INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Hybrid Method for Power Transformers Faults Diagnosis Based on Ensemble Bagged Tree Classification and Training Subsets Using Rogers and Gouda Ratios

Arnaud Nanfak[1]* **Charles Hubert Kom[1]** **Samuel Eke[1]**

[1]*Laboratory of Energy, Materials, Modelling and Methods, National Higher Polytechnic School of Douala, University of Douala, PO Box 2701 Douala, Cameroon*
* Corresponding author's Email: nanfak.arnaud@yahoo.fr

Abstract: As a major piece of equipment in the electrical power transmission and distribution, the rapid and accurate assessment of emerging or existing internal faults in power transformers is a key factor in the safe and stable operation of the power grid. This paper proposes a hybrid approach to fault diagnosis for these transformers. This approach is based on ensemble bagged tree classification and training subsets obtain by a conventional pre-processing method. Two pre-processing approaches are performed, the first based on the maximum concentrations of the dissolved gas samples and the second based on the minimum concentrations of the dissolved gas samples. For each training subset, an ensemble classifier is constructed with as inputs the Rogers ratios, Gouda ratios, dissolved gas concentrations and their associations. The proposed hybrid methods are established with 475 samples of training dataset, tested on 117 samples dissolved gas analysis (DGA) data and validated on International Electrotechnical Commission (IEC) TC10 database. The performances of the proposed diagnostic models are evaluated and a comparison is done compared with the following diagnostic methods: IEC ratios method (IRM), Rogers ratios method (RRM), three ratios technique (TRT), Gouda's triangle (GT), and self-organizing map (SOM) clusters. The results found by computer simulations carried out by the matrice laboratory (MATLAB) software show that, of the two pre-processing approaches used, the one based on the minimum sample concentration gives better results than the one based on the maximum concentration. In terms of fault type, the best diagnostic model using the minimum concentration-based pre-processing approach has a diagnostic accuracy of 94.02%, compared to 92.31% for the best diagnostic model using the maximum concentration-based pre-processing approach. This is lower than the 97.25% for SOM clusters and 96.58% for GT but higher than the 59.83% for RRM, 81.19% for IRM and 93.16% for TRT. In terms of fault severity, the best diagnostic model using the minimum concentration-based pre-processing approach has a diagnostic accuracy of 81.20%, compared to 74.36% for the best diagnostic model using the minimum concentration-based pre-processing approach. This result is lower than the 83.76% for TRT and 89.74% for GT but higher than the 49.57% for RRM, 66.67% for IRM and 78.90% for SOM clusters.

Keywords: Hybrid method, Power transformers faults diagnosis, Ensemble bagged trees algorithm, Training subsets, Rogers ratios, Gouda ratios.

1. Introduction

The power transformer is one of the most expensive and important pieces of equipment in the electrical system. As such, it plays a major role in the safe and stable operation of the power grid [1]. The failure of a power transformer during operation can result in a serious breakdown of the power grid, leading to interruptions in the transmission and distribution of electricity, environmental damage, risks of explosion and fire, and costly financial losses due to repairs or replacements [2]. Therefore, transformer condition analysis for early fault detection is extremely important in the process of operating and maintaining power networks [2, 3]. The dissolved gas analysis (DGA), is one of the most widely used techniques for the early detection of faults in the active parts of transformers [4]. It consists of examining the gases dissolved in the transformer oil to diagnose the condition of the transformer. Once the different gases have been identified and quantified, the result still needs to be interpreted to assess the condition of the transformer. Several methods are proposed in the literature to predict the occurrence of faults and to determine their types by interpreting the concentrations of the gases detected. Existing conventional methods for diagnosing faults in power transformers mainly include rule-based methods documented in Institute of Electrical and Electronics Engineers (IEEE) C57.104-1991, International Electrotechnical Commission (IEC) 60599-1999, and International Council on Large Electric Systems (CIGRE) TF 15.01.01. These methods are based on the analysis of the main gases, gas concentration ratios or certain proportions of gases such as Rogers ratios method (RRM) [5], Doernenburg ratios method [6], Key Gas method [7], or conventional IEC ratios method [8].

However, conventional DGA methods of interpretation have certain drawbacks in terms of precision and uncertainty. Limited in terms of data for learning because they require interpretation by human experts, some measurements may be unidentifiable [9]. In order to overcome the difficulties posed by traditional methods in interpreting test results, a major effort has been made to develop intelligent diagnosis in this area. For this purpose, several methods have used artificial intelligence (AI) including expert system (EPS) [10], artificial neural network (ANN) [11], fuzzy logic theory [12], rough sets theory (RST) [13], grey system theory (GST) [14], swarm intelligence (SI) algorithms [15], data mining technology [16], Self-organizing map (SOM) [17], machine learning (ML) [18] and optimized machine learning (OML) [19], to the diagnosis of transformer faults based on DGA data.

With the increased ability of computers to process data, machine learning (ML) algorithms are increasingly being developed for the state of power transformers. In [20], SVM is introduced for the first time for fault diagnosis of power transformer. The diagnosis model includes three-layer SVM classifier to identify the fault types. Zang et *al*. [21] proposed a new multiple SVMs to predict the fault of power transformer. They are demonstrated that, although a simple SVM has a superior generalisation capability, the new multiple SVM method has the best performance in learning ability and the generalisation ability. Li et *al*. [22] proposed an optimized model combining SVM with a genetic algorithm (SVMG) to diagnose power transformer faults. The experimental results indicated that the SVMG method can achieve higher diagnostic accuracy than IEC three ratios, normal SVM classifier, and artificial neural network. Senoussaoui et *al*. [23] proposed a synthesis study that tests the performance of four machine learning algorithms for DGA interpretation, namely, the Bayes network, the Multi-layer perceptron, the K-nearest neighbours and the J48 decision tree. In this study, the basic algorithms were tested and compared with data pre-processing to the improved version of these algorithms using an ensemble approach such as boosting and bagging. However, intelligent methods have certain drawbacks, including the need for a large training dataset, dependence on the parameters of the algorithm used and dependence on the data used [24]. These drawbacks have a negative impact on the learning process. In order to minimize their impact on the learning process, optimization algorithms are used to determine the optimal parameters of the AI algorithm, as well as data preprocessing methods [25, 26]. Although these solutions increase the diagnostic accuracy of the proposed methods, they greatly complicate them.

