EMITTER International Journal of Engineering Technology
Vol. 10, No. 2, December 2022, pp. 233~245
DOI : 10.24003/emitter.v10i2.702

Classification Method in Fault Diagnosis of Oil-Immersed Power Transformers by Considering Dissolved Gas Analysis

Rosmaliati [1,2], Bernandus Anggo Seno Aji[3], Isa Hafidz[1,4], Ardyono Priyadi[1], Mauridhi Hery Purnomo[1,5]

[1]Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2]Department of Eletrical Engineering, Universitas Mataram, Mataram, Indonesia
[3]Department of Information Technology, Institut Teknologi Telkom, Surabaya, Indonesia
[4]Department of Eletrical Engineering, Institut Teknologi Telkom, Surabaya, Indonesia
[5]Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
Corresponding author: hery@ee.its.ac.id

Received August 20, 2022; Revised September 22, 2022; Accepted October 25, 2022

Abstract

Fault detection in the incipient stage is necessary to avoid hazardous operating conditions and reduce outage rates in transformers. Fault-detected dissolved gas analysis is widely used to detect incipient faults in oil-immersed transformers. This paper proposes fault diagnosis transformers using an artificial neural network based on classification techniques. Data on the condition of transformer oil is assessed for dissolved gas analysis to measure the dissolved gas concentration in the transformer oil. This type of disturbance can affect the gas concentration in the transformer oil. Fault diagnosis is implemented, and fault reference is provided. The result of the NN method is more accurate than the Tree and Random Forest method, with CA and AUC values 0.800 and 0.913. This classification approach is expected to help fault diagnostics in power transformers.

Keywords: Fault Diagnostic, Neural Network, Power Transformer

1. INTRODUCTION

One of the essential devices in distributing electrical power to consumers is a transformer. Some things that can interfere with this process can be caused by short circuits, harmonics, extreme load changes, and chemical conditions of the insulation of the device. Insulation on oil-paper type operates with high temperature and solid electromagnetic environment. Over time, the insulating media can decompose slowly. It can even dissolve more quickly in certain disturbances [1][2].

For oil-immersed transformers, oil and cellulose insulators are generally used. These disturbances can result in the appearance of gases and molecules that are difficult to dissolve in the transformer oil. In order to detect these conditions, the Gas chromatography technique is used [3]. The researchers also developed methods for detecting errors. One of the techniques for detecting these disturbances is Dissolved Gas Analysis (DGA). The procedure for this method is carried out by removing dissolved gas from the sample oil and then determining the gas component content. After knowing the type of fault, the next step is to identify the type of internal fault of the transformer. Several other approaches to detecting gases include the IEC 60599 method [4], Duval Triangles [5] and Duval Pentagon [6]. Other work related to gas analysis was also carried out by [7] and [8] to identify dissolved gas patterns. The analysis process uses computational techniques by considering the ratio of gas components. The method was developed and validated using large datasets on certain gas type thresholds.

2. RELATED WORKS

The decomposable gases in the oil are labelled and further classified by the researchers using intelligent approaches. In this method, numerical limitation of gas parameters is carried out to classify the features contained in the gas. However, in some values, an error occurred in the analysis process because the data ratio was too close, and there was a possibility that interference could occur continuously. Researchers have also employed several other approaches related to fault diagnosis by considering oil. These methods include fuzzy [9]. Work [10] proposes an SVM model based on an imperialist competitive algorithm. This approach is used to classify non-linear data to justify transformer oil diagnosis. However, there are further challenges in the selection process considering other methods. Research [11] discusses feature selection to diagnose faults in transformers. The feature selection method discussed is a combination of genetic algorithms and SVM. The part that can be explored further is the use of soluble gas data analysis to improve the accuracy of diagnosis. Research [12] describes artificial intelligence techniques to classify DGA with power transformers. The approaches used include fuzzy logic, neural network (NN), and support vector machine (SVM).

