponsawatline K, Siriworachanyadee J, Angusatira W (2019) The conformity of DGA interpretation techniques: experience from transformer 132 units. In: 2019 IEEE 20th international conference on dielectric liquids, pp 1–4. <https://doi.org/10.1109/icdl.2019.8796588>
- Faiz J, Soleimani M (2017) Dissolved gas analysis evaluation in electric power transformers using conventional methods a review. *IEEE Trans Dielectr Electr Insul* 24:1239–1248. <https://doi.org/10.1109/TDEI.2017.005959>
- Abu-Siada A, Hmood S, Islam S (2013) A new fuzzy logic approach for consistent interpretation of dissolved gas-in-oil analysis. *IEEE Trans Dielectr Electr Insul* 20:2343–2349. <https://doi.org/10.1109/TDEI.2013.6678888>
- Shang H, Xu J, Zheng Z et al (2019) 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*. <https://doi.org/10.3390/en12204017>
- Zeng B, Guo J, Zhu W et al (2019) A transformer fault diagnosis model based on hybrid grey wolf optimizer and LS-SVM. *Energies*. <https://doi.org/10.3390/en12214170>
- Prasojo RA, Suwarno, Abu-Siada A (2021) Dealing with data uncertainty for transformer insulation system health index. *IEEE Access* 9:1–1. <https://doi.org/10.1109/access.2021.3081699>
- Ghoneim SSM, Taha IB (2015) Artificial neural networks for power transformers fault diagnosis based on iec code using dissolved gas analysis. *Int J Control Autom Syst* 4:18–21
- Rediansyah D, Prasojo RA, Suwarno, Abu-Siada A (2021) Artificial intelligence-based power transformer health index for handling data uncertainty. *IEEE Access* 9:150637–150648. <https://doi.org/10.1109/access.2021.3125379>
- IEEE Std C57.104–2019 (2019) IEEE guide for the interpretation of gases generated in mineral oil-immersed transformers
- Irungu GK, Akumu AO, Munda JL (2016) A new fault diagnostic technique in oil-filled electrical equipment; the dual of Duval triangle. *IEEE Trans Dielectr Electr Insul* 23:3405–3410. <https://doi.org/10.1109/TDEI.2016.005927>
- Akbari A, Setayeshmehr A, Borsi H, Gockenbach E (2008) A software implementation of the Duval triangle method. In: Conference record of the 2008 IEEE international symposium on electrical insulation, pp 124–127. <https://doi.org/10.1109/ELINSL.2008.4570294>
- Geetha R (2016) Fault diagnosis of power transformer using duval triangle based artificial intelligence techniques, pp 78–88
- Duval M (2008) The Duval triangle for load tap changers, non-mineral oils and low. *IEEE Electr Insul Mag* 24:22–29
- Aciu AM, Nicola CI, Nicola M, Nițu MC (2021) Complementary analysis for DGA based on Duval methods and furan compounds using artificial neural networks. *Energies* 14:1–22. <https://doi.org/10.3390/en14030588>
- Prasojo RA, Diwyacitta K, Suwarno, Gumilang H (2017) Transformer paper expected life estimation using ANFIS based on oil characteristics and dissolved gases (Case study: Indonesian transformers). *Energies*. <https://doi.org/10.3390/en10081135>
- Ghoneim SSM, Taha IBM, Elkalashy NI (2016) Integrated ANN-based proactive fault diagnostic scheme for power transformers using dissolved gas analysis. *IEEE Trans Dielectr Electr Insul* 23:1838–1845. <https://doi.org/10.1109/TDEI.2016.005301>
- Forouhari S, Abu-Siada A (2018) Application of adaptive neuro fuzzy inference system to support power transformer life estimation and asset management decision. *IEEE Trans Dielectr Electr Insul* 25:845–852. <https://doi.org/10.1109/TDEI.2018.006392>
- Nasution EF, Prasojo RA, et al (2018) Estimation of the degree of polymerization (DP) in oil-paper insulation system based on dielectric and DGA data using fuzzy logic algorithm. In: 4th IEEE conference on power engineering and renewable energy, ICPERE 2018—proceedings, pp 1–6. <https://doi.org/10.1109/ICPERE.2018.8739284>