In this paper, another way to optimize the learning process of intelligent methods proposed. This approach based not on the parameters of the algorithm used or on the data, but on the shape of the training dataset. The idea is to divide the main training dataset into five training subsets defined according to conventional pre-processing. The conventional pre-processing approach used is to group the samples into subsets defined according to the maximum or minimum concentrations of the samples. The training subsets thus formed constitute the new datasets that will be used to build the diagnostic model. The subdivision of the main training dataset inspired by previous work in the field [27] on power transformer fault diagnosis using a conventional approach based on the creation of training subsets. The proposed diagnostic models are established by 475 samples training dataset, tested on 117 samples DGA data. The classification performance of the proposed method is validated on IEC TC10 database and compared with the following diagnostic methods: IEC ratios method (IRM), Rogers ratios method (RRM), three ratios technique (TRT), Gouda’s triangle (GT), and self-organizing map (SOM) clusters.

After the Introduction section, the remaining part of this paper is organized as follows: A brief description of the types of faults detectable by DGA, the relationships between the gases produced and the corresponding faults, and a presentation of RRM and TRT is given in Section 2. Section 3 is devoted to the description ensemble bagged tree algorithm. The principle and the flowchart of hybrid diagnosis approach are presented at Section 4. The performances of the classifiers proposed and its comparison with conventional and intelligent methods are presented in Section 5. Finally a conclusion is given in Section 6.

2. Fault types and DGA

2.1 Transformer fault types

The classification of faults in Publication IEC 60599 is based on the main types of faults that can be reliably identified by visual inspection of the equipment after the fault has occurred in service. IEC 60599 and IEEE C57.104 standards classify transformer faults detectable by gas analysis into two categories: electrical and thermal faults. Electrical faults refer to the deterioration of insulation caused by high electrical stress. Thermal faults refer to the deterioration of the insulation system because of abnormal temperature rises and result from overheating of conductors, short circuits, overheating of windings due to Foucault's currents, loose connections and insufficient cooling [28]. Based on IEC 60599, these major fault types can be further classified into 6 types of transformer faults, summarized in Table 1.

2.2 Relationship between faults and dissolved gas produced

The two principal causes of gas formation within an operating transformer are electrical disturbances and thermal decomposition. Each type of fault degrades the oil or paper differently, producing the relative amounts of dissolved gas that characterize the fault. Gas production is favoured by the temperature level and/or the energy produced by the fault. Depending on the type of fault, different types of decomposition processes may occur. When electrical and thermal faults occur in the transformer oil, it degrades, generating combustible gases, such as hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4) and acetylene (C_2H_4). When decomposition occurs in cellulosic insulation, the gases generated are carbon monoxide (CO) and carbon dioxide (CO_2), and these gases indicate a thermal fault. Other gases such as oxygen (O_2) and nitrogen (N_2) are also produce [4]. The nature of the gases formed and their relative proportions provide information on the nature of the stress, its intensity and the type of materials affected [29]. Table 2 summarizes the main gases produced according to the type of transformer faults.

Table 1. Fault classification according to IEC 60599 and IEEE C57.104 standard

acronyms	Faults
PD	Partial discharge
D_1	Low energy discharge
D_2	High energy discharge
T_1	Low temperature thermal fault T < 300° C
T_2	Medium temperature thermal fault 300° C < T < 700° C
T_3	High temperature thermal fault T > 700° C

Table 2. Gas generated according to the type of transformer fault [4], [30]

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

2.3 Rogers ratios method

The Rogers ratios method uses five key gases dissolved in oil namely H_2, CH_4, C_2H_6, C_2H_4 and C_2H_2 to compute three ratios and to develop a code that is supposed to give an indication of what causes the gases to evolve. This method is used to analyse six conditions of power transformer. In Table 3, ratio range and corresponding code are listed. The

Table 3. Rogers codes

Ratio	Ratio range	Code
$R_1 = C_2H_2/C_2H_4$	$R_1 < 0.1$	0
	$0.1 \leq R_1 \leq 3$	1
	$R_1 > 3$	2
$R_2 = CH_4/H_2$	$R_2 < 0.1$	1
	$0.1 \leq R_2 \leq 1$	0
	$R_2 > 1$	2
$R_3 = C_2H_4/C_2H_6$	$R_3 < 1$	0
	$1 \leq R_3 \leq 3$	1
	$R_3 > 3$	2

Table 4. Fault diagnosis by RRM

Fault type	R_1	R_2	R_3
Normal	0	0	0
Low energy density arcing-PD	0	1	0
Arcing-High energy discharge	1	0	2
Low temperature thermal	0	0	1
Thermal < 700° C	0	2	1
Thermal > 700° C	0	2	2

Received: February 23, 2022. Revised: April 26, 2022.

Table 5. Gouda codes

Ratio	Ratio range	Code
$R_1 = \frac{C_2H_6 + C_2H_4}{H_2 + C_2H_2}$	$R_1 < 0.05$	0
	$0.05 \le R_1 \le 0.9$	1
	$R_1 > 0.9$	2
$R_2 = \frac{C_2H_2 + CH_4}{C_2H_4}$	$R_2 < 1$	0
	$1 \le R_2 \le 3.5$	1
	$R_2 > 3.5$	2
$R_3 = \frac{C_2H_2}{C_2H_4}$	$R_3 < 0.05$	0
	$0.05 \le R_3 \le 0.5$	1
	$R_3 > 0.5$	2

Table 6. Fault diagnosis by TRT

Fault type	Severity of fault	R_1	R_2	R_3
High temperature thermal $T > 700°$ C	T_3	1 or 2	0	0 or 1
Medium temperature thermal $300°$ C $< T < 700°$ C	T_2	1 or 2	1	0 or 1
Low temperature thermal $150°$ C $< T < 300°$ C	T_1	1 or 2	2	0 or 1
Low temperature thermal $T < 150°$ C	T_0	1	/	0
Low partial discharge	PD_1	0	1 or 2	0 or 1
High partial discharge	PD_2	0	1 or 2	2
High arcing discharge	D_2	0 or 1	0 or 1	2
Low arcing discharge	D_1	1 or 2	2	2
Mix of electrical and thermal fault	DT	2	0 or 1	2

corresponding diagnostics for the various code combinations are presented in Table 4 [5].

2.4 Three ratios technique (TRT)

The TRT diagnosis technique proposed by Gouda et al. [31] suggests three new gas-ratio combinations that are able to classify the fault type and its severity clearly. The codes for the three-ratio used are given in Table 5.