Some things that have not been considered are arc discharge and spark discharge conditions in DGA analysis. Failure statistics of transformers with dissolved gas analysis technique In [13] describes the results of testing the performance of the Bayes network, multilayer perceptron, k-nearest neighbor, and j48. The results were obtained to explain the efficiency and stability of the j48 and Bayes network methods. However, the proposed method has not considered the NN and random forest. Work on the use of fuzzy to classify the types of disturbances present in transformers was carried out in [14]. DGA analysis was used to consider gas density, which consisted of 13 concentrate samples. The classification of transformer fault types is carried out using the fuzzy method with DGA data. The other is by implementing rough set theory [15]. However, complex data processing has not been further elaborated, and the limited DGA data is a challenge.

Figure 1 describes the comparison between the operating level of the transformer under normal conditions, defects, and errors to the safety factor of its operation. The operational safety factor can be represented in two critical and safety conditions. Section 1 describes the normal operating conditions. Part 2 is a damaged condition due to damage to the equipment. It is in critical condition, and part 3 describes the transformer life. The faulty operation causes the transformer to enter a critical condition faster than the faulty operation. Users can classify whether the transformer is in the category of defect or faulty.

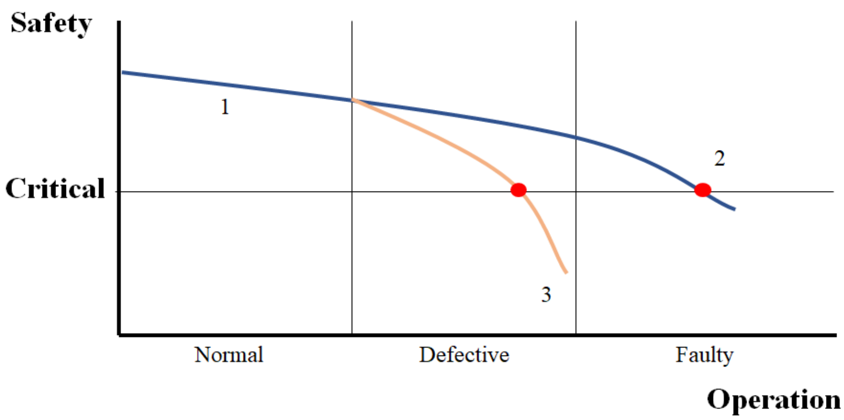

Figure 1. Transformer operation compared with level safety of transformer

Dissolved gas analysis (DGA) is used to assess the condition of power transformers. As has been widely applied in the world, this technique is used to measure the concentration of various types of gases dissolved in transformer oil. The oil in the transformer has an electrically insulating mineral, which functions as a heat transfer medium. When electrical and thermal disturbances occur, the molecules decompose and release gases [12, 16, 17]. Some gases are trapped in the oil, but some are soluble. In the DGA test, the oil must be sampled according to the IEC60567 regulations in the laboratory.

The analysis process for the separation of the solution is carried out with a gas chromatography device. Then the gas in the solution is measured and recorded. This process occurs several times under transformer conditions. Fault in transformer that affects its internal condition. In some cases, the fault is a combination of several types. The results of the DHA Research by experts to calculate the transformer index. The gases produced include CO2, CO, H2, CH4, C2H2, C2H4, and C2H6. The gas produced can trigger a fire, except for CO2. Different methods for diagnosing internal failures include Dornenburg, Rogers, key gas, Duvaltriangle, grading system, IEEEC57.104, and IEC standard 60599. The disaggregated data relates to a chromatographic analysis of the fault types.

Classification algorithms help build classification models. The investigation focuses not only on one approach but also on other machine learning algorithms. Classification algorithms help build a classification model. Investigation of the use of algorithms focuses on one approach and compares it with other machine learning. These techniques include using decision trees, random forests, and NNs. Decision trees have the advantage of solving a series of features in a one-step decision. Random forests have an approach to solving problems by combining classifiers. NN completes the classification process by using layers to get the best solution. The selection of the algorithm took into account its popularity for comparing different learning paradigms [13]. In this study, only supervised investigations were used from labelled data samples. The validation of the machine learning model was carried out by dividing the training data and testing datasets to test the model reliability.