- Basuki A (2018) Online dissolved gas analysis of power transformers based on decision tree model. In: 4th IEEE conference on power engineering and renewable energy, ICPERE 2018—proceedings
- Li Z, Zhang Y, Abu-Siada A et al (2021) Fault diagnosis of transformer windings based on decision tree and fully connected neural network. *Energies*. <https://doi.org/10.3390/en14061531>
- Ashkezari AD, Ma H, Saha T, Ekanayake C (2013) Application of fuzzy support vector machine for determining the health index of the insulation system of in-service power transformers. *IEEE Trans Dielectr Electr Insul* 20:965–973. <https://doi.org/10.1109/TDEI.2013.6518966>
- Liu J, Zhao Z, Tang C et al (2019) Classifying transformer winding deformation fault types and degrees using FRA based on support vector machine. *IEEE Access* 7:112494–112504. <https://doi.org/10.1109/access.2019.2932497>
- Prasojo RA, Suwarno (2018) Power transformer paper insulation assessment based on oil measurement data using SVM-classifier. *Int J Electr Eng Informatics* 10:661–673. <https://doi.org/10.15676/ijeei.2018.10.4.4>
- Bigdeli M, Vakilian M, Rahimpour E (2012) Transformer winding faults classification based on transfer function analysis by support

- vector machine. *IET Electr Power Appl* 6:268–276. <https://doi.org/10.1049/iet-epa.2011.0232>
27. Breiman L (1999) Random forests—random features, Technical Report 567, Statistic Department, University of California, Berkeley, (<https://www.stat.berkeley.edu/~breiman/random-forests.pdf>, 08.10.2018'de erişildi), pp 1–29
 28. Kartojo IH, Wang Y-B, Zhang G-J (2019) Partial discharge defect recognition in power transformer using random forest. In: 2019 IEEE 20th international conference on dielectric liquids (ICDL). IEEE, pp 1–4
 29. Shah AM, Bhalja BR (2016) Fault discrimination scheme for power transformer using random forest technique. *IET Gener Transm Distrib* 10:1431–1439. <https://doi.org/10.1049/iet-gtd.2015.0955>
 30. Wang S-C (2003) Artificial neural network. In: *Interdisciplinary computing in java programming*. In: The Springer international series in engineering and computer science
 31. Wang Z, Liu Y, Griffin PJ (2000) Neural net and expert system diagnose transformer faults. *IEEE Comput Appl Power* 13:50–55. <https://doi.org/10.1109/67.814667>
 32. Webb GI (2017) Naive Bayes. In: *Encyclopedia of machine learning and data mining*
 33. Schapire RE (2013) Explaining adaboost. *Empir Inference*
 34. Zhang C, Hu C, Zhang Z, Cao S (2020) Research on main transformer defect detection methods based on Conditional Inference Tree and AdaBoost Algorithm. In: *The 2nd international seminar on computer science and engineering technology (SCSET)*
 35. Yan C, Li M, Liu W (2019) Transformer fault diagnosis based on BP-Adaboost and PNN series connection. *Math Probl Eng*
 36. Kim SW, Kim SJ, Seo HD et al (2013) New methods of DGA diagnosis using IEC TC 10 and related databases Part 1: application of gas-ratio combinations. *IEEE Trans Dielectr Electr Insul* 20:685–690. <https://doi.org/10.1109/TDEI.2013.6508773>
 37. Ibrahim K, Sharkawy RM, Temraz HK, Salama MMA (2016) Selection criteria for oil transformer measurements to calculate the Health Index. *IEEE Trans Dielectr Electr Insul* 23:3397–3404. <https://doi.org/10.1109/TDEI.2016.006058>
 38. Mharakurwa ET, Nyakoe GN, Akumu AO (2019) Power transformer fault severity estimation based on dissolved gas analysis and energy of fault formation technique. *J Electr Comput Eng*. <https://doi.org/10.1155/2019/9674054>