In this method, the R_1 ratio is used to classify thermal, arcing and partial discharge faults. The R_3 ratio, used in existing diagnostic techniques by Doernenburg, Rogers and IEC, is able to separate thermal and electrical faults, so it is used to confirm the type of R_1 ratio fault. The R_2 ratio is used to assess the degree of severity of thermal, electrical and partial discharge faults. It is used to distinguish between low (PD_1) and high (PD_2) partial discharge faults, low (D_1) and high (D_2) energy discharge faults and also very low (T_0), low (T_1), medium (T_2) and high (T_3) temperature thermal energy faults [31]. The corresponding diagnostics for the various code combinations inspired of flowchart describe in [31] are presented in Table 6.

This technique shall be applied when at least one of the concentrations of dissolved gases exceeds the normal limits as shown in Table 7.

Table 7. Limit concentrations of dissolved gases for the application of TRT [31]

Gas	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	CO	CO_2
Limit (ppm)	100	120	65	50	1	350	2500

3. Ensemble bagged tree algorithm

Machine learning is a field in computer science where existing data are used to predict, or respond to, future data. In others words, it is the practice of programming computers to learn from data called training set or training dataset. The goal of ML is to find this optimal mapping to enable more accurate predictions and judgments of outputs [32]. An obvious approach to make decisions more reliable is to combine the results of several basic learners as members of a set and combine their predictions into a single result [33]. The final decision is reached based on these diverse opinions. These algorithms, called ensemble learning (EL) provides better performance compared to using a single classifier [34]. Strictly speaking, EL is not considered a type of ML algorithm but more of an optimization method or strategy [32]. In an ensemble classifier, each base classifier can be any kind of supervised classification

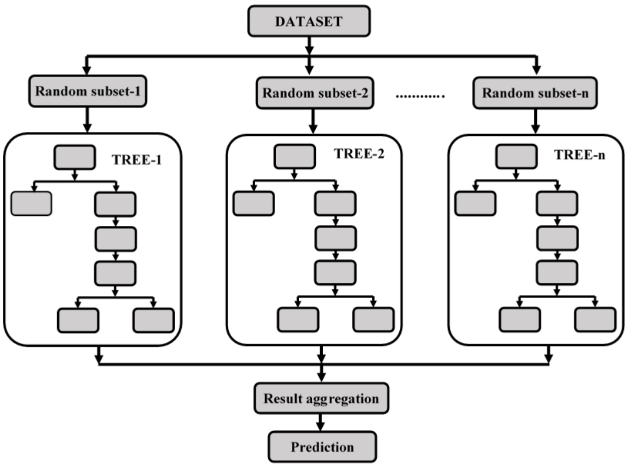

Figure. 1 Flow chart of ensemble bagged trees training [34]

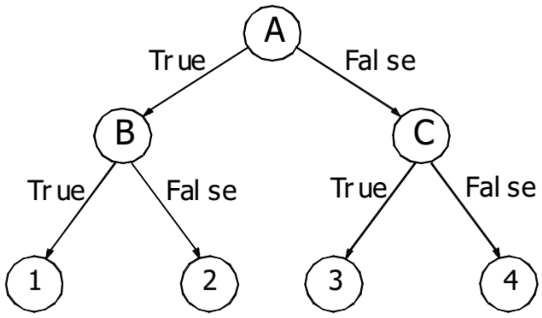

Figure. 2 Structure of a simple decision tree [37]

such as decision trees, neural networks, or SVMs. The most popular techniques for constructing ensembles are boosting, bagging, and stacking. Among these, Bootstrap aggregating, also known as bagging is the simplest, efficient and highly accurate ensemble-based algorithms [34, 35]. Each basic classifier in the ensemble is trained on a subset of the initial training set. The training subsets are sampled by bootstrapping, i.e., by randomly selecting a subset of the given data set with replacement. Classification of a new sample is based on the simple majority voting scheme, i.e., each classifier gives a classification for the sample, and the class that has the maximum number of votes cast by the base classifiers is the final classification [36]. In ensemble bagged trees, the basic classifiers are the decision trees. The Fig. 1 shows the principle of an ensemble bagging trees algorithm.

The decision tree is a kind of tree structure similar to a flowchart, where each inner node expresses test or selection for an attribute, and every branch represents a tested output, but each leaf node all represents class or class distribution [37]. The top node in tree is called root node. In the structure of a simple decision tree shown in Fig. 2, A, B and C represent test attributes, 1, 2, 3, 4 respectively express leaf nodes.

4. Proposed scheme for transformers fault diagnosis

In this paper, a hybrid approach that combines a conventional pre-processing and a machine learning algorithm to build models for power transformer faults diagnosis is proposed. Pre-processing allows the creation of training subsets consisting of samples with a common characteristic. Two pre-processing approaches are performed, the first based on the maximum concentrations of the dissolved gas samples and the second based on the minimum concentrations of the dissolved gas samples. The machine learning used is the ensemble bagged tree algorithm with Rogers and Gouda ratios as input features. As shown in Fig. 3, the main training dataset is divided into five subsets defined according to the chosen conventional pre-processing approach. The training subsets thus formed constitute the new training datasets to which the different algorithms of the ensemble bagged tree learning will be applied. The final diagnostic model includes five ensemble bagged tree classifiers (EBTC) which are used to identify the faults in the five training subsets. The $EBTC_1$ is trained to separate the faults of training subset having hydrogen as maximum or minimum concentration. The $EBTC_2$ is trained to separate the faults of training subset having methane as maximum or minimum concentration. The $EBTC_3$ is trained to separate the faults of training subset having ethane as maximum or minimum concentration. The $EBTC_4$ is trained to separate the faults of training subset having ethylene as maximum or minimum concentration and The $EBTC_5$ is trained to separate the faults of training subset having acetylene as maximum or minimum concentration.

Fig. 4 shows the flow chart of the proposed hybrid diagnostic model. The different steps in the implementation of the method are summarized as follows:

Step 1: Input dissolved gas sample concentrations
Step 2: Compute the Rogers and Gouda gas ratios
Step 3: Determination of the sample’s subset
Step 4: Fault diagnosis using the corresponding classifier

5. Results and discussion

5.1 Data collection

The present study was carried out using 592-labelled samples, collected from several sources and covering the six faults classes as presented in Table 8 below. 144 data samples from [38], 339 data samples collected from [39], 64 data samples from [9], 20 data

Received: February 23, 2022. Revised: April 26, 2022.