3. ORIGINALITY

From a review of existing methods, it is known that the decision tree algorithm is not adequate to be applied to regression and predictive values that have a sustainable nature. Then, the random forest method has a challenge in terms of interpretability and determining the significance of each variable. This refers to the ensemble nature of the decision tree. So the use of NN is an alternative to the two methods. The NN method can be applied to solve complex non-linear problems. The same calculation accuracy ratio can be obtained with smaller data.

This study carried out research in classifying fault types by considering the composition of dissolved gas analysis in Transformer oil. The data are based on the use of a chromatographic approach. After the labelling process has been carried out, the next step is the classification process using a three-algorithm approach. These include Random Forest, Decision Tree, and NN and the validation process. This paper is organized as follows: dissolved gas analysis in Section 2. Section 3 describes materials and methods. Then, Section 4 reports a detailed experiment and research. At last, Section 5 concludes the entire paper.

4. SYSTEM DESIGN

The DGA dataset initialization process is carried out by calculating the proportion of gases associated with various disturbances. The average value is used to eliminate data, especially for data whose value is more significant than twice the deviation. There are several factors considered in the classification process. AUC is Area Under the Curve, CA is Classification Accuracy, F1 is a weighted average, Precision is a proportion of correct identifications, and Recall is related to the proportion of actual positives identified correctly. Each analysis can be calculated as follow.

$$F1 = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)} \tag{1}$$

Where

$$Precision = \frac{True}{True + False} \tag{2}$$

$$Recall = \frac{True\ Positive}{True\ Positive\ + False\ Negative} \tag{3}$$

4.1 Decision Tree

A decision tree algorithm is a popular tool in machine learning, where the tool can classify feature sets in a one-step decision step [18], [19]. This method works based on multistage schemes or hierarchical decisions. Trees consist of internal nodes and terminal nodes. Each node forms a binary decision, whether to merge or separate classes. The process carried out has a top-down approach. Internal nodes have a decision function to indicate which node to visit forward. In contrast, terminal nodes describe the learning scheme output as input from the vector.

The process starts with tree building and then pruning [20]. The procedure is based on binary recursive partitioning, and it is repeated to sort the data into partitions. All training data examples determine the tree structure [21]. The algorithm is then split using binary split and chooses to partition the data into two parts to reduce the deviation. The split process is used in the next branch. This process continues until each node reaches the user-specified minimum node size. This condition can also be explained as the number of training on each node. In the final stage, the node that has been formed is called a terminal node.

4.2 Random Forest

The Random Forest algorithm is a learning algorithm that works by combining vulnerable classifiers. The result will have higher accuracy and relatively stable performance [22-24]. The sequence in the random forest procedure begins with the Bootstrap resampling technique. In the bootstrap process, all samples are not used in making the decision tree. Samples that are not used are suspected as samples outside the assessment category. Each sample is started randomly to generate a new training sample. Then, each decision tree is generated to form a random forest. Furthermore, the average of the output of each tree is used to determine the final result. The other tree forms are evaluated based on the average final result. The architecture of the random forest method can be described in Figure 2.