 39. Abu-Siada A, Islam S (2012) A new approach to identify power transformer criticality and asset management decision based on dissolved gas-in-oil analysis. *IEEE Trans Dielectr Electr Insul* 19:1007–1012. <https://doi.org/10.1109/TDEI.2012.6215106>
 40. Ranga C, Chandel AK (2017) Expert system for health index assessment of power transformers. *Int J Electr Eng Inform* 9:850–865. <https://doi.org/10.15676/ijeii.2017.9.4.16>
 41. Prasojo RA, Gumilang H et al (2020) A fuzzy logic model for power transformer faults' severity determination based on gas level, gas rate, and dissolved gas analysis interpretation. *Energies* 13:1009
 42. Ranga C, Chandel AK, Chandel R (2017) Expert system for condition monitoring of power transformer using fuzzy logic. *J Renew Sustain Energy*. <https://doi.org/10.1063/1.4995648>
 43. Kadim EJ, Azis N, Jasni J et al (2018) Transformers health index assessment based on neural-fuzzy network. *Energies*. <https://doi.org/10.3390/en11040710>
 44. Ranga C, Chandel AK, Chandel R (2017) Condition assessment of power transformers based on multi-attributes using fuzzy logic. *IET Sci Meas Technol* 11:983–990. <https://doi.org/10.1049/iet-smt.2016.0497>
 45. Bacha K, Souahlia S, Gossa M (2012) Power transformer fault diagnosis based on dissolved gas analysis by support vector machine. *Electr Power Syst Res* 83:73–79. <https://doi.org/10.1016/j.epr.2011.09.012>
 46. Ghoneim S, Ward S (2012) Dissolved gas analysis as a diagnostic tools for early detection of transformer faults. *Adv Electr* 1:152–156
 47. Mansour DEA (2015) Development of a new graphical technique for dissolved gas analysis in power transformers based on the five combustible gases. *IEEE Trans Dielectr Electr Insul* 22:2507–2512. <https://doi.org/10.1109/TDEI.2015.004999>
 48. Mehta AK, Sharma RN, Chauhan S, Saho S (2013) Transformer diagnostics under dissolved gas analysis using Support vector machine. In: *Proceedings 2013 international conference on power, energy and control (ICPEC)*, pp 181–186. <https://doi.org/10.1109/ICPEC.2013.6527647>
 49. Tang WH, Goulermas JY, Wu QH et al (2008) A probabilistic classifier for transformer dissolved gas analysis with a particle swarm optimizer. *IEEE Trans Power Deliv* 23:751–759. <https://doi.org/10.1109/TPWRD.2008.915812>
 50. Su Q, Lai LL, Austin P (2001) A fuzzy dissolved gas analysis method for the diagnosis of multiple incipient faults in a transformer. *IEE Conf Publ* 15:344–348. <https://doi.org/10.1049/cp:20000420>
 51. Forouhari M (2017) Remnant life estimation of power transformers based on chemical diagnostic parameters using adaptive neuro-fuzzy inference system
 52. Duval M (2002) A review of faults detectable by gas-in-oil analysis in transformers. *IEEE Electr Insul Mag* 18:8–17
 53. Duval M, DePablo A (2001) Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases. *IEEE Electr Insul Mag* 17:31–41. <https://doi.org/10.1109/57.917529>
 54. Gouda OE, El-Hoshy SH, Tamaly HHEL (2019) Condition assessment of power transformers based on dissolved gas analysis. *IET Gener Transm Distrib* 13:2299–2310. <https://doi.org/10.1049/iet-gtd.2018.6168>
 55. Gouda OE, El-Hoshy SH, El-Tamaly HH (2018) Proposed three ratios technique for the interpretation of mineral oil transformers based dissolved gas analysis. *IET Gener Transm Distrib* 12:2650–2661. <https://doi.org/10.1049/iet-gtd.2017.1927>
 56. Scatiggio F, Fraioli A, Iuliani V, Pompili M (2014) Health index: the TERNAs practical approach for transformers fleet management. In: *CIGRE Session 45—45th international conference on large high voltage electric systems*, pp 178–182