Table 8. Distribution of collected data according to references

	Fault types						
Ref.	**PD**	**D_1**	**D_2**	**T_1**	**T_2**	**T_3**	**Total**
[38]	16	35	15	29	19	30	144
[39]	32	51	74	85	41	56	339
[9]	0	32	32	0	0	0	64
[40]	7	2	2	0	5	4	20
[20]	0	7	18	0	0	0	25
Total	55	127	141	114	65	90	592

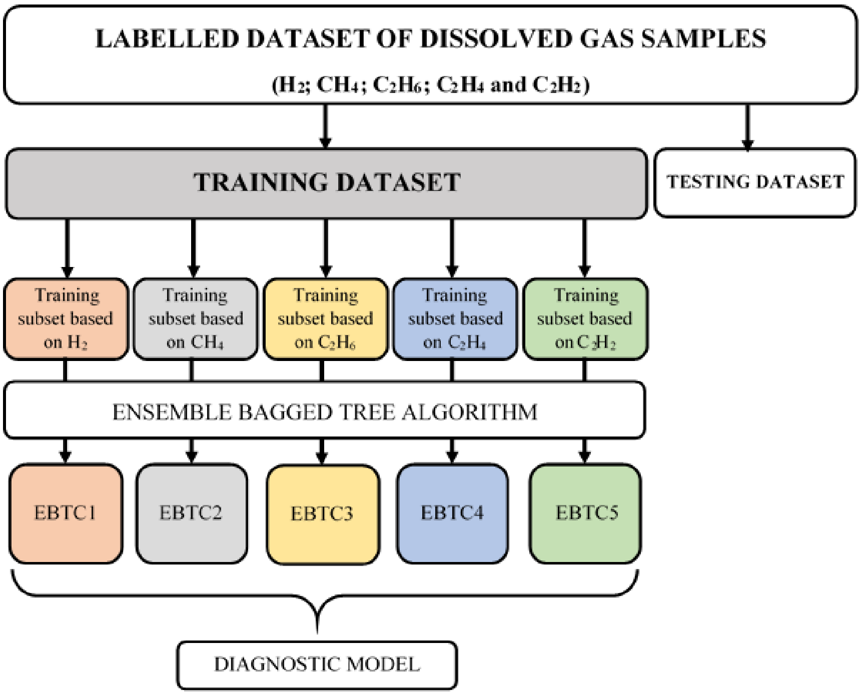

Figure. 3 Schematic view of the hybrid diagnostic method

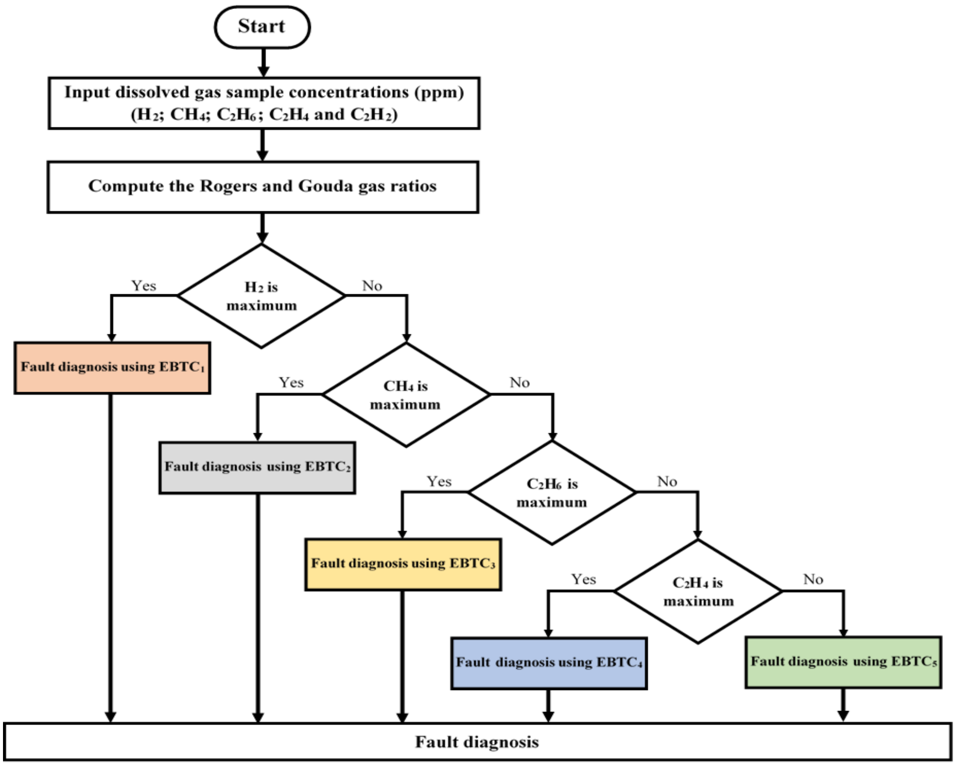

Figure. 4 Flow chart of proposed diagnostic model of power transformer according to maximum concentrations approach

Table 9. Composition of training and testing dataset

	Fault types						
	PD	**D_1**	**D_2**	**T_1**	**T_2**	**T_3**	**Total**
Training dataset	45	100	114	92	52	72	475
Testing dataset	11	25	28	22	13	18	117
Total	55	127	141	114	65	90	592

Table 10. Composition of training subsets according to fault severity

Major gas	**Fault type**						
	PD	**D_1**	**D_2**	**T_1**	**T_2**	**T_3**	**Total**
H_2	40	82	43	19	3	2	189
CH_4	1	3	3	22	20	2	51
C_2H_6	4	0	11	51	0	1	67
C_2H_4	0	3	6	3	29	68	109
C_2H_2	0	14	51	0	0	0	65

Table 11. Composition of training subsets according to fault severity

Minor gas	**Fault type**						
	PD	**D_1**	**D_2**	**T_1**	**T_2**	**T_3**	**Total**
H_2	0	1	6	3	3	5	18
CH_4	0	2	31	0	0	2	35
C_2H_6	8	55	57	1	0	6	127
C_2H_4	14	24	18	3	0	0	59
C_2H_2	42	14	5	86	51	61	259

samples from table 2 of [40] and 25 data samples from Table 1 and Table 2 of [20].

In order to conduct the diagnostic models proposed, the DGA data is divided into training and testing data for the implementation of flow chart and verification as shows in Table 9. Table 10 and Table 11 show the composition of training subsets according to fault severity. Table 10 for the conventional pre-treatment approach based on maximum dissolved gas sample concentrations and Table 11 for that based on minimum dissolved gas sample concentrations.