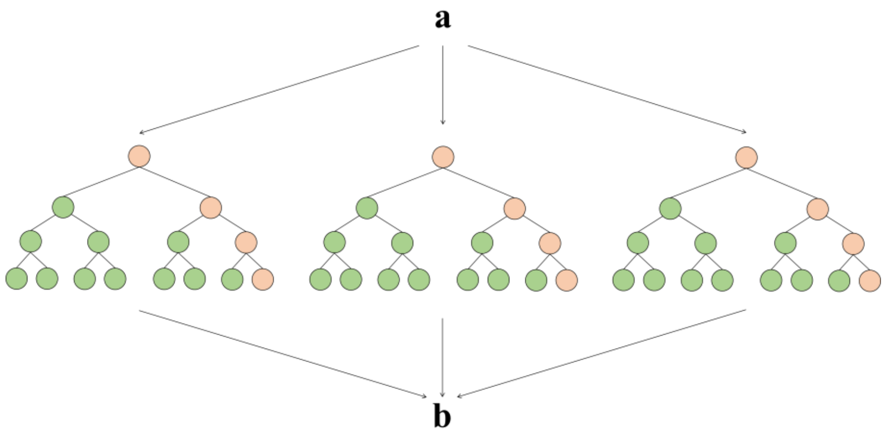

Figure 2. The architecture of the random forest method

4.3 Artificial NN

The Artificial NN (ANN) algorithm comprises several layers that are connected in a feed-forward manner. Each neuron is directly connected to the layers [25-28]. The algorithm determines the best layout of neurons and hidden layers to reduce the Mean Square Error (MSE). A correlation curve (R) between input and output data is also obtained using this method. The number of neurons (NN) per layer and the number of hidden layers (HL) serve as a starting point for various potential combinations, with the MSE

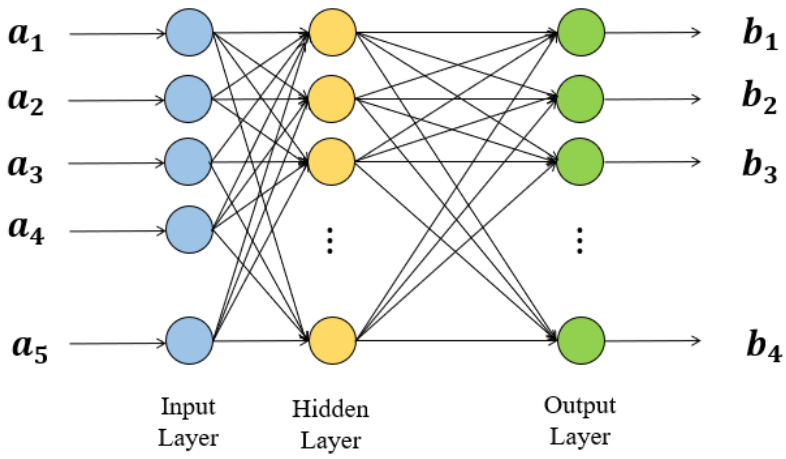

Figure 3. The architecture of the Artificial NN

eventually reaching zero. The R between the input and output data must be near one. The architecture of the Artificial NN is represented in Figure 3.

4.4 Data Preparation

Data preparation is essential in ensuring the data mining process success rate. The condition of DGA data was identified in the public literature. Furthermore, data variables for the classification process are obtained by screening the data. Low-temperature disturbance, medium temperature, high temperature when a fault occurs, partial discharge, arc discharge, spark discharge, and arc discharge are all covered in the classification process. Each part is grouped into several attributes H2, CH4, C2H6, C2H4, C2H2 in 201 data.

Table 1. Matrix of All Dataset of Transformer

No	H2	CH4	C2H6	C2H4	C2H2	Type
1	3930	2397	157	0	0	Partial Discharge
2	37800	1740	249	8	8	Partial Discharge
…	…	…	…	…	…	…
201	40000	400	70	600	6	Low Medium Temperature Overheating

Table 2. Matrix of Testing Dataset of Transformer

No	H2	CH4	C2H6	C2H4	C2H2	Type
1	36036	4704	554	5	10	Partial Discharge
2	33046	619	58	2	0	Partial Discharge
…	…	…	…	…	…	…
15	12	18	4	4	0	Low Medium Temperature Overheating

The number of testing datasets used is a total of 15 units. Tables 1 and 2 explain the matrix Transformer dataset and testing dataset. Furthermore, the classification algorithm setup, both with the Tree, Random Forest, and NN methods, can be described in tables 3 to 5. Parameters of each algorithm are obtained from several tests to get the smallest error for the processing procedure.