 57. Ma C, Xie R, Zhang L et al (2018) A novel condition assessment method based on dissolved gas in transformer oil. *IEEE Electr Insul Conf EIC* 2018:232–235. <https://doi.org/10.1109/EIC.2018.8481079>
 58. Ghoneim SSM (2018) Intelligent prediction of transformer faults and severities based on dissolved gas analysis integrated with thermodynamics theory. *IET Sci Meas Technol* 12:388–394. <https://doi.org/10.1049/iet-smt.2017.0450>
 59. Prasojo RA, Gumilang H et al (2020) A fuzzy logic model for power transformer faults' severity determination based on gas level, gas rate, and dissolved gas analysis interpretation. *Energies*. <https://doi.org/10.3390/en13041009>
 60. Taecharoen P, Kunagonnimrattana P, Chotigo S (2019) Development of Dissolved gas analysis analyzing program using visual studio program. In: *2019 IEEE PES GTD Gd International Conference Expo Asia, GTD Asia 2019*, pp 785–790. <https://doi.org/10.1109/GTDA.2019.8715892>
 61. Mansour DEA (2012) A new graphical technique for the interpretation of dissolved gas analysis in power transformers. In: *Annual report—conference on electrical insulation and dielectric phenomena, CEIDP*, pp 195–198. <https://doi.org/10.1109/CEIDP.2012.6378754>
 62. Li E, Wang L, Song B, Jian S (2018) Improved fuzzy C-means clustering for transformer fault diagnosis using dissolved gas analysis data. *Energies*. <https://doi.org/10.3390/en11092344>

63. Abu-Siada A (2019) Improved consistent interpretation approach of fault type within power transformers using dissolved gas analysis and gene expression programming. *Energies*. <https://doi.org/10.3390/en12040730>
64. Abu-Siada A, Arshad M, Islam S (2010) Fuzzy logic approach to identify transformer criticality using dissolved gas analysis. *IEEE PES Gen Meet PES 2010*:1–5. <https://doi.org/10.1109/PES.2010.5589789>
65. Ahmed MR, Geliel MA, Khalil A (2013) Power transformer fault diagnosis using fuzzy logic technique based on dissolved gas analysis. In: 21st Mediterranean conference on control and automation, pp 584–589. <https://doi.org/10.1109/MED.2013.6608781>
66. Tamma WR, Prasajo RA (2021) High voltage power transformer condition assessment considering the health index value and its decreasing rate. *High Volt*. <https://doi.org/10.1049/hve2.12074>
67. Taha IBM, Mansour DEA, Ghoneim SSM, Elkalashy NI (2017) Conditional probability-based interpretation of dissolved gas analysis for transformer incipient faults. *IET Gener Transm Distrib* 11:943–951. <https://doi.org/10.1049/iet-gtd.2016.0886>
68. Farooque MU, Wani SA, Khan SA (2016) Artificial neural network (ANN) based implementation of Duval pentagon. In: 2015 International conference on condition assessment techniques in electrical systems CATCON 2015—Proceedings, pp 46–50. <https://doi.org/10.1109/CATCON.2015.7449506>
69. Hongzhong M, Zheng L, Ju P, et al (2008) Diagnosis of power transformer faults on fuzzy three-ratio method, pp 1–456. <https://doi.org/10.1109/ipecc.2005.206897>
70. Poiss G (2016) Development of DGA indicator for estimating risk level of power transformers. In: Proc—2016 17th international scientific conference on electric power engineering, EPE 2016, pp 4–7. <https://doi.org/10.1109/EPE.2016.7521813>
71. Mendes Barbosa T, Goncalves Ferreira J, Antonio Ferreira Finocchio M, Endo W (2017) Development of an application based on the Duval triangle method. *IEEE Lat Am Trans* 15:1439–1446. <https://doi.org/10.1109/TLA.2017.7994790>
72. Benmahamed Y, Tegar M, Boubakeur A (2017) Application of SVM and KNN to Duval Pentagon 1 for transformer oil diagnosis. *IEEE Trans Dielectr Electr Insul* 24:3443–3451. <https://doi.org/10.1109/TDEI.2017.006841>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.