5.2 Analysis and comment

The different diagnostic models, summarised in Table 12 and 13, were made based on input features. All proposed diagnostic models and conventional methods were implemented using MATLAB software and the algorithms were programmed in .m code. Table 14 gives an overview of the faults diagnostic accuracy obtained with training and testing datasets.

Table 12. Different diagnostic models according to maximum concentrations approach

Input features			**Diagnostic model**
Key gases	**Rogers ratios**	**Gouda ratios**	
✓	✓	✓	DM_1
✓	✓		DM_2
✓		✓	DM_3
	✓	✓	DM_4
	✓		DM_5
		✓	DM_6
✓			DM_7

Table 13. Different diagnostic models according to minimum concentrations approach

Input features			**Diagnostic model**
Key gases	**Rogers ratios**	**Gouda ratios**	
✓	✓	✓	DM_8
✓	✓		DM_9
✓		✓	DM_{10}
	✓	✓	DM_{11}
	✓		DM_{12}
		✓	DM_{13}
✓			DM_{14}

5.3 Validation and comparison with other conventional methods using IEC TC10 database

The IEC TC10 database contains 117 cases of faults in transformers in service, which identified by visual inspection [41]. These data are not part of the training and testing dataset. In order to validate the diagnostic models proposed, this DGA database is used and average diagnostic accuracies by equipment type are summarized in Table 15. In this Table, the fault type refers to the 3 main faults, i.e. partial discharges, thermal overheating and arcing while the severity discriminates the 3 main faults i.e. PD for partial discharge, D_1 and D_2 for arcing and T_1, T_2 and T_3 for thermal overheating. The results obtained are compared with those obtained with RRM [5], IRM [8], TRT [31], Gouda's triangle (GT) [42] and SOM clusters [17]. Table 16 shows the abbreviations used for equipment type of IEC TC10 database.

Table 17 and 18 summarize the comparison between proposed diagnostic models and other diagnostic methods obtained with 117 cases of IEC TC10 databases. In Table 17, the comparison is done in terms of fault type and in Table 18 in terms of severity.

Received: February 23, 2022. Revised: April 26, 2022.

Table 14. Fault diagnosis accuracies of different diagnostic model based on subsets

			Fault diagnosis accuracy (%)					
Pre-processing approach	**Ensemble tree models**		**H_2**	**CH_4**	**C_2H_6**	**C_2H_4**	**C_2H_2**	**Total**
Diagnostic models based on maximum concentrations pre-processing	DM_1	Training	98.94	100	100	100	100	99.58
		Testing	86.79	83.33	92.31	95.65	100	90.60
	DM_2	Training	97.88	97.87	100	100	100	98.95
		Testing	86.79	91.67	92.31	91.30	100	90.60
	DM_3	Training	98.94	100	100	99.04	100	99.37
		Testing	84.91	66.67	92.31	73.91	100	83.76
	DM_4	Training	98.94	100	100	99.04	100	99.37
		Testing	79.24	83.33	92.31	95.65	93.33	87.18
	DM_5	Training	98.94	97.87	100	100	100	99.37
		Testing	77.36	100	92.31	95.65	86.67	87.18
	DM_6	Training	98.94	100	100	100	100	99.58
		Testing	62.26	75.00	92.31	73.91	93.33	74.36
	DM_7	Training	98.94	97.87	100	100	100	99.37
		Testing	84.62	75.00	92.31	82.61	100	86.32
Diagnostic models based on maximum concentrations	DM_8	Training	100	100	100	100	98.84	99.36
		Testing	87.50	85.71	88.24	88.00	82.54	87.18
	DM_9	Training	100	100	100	100	99.23	99.58
		Testing	87.50	92.86	88.24	92.00	85.71	89.74
	DM_{10}	Training	100	100	98.43	100	99.23	99.16
		Testing	87.50	92.86	76.47	84.00	82.54	84.61
	DM_{11}	Training	100	100	98.43	100	99.23	99.16
		Testing	87.50	92.86	70.59	88.00	84.13	86.32
	DM_{12}	Training	100	100	99.21	98.31	98.84	99.16
		Testing	87.50	85.71	70.59	88.00	85.71	85.47
	DM_{13}	Training	100	97.50	100	100	98.84	99.16
		Testing	87.50	92.86	70.59	80.00	76.19	79.49
	DM_{14}	Training	100	100	99.21	100	99.23	99.16
		Testing	87.50	85.71	94.12	88.00	82.54	86.32

The diagnostic accuracies obtained with the different methods on the IEC TC10 database are presented according to the equipment and distributed according to severity and fault type. Considering the diagnostic accuracies obtained by equipment, it appears that for P-type and U-type transformers, the proposed diagnostic models allow the classification of faults in terms of fault type. Indeed, 100% diagnostic accuracy is achieved with the DM_4, DM_{10}, DM_{11}, DM_{12} and DM_{13} diagnostic models for P-type transformers and with the DM_6, DM_{11} and DM_{13} models for U-type transformers. For these types of equipment, less good performances are achieved in terms of fault severity. For the maximum concentration-based pre-processing approach, the best performances are achieved with the DM_5 diagnostic model, which obtained diagnostic accuracies of 77.78% for P-type transformers and 68.18% for U-type transformers. For the minimum concentration-based pre-processing approach, the best performances are achieved with the DM_{10} and DM_8 diagnostic models, which respectively obtained diagnostic accuracies of 80.56% for P-type transformers and 81.82% for U-type transformers. Out the 117 cases including all types of equipment, the best results with the maximum concentration-based pre-processing approach are achieved in terms of fault type by DM_5 and DM_6 with a diagnostic accuracy of 92.31% and in terms of fault severity by DM_4 and DM_5 with a diagnostic accuracy of 74.36%. For minimum concentration-based pre-processing

Received: February 23, 2022. Revised: April 26, 2022.