Table 3. Tree Decision Parameter Setup

Parameter	Value/Function
Instances Leaves	2, 5, 10
Limit Split Subset	2, 3
Maximal Tree Depth	1000
Classification Stop (%)	85

Table 4. Random Forest Parameter Setup

Parameter	Value/Function
Number of Trees	10, 20, 40
Attribute at each split	2
Limit depth individual trees	3, 10
Limit Split Subset	3

Table 5. NN Parameter Setup

Parameter	Value/Function
Neuron of Hidden Layers	1, 2, 3, 4, 5
Activation	Logistic
Solver	L-BFGS-B
Maximal Interation	230

5. EXPERIMENT AND ANALYSIS

The experimental setup of the fault diagnosis classification process, considering the diagnosis of DGA, showed in Figure 4. First, import the 201 datasets. Then perform a feature selection of the missing data. Fault detection is carried out to classify each disturbance into the Decision Tree, Random Forest, and ANN methods. Then evaluate to ensure the credibility of the proposed approach method is valid. Next, do testing and training and compare the methods used to determine the best method. The condition of DGA data was identified in the public literature [27].

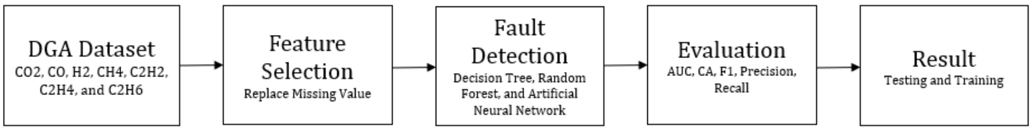

Figure 4. Experiment Setup of Classification Method in Fault Diagnosis by Considering Dissolved Gas Analysis

Table 6. Result of Classification Process by Using Neural Network

Model	Activation	Hidden Layer	AUC	CA	F1	Precision	Recall
Neural Network	Logistic	1	0.739	0.330	0.237	0.189	0.330
		2	0.891	0.800	0.740	0.742	0.800
		3	0.897	0.667	0.659	0.680	0.667
		4*	0.913	0.800	0.790	0.822	0.800
		5	0.913	0.733	0.680	0.669	0.733

Table 7. Result of Classification Process by Using Tree Method

Model	Min Number of Instances in Leaves	Do not Split subset smaller than	AUC	CA	F1	Precision	Recall
Decision Tree	2	2	0.810	0.600	0.566	0.569	0.600
	5	2	0.832	0.667	0.622	0.633	0.667
	10	2	0.829	0.467	0.444	0.500	0.467
	2	3	0.810	0.600	0.566	0.590	0.600
	5	3*	0.832	0.667	0.622	0.633	0.667

Table 8. Result of Classification Process by Using Random Forest

Model	Number of Trees	Limit Depth	AUC	CA	F1	Precision	Recall
Random Forest	10	3	0.821	0.667	0.610	0.569	0.667
	20	3	0.859	0.600	0.530	0.480	0.600
	40	3	0.848	0.600	0.541	0.500	0.600
	10	10	0.845	0.600	0.568	0.544	0.600
	20	10*	0.829	0.667	0.648	0.633	0.667

Model performance is determined by examining the area below the ROC curve. A good model has an AUC close to 1.0, which has a good separability measure. F1 score combines Recall and fits into a single performance metric, a weighted average of fit and Recall. Therefore, this score considers both false positives and false positives. F1 is usually more helpful than accuracy, especially with an uneven class distribution.

Detailed parameters between the three classification methods are explained in Figure 5 and Table 3-6 by exploring the variation of parameters in the classification process to solve the problem. The "*" sign indicates the best value from each experiment used to compare with other methods. The variations in the selection of hidden layers of NN method are used with Logistics activation. The iteration process remains at number 250. The highest CA and AUC value is obtained with 0.800 and 0.913, respectively, using four hidden layers. In Decision Tree, an experiment is conducted by comparing the number of instance leaves so that it can affect the split process. Max tree depth is 1000. The best CA was obtained using the Number of Instances in Leaves 5 and the split subset 3. In this experiment, the CA and AUC results were 0.667 and 0.832, respectively. The Random Forest method sets the number of trees, thus affecting the depth limit of individual trees and subsets. Subset not split smaller than 3. The calculation process produces the highest CA and AUC of 0.667 and 0.848, respectively. The number of trees is 20 with a depth limit of 10. Therefore, NN has the highest CA and AUC level from the various experiments compared to Decision Tree and Random Forest.