Table 15. Average diagnosis accuracy of diagnosis models validated with IEC TC10 database

		Equipment type								
Ensemble tree models		P	U	R	I	B	C	S	**Vide**	**Total**
DM_1	Severity	75.00	68.18	81.25	50.00	40.00	100.00	85.71	100.00	72.65
	Fault type	97.22	90.91	90.63	58.33	80.00	100.00	100.00	100.00	89.74
DM_2	Severity	69.44	68.18	71.88	58.33	40.00	100.00	85.71	100.00	69.23
	Fault type	94.44	90.91	87.50	66.67	80.00	100.00	100.00	100.00	88.89
DM_3	Severity	69.44	63.64	81.25	58.33	40.00	100.00	85.71	100.00	70.94
	Fault type	97.22	86.36	90.63	66.67	80.00	100.00	100.00	100.00	89.74
DM_4	Severity	75.00	68.18	81.25	66.67	40.00	100.00	85.71	100.00	74.36
	Fault type	100.00	90.91	90.63	66.67	80.00	100.00	100.00	100.00	91.45
DM_5	Severity	77.78	68.18	75.00	75.00	40.00	100.00	85.71	100.00	74.36
	Fault type	97.22	95.45	90.63	75.00	80.00	100.00	100.00	100.00	92.31
DM_6	Severity	69.44	68.18	81.25	66.67	40.00	50.00	85.71	100.00	71.79
	Fault type	97.22	100.00	90.63	66.67	80.00	100.00	100.00	100.00	92.31
DM_7	Severity	61.11	59.09	75.00	58.33	20.00	50.00	85.71	100.00	64.10
	Fault type	91.67	86.36	90.63	66.67	60.00	100.00	100.00	100.00	87.18
DM_8	Severity	75.00	81.82	87.50	58.33	20.00	100.00	85.71	100.00	76.92
	Fault type	97.22	95.45	93.75	58.33	60.00	100.00	100.00	100.00	89.74
DM_9	Severity	77.78	72.73	78.13	41.67	20.00	100.00	85.71	100.00	71.79
	Fault type	97.22	86.36	90.63	41.67	60.00	100.00	100.00	100.00	86.32
DM_{10}	Severity	80.56	63.64	78.13	58.33	20.00	50.00	85.71	100.00	71.79
	Fault type	100.00	86.36	90.63	66.67	80.00	100.00	100.00	100.00	90.60
DM_{11}	Severity	75.00	68.18	87.50	75.00	40.00	50.00	85.71	100.00	76.07
	Fault type	100.00	100.00	93.75	75.00	80.00	100.00	100.00	100.00	94.87
DM_{12}	Severity	77.78	77.27	90.63	75.00	40.00	100.00	100.00	100.00	81.20
	Fault type	100.00	95.45	93.75	75.00	80.00	100.00	100.00	100.00	94.02
DM_{13}	Severity	66.67	77.27	87.50	66.67	20.00	5000	85.71	100.00	73.50
	Fault type	100.00	100.00	93.75	66.67	60.00	100.00	100.00	100.00	93.16
DM_{14}	Severity	63.89	50.00	65.63	50.00	20.00	100.00	85.71	100.00	60.68
	Fault type	94.44	86.36	90.63	66.67	80.00	100.00	100.00	100.00	88.89
RRM	Severity	52.77	59.09	56.25	41.66	00.00	50.00	28.57	00.00	49.57
	Fault type	69.44	63.63	65.62	41.66	40.00	50.00	28.57	00.00	59.83
TRT	Severity	86.11	95.45	84.37	83.33	20.00	100	71.42	100	83.76
	Fault type	94.44	100	93.75	83.33	60.00	100	100	100	93.16
IRM	Severity	63.88	77.27	75.00	58.33	00.00	100	71.42	00.00	66.67
	Fault type	86.11	86.36	87.50	58.33	40.00	100	85.71	00.00	81.19
GT	Severity	86.11	95.45	93.75	91.67	60.00	100	85.71	100	89.74
	Fault type	94.44	100	96.88	91.67	100	100	100	100	96.58
SOM cl.	Severity	77.78	72.73	84.36	91.67	60.00	50.00	/	/	78.90
	Fault type	100	95.45	96.88	100	100	50.00	/	/	97.25

approach, the best results are achieved in terms of fault type by DM_{11} with a diagnostic accuracy of 94.87% and in terms of fault severity by DM_{12} with a diagnostic accuracy of 81.20%. Of the two pre-processing approaches used, the one based on the minimum sample concentration gives better results than the one based on the maximum sample concentration. Indeed, in terms of fault type, the

Table 16. Abbreviations used for equipment type

Abbreviations	**Equipment**
P	Power transformer without communication OLTC
U	Power transformer with communication OLTC
R	Reactor
I	Instrument transformer
B	Bushing
C	Cable

Table 17. Comparison between proposed method and conventional methods in terms of fault type

DGA models	Unresolved diagnostic	Wrong diagnostic	Error	Diagnostic accuracy
DM_1	00.00	10.26	10.26	89.74
DM_2	00.00	11.11	11.11	88.89
DM_3	00.00	10.26	10.26	89.74
DM_4	00.00	08.55	08.55	91.45
DM_5	00.00	07.69	07.69	92.31
DM_6	00.00	07.69	07.69	92.31
DM_7	00.00	12.82	12.82	87.18
DM_8	00.00	10.26	10.26	89.74
DM_9	00.00	13.68	13.68	86.32
DM_{10}	00.00	09.40	09.40	90.60
DM_{11}	00.00	05.13	05.13	94.87
DM_{12}	00.00	05.98	05.98	94.02
DM_{13}	00.00	06.84	06.84	93.16
DM_{14}	00.00	11.11	11.11	88.89
RRM	21.89	18.28	40.17	59.83
TRT	00.85	05.99	06.84	93.16
IRM	14.53	04.28	18.81	81.19
GT	00.00	03.42	03.42	96.58
SOM cl.	00.00	02.75	02.75	97.25

Table 18. Comparison between proposed method and conventional methods in terms of severity

DGA models	Unresolved diagnostic (%)	Wrong diagnostic (%)	Error (%)	Diagnostic accuracy (%)
DM_1	00.00	27.35	27.35	72.65
DM_2	00.00	30.77	30.77	69.23
DM_3	00.00	29.06	29.06	70.94
DM_4	00.00	25.64	25.64	74.36
DM_5	00.00	25.64	25.64	74.36
DM_6	00.00	28.21	28.21	71.79
DM_7	00.00	35.90	35.90	64.10
DM_8	00.00	23.08	23.08	76.92
DM_9	00.00	28.21	28.21	71.79
DM_{10}	00.00	28.21	28.21	71.79
DM_{11}	00.00	23.93	23.93	76.07
DM_{12}	00.00	18.80	18.80	81.20
DM_{13}	00.00	26.50	26.50	73.50
DM_{14}	00.00	39.32	39.32	60.68
RRM	21.89	28.54	50.43	49.57
TRT	00.85	15.38	16.24	83.76
IRM	14.53	18.80	33.33	66.67
GT	00.00	10.26	10.26	89.74
SOM cl.	00.00	21.10	21.10	78.90

DM_{12} diagnostic model has a diagnostic accuracy of 94.02%, compared to 92.31% for the DM_5 diagnostic model. This is lower than the 97.25% for SOM clusters and 96.58% for GT but higher than the 59.83% for RRM, 81.19% for IRM and 93.16% for TRT. In terms of fault severity, the DM_{12} diagnostic model has a diagnostic accuracy of 81.20%, compared to 74.36% for the DM_5 diagnostic model. This result is lower than the 83.76% for TRT and 89.74% for GT but higher than the 49.57% for RRM, 66.67% for IRM and 78.90% for SOM clusters.