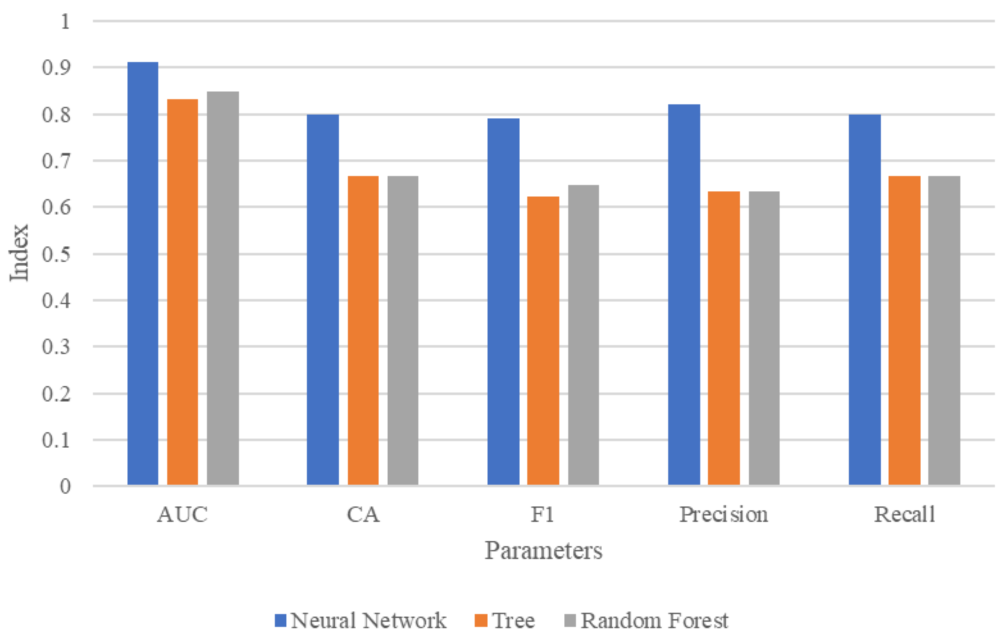

Figure 5. The Accurate Result between Three Classification Methods

The Random Forest method has a higher AUC value of 0.848 than 0.832 Random Forest. NN has the highest AUC at 0.913. By considering the confusion matrix, the highest score was obtained by NN, followed by Random Forest, Tree Method. They are 0.800, 0.667 and 0.667, respectively. In other words, the precision levels are 80%, 66.7%, and 66.7%. F1 in Decision Tree has the lowest value compared to other methods, which is 0.622, where the highest value is obtained on the NN at 0.79. The level of Precision and Recall is in the range of 0.6 to 0.8, where both Random Forest and Decision Tree have the lowest with the same value. A comparison of accuracy in classification between three methods can be described in Figure 6. The accurate Neural Network, Random Forest, and decision tree are 80.0%, 66.7%, and 66.7%, respectively. Overall, the NN method is more accurately used as a DGA classification method for transformer fault diagnosis than Tree and Random Forest.