6. Conclusion

In this paper, a hybrid approach for power transformers fault diagnosis is proposed. This approach combines a conventional pre-processing and a machine learning algorithm. Two pre-processing approaches are performed, the first based on the maximum concentrations of the dissolved gas samples and the second based on the minimum concentrations of the dissolved gas samples. This pre-processing allows us to obtain new training subsets with common characteristics. The Ensemble bagged tree algorithm is used as machine learning algorithm with Rogers and Gouda ratios as input features for faults classification. The dataset used in this paper contains 709 labelled samples covering six fault types. The first group of 592 samples for implementation and evaluation of diagnostic models proposed. The performances of proposed diagnostic models are validated using the second group of data consisting of the 117 samples from the IEC TC10 database. The use of the subsets proposed in this work reduces the size of the calculations required for fault pattern recognition during training and the number of fault patterns to be identified in each subset. A hybrid diagnostic approach combining an intelligent pre-processing allowing the creation of subsets followed by a conventional processing of the different subsets is a perspective of this work.

Received: February 23, 2022. Revised: April 26, 2022.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, software, formal analysis, investigation, writing—original draft preparation, writing—review and editing have been done by 1st author (Arnaud NANFAK). Methodology, validation, supervision have been done by 2nd and 3rd author (Charles Hubert KOM & Samuel EKE).

References

[1] Y. Liu, B. Song, L. Wang, J. Gao, and R. Xu, "Power Transformer Fault Diagnosis Based on Dissolved Gas Analysis by Correlation Coefficient-DBSCAN", *Applied Sciences*, Vol. 10, No. 3, pp. 1-18, 2020.

[2] F. E. Bezerra, F. A. Z. Garcia, S. I. Nabeta, G. F. M. D. Souza, I. E. Chabu, J. C. Santos, S. N. Junior, and F. H. Pereira, "Wavelet-Like Transform to Optimize the Order of an Autoregressive Neural Network Model to Predict the Dissolved Gas Concentration in Power Transformer Oil from Sensor Data", *Sensors*, Vol. 20, No. 9, pp. 597-610, 2020.

[3] J. I. Aizpurua, V. M. Catterson, B. G. Stewart, S. D. J. McArthur, B. Lambert, B. Ampofo, G. Pereira, and J. G. Cross, "Selecting appropriate machine learning classifiers for DGA diagnosis", In: *Proc. of 2017 IEEE Conference on Electrical Insulation and Dielectric Phenomenon (CEIDP), Fort Worth*, pp. 153-156, 2017.

[4] L. Cheng and T. Yu, "Dissolved Gas Analysis Principle-Based Intelligent Approaches to Fault Diagnosis and Decision Making for Large Oil-Immersed Power Transformers: A Survey", *Energies*, Vol. 11, No. 4, pp. 913-978, 2018.

[5] R. R. Rogers, "IEEE and IEC Codes to Interpret Incipient Faults in Transformers, Using Gas in Oil Analysis", *IEEE Transactions on Electrical Insulation*, Vol. EI-13, No. 5, pp. 349-354, 1978.

[6] "IEEE Guide for the Detection and Determination of Generated Gases in Oil-Immersed Transformers and Their Relation to the Serviceability of the Equipment", *ANSI/IEEE Std C57.104-1978*, pp. 1-27, 1978.

[7] "IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers", *IEEE Std C57.104-2019 (Revision of IEEE Std C57.104-2008)*, pp. 1-98, 2019.

[8] "Interpretation of the analysis of gases in transformers and other Oil filled electrical equipment in service", *IEC Publication 60599*, 1999.

[9] H. A. Illias and W. Z. Liang, "Identification of transformer fault based on dissolved gas analysis using hybrid support vector machine-modified evolutionary particle swarm optimisation", *PLOS ONE*, Vol. 13, No. 1, pp. 1-15, 2018.

[10] C. Ranga, A. K. Chandel, and R. Chandel, "Expert system for condition monitoring of power transformer using fuzzy logic", *Journal of Renewable and Sustainable Energy*, Vol. 9, No. 4, pp. 1-16, 2017.

[11] Q. Wang, S. Wang, R. Shi, and Y. Li, "A Power Transformer Fault Diagnosis Method Based on Random Vector Functional-Link Neural Network", *Mathematical Problems in Engineering*, Vol. 2021, pp. 1-18, 2021.

[12] R. N. Patel and H. S. Thakur, "Integrated fuzzy approach for incipient fault detection in power transformers", In: *Proc. of IEEE Industrial Electronics and Applications Conference (IEACon)*, pp. 211-218, 2016.

[13] S. M. Shaaban and H. A. Nabwey, "Transformer Fault Diagnosis Method Based on Rough Set and Generalized Distribution Table", *International Journal of Intelligent Engineering and Systems*, Vol. 5, No. 2, pp. 17-24, 2012.

[14] M. Xue, "State Evaluation for Transformer Based on Grey System Theory", *Applied Mechanics and Materials*, Vol. 488-489, pp. 983-987, 2014.

[15] W. Mo, T. Kari, H. Wang, L. Luan, and W. Gao, "Power Transformer Fault Diagnosis Using Support Vector Machine and Particle Swarm Optimization", In: *Proc. of 2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, pp. 511-515, 2017.

[16] C. H. Liu, T. L. Chen, L. T. Yao, and S. Y. Wang, "Using data mining to dissolved gas analysis for power transformer fault diagnosis", In: *Proc. of International Conference on Machine Learning and Cybernetics*, pp. 1952-1957, 2012.

[17] S. Misbahulmunir, V. K. Ramachandaramurthy, and Y. H. M. Thayoob, "Improved Self-Organizing Map Clustering of Power Transformer Dissolved Gas Analysis Using Inputs Pre-Processing", *IEEE Access*, Vol. 8, pp. 71798-71811, 2020.

[18] S. Eke, G. Clerc, T. A. Ngnui, and I. Fofana, "Transformer condition assessment using fuzzy C-means clustering techniques", *IEEE Electrical Insulation Magazine*, Vol. 35, No. 2, pp. 47-55, 2019.

[19] I. B. Taha and D. A. Mansour, "Novel power transformer fault diagnosis using optimized machine learning methods", *Intelligent Automation & Soft Computing*, Vol. 28, No. 3, pp. 739-752, 2021.

[20] G. Lv, H. Cheng, H. Zhai, and L. Dong, "Fault diagnosis of power transformer based on multi-layer SVM classifier", *Electric Power Systems Research*, Vol. 75, No. 1, pp. 9-15, 2005.

[21] M. H. Zhang, K. Li, and H. Tian, "Multiple SVMs Modelling Method for Fault Diagnosis of Power Transformers", *Electrotechnical Journal*, Vol. 88, No. 7, pp. 232-234, 2012.

[22] J. Li, Q. Zhang, K. Wang, J. Wang, T. Zhou, and Y. Zhang, "Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine", *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 23, No. 2, pp. 1198-1206, 2016.

[23] M. E. A. Senoussaoui, M. Brahami, and I. Fofana, "Combining and comparing various machine-learning algorithms to improve dissolved gas analysis interpretation", *IET Generation, Transmission & Distribution*, Vol. 12, No. 15, pp. 3673-3679, 2018.

[24] K. Liu, W. Shizhe, Z. Luo, Z. Gongze, X. Ma, Z. Cao, and H. Li, "An intelligent fault diagnosis method for transformer based on IPSO-gcForest", *Mathematical Problems in Engineering*, Vol. 2021, pp. 1-12, 2021.

[25] H. Ma, W. Zhang, R. Wu, and C. Yang, "A power transformers fault diagnosis model based on three DGA ratios and PSO optimization SVM", *IOP Conference Series: Materials Science and Engineering*, Vol. 339, pp. 1-6, 2018.

[26] J. Li, C. Hai, Z. Feng, and G. Li, "A Transformer Fault Diagnosis Method Based on Parameters Optimization of Hybrid Kernel Extreme Learning Machine", *IEEE Access*, Vol. 9, No., pp. 126891-126902, 2021.

[27] A. Nanfak, S. Eke, C. H. Kom, R. Mouangue, and I. Fofana, "Interpreting dissolved gases in transformer oil: A new method based on the analysis of labelled fault data", *IET Generation Transmission & Distribution*, Vol. 15, No. 21, pp. 3032-3047, 2021.

[28] V. A. Thiviyanathan, P. J. Ker, Y. S. Leong, F. Abdullah, A. Ismail, and M. Z. Jamaludin, "Power transformer insulation system: A review on the reactions, fault detection, challenges and future prospects", *Alexandria Engineering Journal*, Vol. 61, No. 10, pp. 7697-7713, 2022.

[29] S. Bustamante, M. Manana, A. Arroyo, P. Castro, A. Laso, and R. Martinez, "Dissolved Gas Analysis Equipment for Online Monitoring of Transformer Oil: A Review", *Sensors*, Vol. 19, No. 19, pp. 1-21, 2019.

[30] J. S. N'cho, I. Fofana, Y. Hadjadj, and A. Beroual, "Review of Physicochemical-Based Diagnostic Techniques for Assessing Insulation Condition in Aged Transformers", *Energies*, Vol. 9, No. 5, pp. 1-29, 2016.

[31] O. E. Gouda, S. H. E. Hoshy, and H. H. E. L. Tamaly, "Proposed three ratios technique for the interpretation of mineral oil transformers based dissolved gas analysis", *IET Generation, Transmission & Distribution*, Vol. 12, No. 11, pp. 2650-2661, 2018.

[32] L. Cheng and T. Yu, "A new generation of AI: A review and perspective on machine learning technologies applied to smart energy and electric power systems", *International Journal of Energy Research*, Vol. 43, No. 6, pp. 1928-1973, 2019.

[33] R. Mousavi and M. Eftekhari, "A new ensemble learning methodology based on hybridization of classifier ensemble selection approaches", *Applied Soft Computing*, Vol. 37, No. C, pp. 652-666, 2015.

[34] M. S. Saeed, M. W. Mustafa, U. U. Sheikh, T. A. Jumani, and N. H. Mirjat, "Ensemble Bagged Tree Based Classification for Reducing Non-Technical Losses in Multan Electric Power Company of Pakistan", *Electronics*, Vol. 8, No. 8, pp. 1-16, 2019.

[35] K. Mishra, A. Yadav, and M. Pazoki, "A Novel Fault Classification Scheme for Series Capacitor Compensated Transmission Line Based on Bagged Tree Ensemble Classifier", *IEEE Access*, Vol. 6, pp. 27373-27382, 2018.

[36] M. O. Elish and K. Elish, "An empirical comparison of resampling ensemble methods of deep learning neural networks for cross-project software defect prediction", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 3, pp. 201-209, 2021, doi: 10.22266/ijies2021.0630.18.

[37] F. Zhao and H. Su, "A decision tree approach for power transformer insulation fault diagnosis", In: *Proc. of 2008 7th World Congress on Intelligent Control and Automation*, pp. 6882-6886, 2008.

[38] E. Li, L. Wang, and B. Song, "Dissolved gas data in transformer oil Fault Diagnosis of Power Transformers with Membership Degree", *IEEE Access*, Vol. 7, pp. 28791-28798, 2019.

[39] "DGA.", Jun. 01, 2020. https://github.com/Saleh860/DGA/tree/master/datasets (accessed Jun. 01, 2020).

Received: February 23, 2022. Revised: April 26, 2022.

[40] B. Zeng, J. Guo, W. Zhu, Z. Xiao, F. Yuan, and S. Huang, "A Transformer Fault Diagnosis Model Based on Hybrid Grey Wolf Optimizer and LS-SVM", *Energies*, Vol. 12, No. 21, pp. 1-18, 2019.

[41] 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, 2001.

[42] O. E. Gouda, S. H. E. Hoshy, and H. H. E. Tamaly, "Condition assessment of power transformers based on dissolved gas analysis", *IET Generation, Transmission & Distribution*, Vol. 13, No. 12, pp. 2299-2310, 2019.