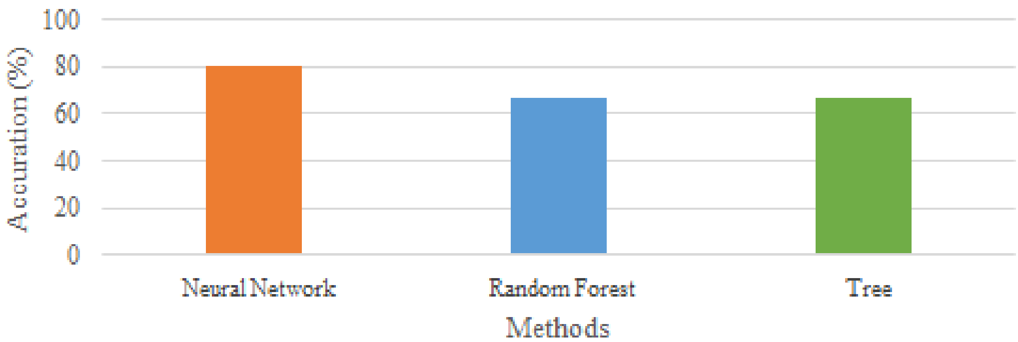

Figure 6. The Accuracy Result between Three Classification Method

6. CONCLUSION

The chromatographic representation of the oil is one of the essential factors in analyzing the dissolved gas in the transformer. Fault diagnosis in oil-immersed transformer was investigated using the classification method. The characteristics of several disturbances in DGA data can be informed more clearly through feature learning. Factors taken into consideration include the condition of the transformer when the temperature is low, medium temperature, high temperature when a fault occurs, partial discharge, arc discharge, spark discharge, and arc discharge. Based on the results of case studies, the accuracy of the correct rate of the NN method by considering CA is 0.800 and AUC is 0.913. This result follows with CA and AUC random forest method 0.667 and 0.848 and decision tree method 0.667 and 0.832, respectively. Therefore, the most accurate results are obtained from several comparative experimental methods using the NN method. Future research from the proposed model aims to develop and apply big data for power system devices, particularly in transformers fault analysis.

Acknowledgements

This research is financially supported by Institut Teknologi Sepuluh Nopember Surabaya.

REFERENCES

[1] **IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers***, IEEE Standard C57.104*, New York, NY, USA, 2009.

[2] **British Standards Institute Staff. Mineral Oil-Impregnated Electrical Equipment in Service-Guide to the Interpretation of Dissolved and Free Gases Analysis**, *IEC Standard 60599, Geneva, Switzerland*, 2015.

[3] H. M. Wilhelm, C. C. Santos and G. B. Stocco, **Dissolved gas analysis (DGA) of natural ester insulating fluids with different chemical compositions,** *IEEE Trans. Dielectr. Electr. Insul*., vol. 21, no. 3, pp. 1071-1078, June 2014.

[4] **Mineral Oil-Impregnated Electrical Equipment In Service-Guide To The Interpretation Of Dissolved And Free Gases Analysis**., *IEC Std. 60599*, 2015.

[5] M. D and L. Lamarre, **The Duval Pentagon-A New Complementary Tool For The Interpretation Of Dissolved Gas Analysis In Transformers**, *IEEE Electr. Insul. Mag*.,vol. 30, no. 6, pp. 9–12, 2014.

[6] M. Duval and A. de Pabla, **Interpretation Of Gas-In-Oil Analysis Using New Iec Publication 60599 and IEC Tc 10 Databases**, *IEEE Electr. Insul. Mag.*, Vol. 17, No. 2, pp. 31-41, 2001.

[7] N. Bakar, A. Abu-Siada, and S. Islam, **A review of dissolved gas analysis measurement and interpretation techniques**, *IEEE Elect. Insul. Mag.,* vol. 30, no. 3, pp. 39–49, May/Jun. 2014.

[8] H. C. Sun, Y. C. Huang, and C. M. Huang, **A review of dissolved gas analysis in power transformers**, *Energy Procedia*, vol. 14, pp. 1220–1225, Mar. 2012.

[9] A. Velasquez, M. Ricardo, M. Lara, and V. Jennifer, **Principal Components Analysis and Adaptive Decision System Based On Fuzzy Logic For Power Transformer**, *Fuzzy Inf. Eng.,* vol. 9, no. 4, pp. 493–514, 2017.

[10] H. Wei, Y. Wang, L. Yang, C. Yan, Y. Zhang, and R. Liao, **A New Support Vector Machine Model Based On Improved Imperialist Competitive Algorithm For Fault Diagnosis Of Oil-Immersed Transformers**, *J. Elect. Eng. Technol.,* vol. 12, no. 2, pp. 830–839, 2017.

[11] T. Kari, W. Gao, and D. Zhao, **Hybrid feature selection approach for power transformer fault diagnosis based on support vector machine and genetic algorithm,** *IET Gener. Transmiss. Distrib.,* vol. 12, no. 21, pp. 5672–5680, 2018.

[12] Seifeddine, S., Khmais, B., Abdelkader, C.: **Artificial Intelligence Toolsaided-Decision For Power Transformer Fault Diagnosis**, *Int. J. Comput. Appl.,* 38, pp. 1–8, 2012.

[13] Mohammed El A. S., Mostefa B., Issouf F., **Combining and Comparing Various Machine-Learning Algorithms To Improve Dissolved Gasanalysis Interpretation**, IET The Institution of Engineering and Technology, vol. 12 pp. 3673-3679, June, 2018.

[14] Yin, Y. Zhu and G. Yu, **Power Transformer Fault Diagnosis Based on Support Vector Machine with Cross Validation and Genetic Algorithm**, 2011 International Conference on Advanced Power System Automation and Protection, pp. 309-313, 2011.

[15] J. Watada, S. Chen, and Y. Yabuuchi, **A Rough Set Approach To Data Imputation And Its Application To A Dissolved Gas Analysis Dataset,** *Proc. IEEE Int. Conf. Comput. Meas. Control Sensor Netw*. pp. 24–27, May 2017.

[16] H.-C. Sun, Y.-C. Huang, and C.-M. Huang, **A review of dissolved gas analysis in power transformers**, *Energy Procedia*, vol. 14, pp. 1220–1225, 2012.

[17] T. Suwanasri, E. Chaidee, and C. Adsoongnoen, **Failure statistics and power transformer condition evaluation by dissolved gas analysis technique**, *International Conference on Condition Monitoring and Diagnosis*, CMD 2008, pp. 492–496, 2008.

[18] Witten, I., Frank, E., **Data mining: practical machine learning tools andtechniques with java implementations**, *Morgan Kaufmann Press*, SanFrancisco 2nd edn, CA, USA, March, 2005,

[19] S. Sardari, M. Eftekhari, and F. Afsari, **Hesitant Fuzzy Decision Tree Approach For Highly Imbalanced Data Classification**, *Appl. Soft Comput.,* vol. 61, pp. 727–741, Dec. 2017.

[20] V. T. Tran, B.-S. Yang, M.-S. Oh, and A. C. C. Tan, **Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference**, *Expert Syst. Appl.,* vol. 36, no. 2, pp. 1840–1849, March. 2009.

[21] Y. Cui, J. Shi, and Z. Wang, **Analog circuit fault diagnosis based on quantum clustering based multi-valued quantum fuzzification decision tree (QC-MQFDT),** *Measurement*, vol. 93, pp. 421–434, Nov. 2016.

[22] W. Deng, Y. Guo, J. Liu, Y. Li, D. Liu and L. Zhu, **A missing power data filling method based on improved random forest algorithm**, *Chinese Journal of Electrical Engineering*, vol. 5, no. 4, pp. 33-39, Dec. 2019.

[23] L Breiman. **Random forests**. *Machine Learning*, 45(1): 5-32, 2001.

[24] T. K. Ho. **The random subspace method for constructing decision forests.** *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 20(8): 832-844, Aug. 1998

[25] B. Yegnanarayana, **Artificial NNs**. *Prentice-Hall of India*, New York, NY, USA, 2009.

[26] Edgar A. J, Joselito M, Juan C. O, Juan C. S., Alejandro R., **Hot-Spot Temperature Forecasting of the Instrument Transformer Using an Artificial NN**, *IEEE Access*, vol. 8, pp. 164392-164406, Aug. 2020.
[27] G. Chakraborty and B. Chakraborty, **A novel normalization technique for unsupervised learning in ANN**, *IEEE Trans. Neural Netw.*, vol. 11, no. 1, pp. 253–257, Jan. 2000.
[28] E. Li, L. Wang and B. Song, **Fault Diagnosis of Power Transformers With Membership Degree**, *IEEE Access*, vol. 7, pp. 28791-28798, March. 2019.