Fault Identification in Power Transformers Using Dissolve Gas Analysis and Support Vector Machine

Hazlee Azil Illias
Department of Electrical Engineering
Centre of Advanced Manufacturing & Material Processing (AMMP Centre)
Faculty of Engineering
Universiti Malaya
Kuala Lumpur, Malaysia
h.illias@um.edu.my

Chan Kai Choon
Department of Electrical Engineering
Faculty of Engineering
Universiti Malaya
Kuala Lumpur, Malaysia
kcchan93@siswa.um.edu.my

Wee Zhao Liang
Department of Electrical Engineering
Faculty of Engineering
Universiti Malaya
Kuala Lumpur, Malaysia
zlwee@siswa.um.edu.my

Hazlie Mokhlis
Department of Electrical Engineering
Faculty of Engineering
Universiti Malaya
Kuala Lumpur, Malaysia
hazli@um.edu.my

Azrul Mohd Ariffin
Department of Electrical & Electronic Engineering
College of Engineering
Universiti Tenaga Nasional
Kajang, Selangor, Malaysia
azrula@uniten.edu.my

Mohd Fairouz Mohd Yousof
Faculty *of Electrical and Electronic Engineering*
Department of Electrical Engineering
Universiti Tun Hussein Onn
Batu Pahat, Johor, Malaysia
fairouz@uthm.edu.my

***Abstract*—Transformer faults need to be identified accurately at the early stage in order to ease the maintenance of power transformer, reduce the cost of maintenance, avoid severe damage on transformer and extend the lifespan of transformer. Dissolved Gas Analysis (DGA) is the most commonly used method to identify the transformer fault in power system. However, the existing transformer fault identification methods based on DGA have a limitation because each method is only suitable for certain conditions. Thus, in this work, one of the artificial intelligence techniques, which is Support Vector Machine (SVM), was applied to determine the power transformer fault type based on DGA data. The accuracy of the SVM was tested with different ratio of training and testing data. Comparison of the results from SVM with artificial neural network (ANN) was done to validate the performance of the system. It was found that fault identification in power transformers based on DGA data using SVM yields higher accuracy than ANN. Therefore, SVM can be recommended for the application of power transformer fault type identification in practice.**

Keywords— Support vector machine, power transformer, fault diagnosis

I. INTRODUCTION

Very often, transformer oils are subject to electrical and thermal stresses when a transformer is operating. These stresses will break the bond of hydrocarbon molecules, releasing a large number of hydrogen and carbon atoms. They will then combine with each other to form various types of gaseous product. For examples, hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), ethane (C_2H_6) and carbon monoxide (CO) [1]. This will then render its ineffectiveness to serve its purpose. Faults that are normally associated with the release of gases are partial discharge, arcing and overheating [2]. Hence, regular monitor on transformer oil shall be conducted to ensure that it is adequate for further service.

Several works on fault identification in power transformers using dissolve gas analysis and artificial intelligence have been performed in the past. J. Faiz and M. Soleimani compared various conventional methods such as Roger ratio method, Duval triangles, IEC 60599, Mansour pentagon and Duval pentagon in term of consistency [3]. They stated that consistency is a reliable indicator to standardize the accuracy of different conventional methods because it does not depend on the total number of cases in each fault. The accuracy of SVM is affected by two factors; kernel function and training samples [4]. Hence, the option of kernel function is essential to SVM. The results show that Gaussian function performs better than polynomial functions. In this work, Multi-Layer Perceptron (MLP), fuzzy approach and radial basis function (RBF) were used to compare with SVM to determine the type of transformer faults. SVM shows the highest diagnostic accuracy than others AI approaches.

Seven types of transformer faults were determined by using ANN in [5]. The work concluded ANN is an appropriate method to predict power transformer fault. In [6], SVM-particle swarm optimization (PSO) was employed to forecast gases content in transformer oil. The result demonstrated SVM-PSO is the best method in forecasting dissolved gases content compared to grey model and ANN under the condition of small sampling data. In [7], ANN and various PSO techniques were used to predict transformer incipient fault. The work concluded that combination of ANN with evolutionary PSO yields the highest accuracy of 98% in predicting power transformer fault.

In [8], improved Imperialist Competitive Algorithm (IICA) combined with SVM was proposed to analyze transformer faults. The proposed method, IICA-SVM, performed better than other diagnosis approaches with the highest accuracy of 92.59%. In [9], SVM and ANN were optimized by Gravitational Search Algorithm (GSA) to enhance the accuracy of the classification of transformer fault type. results. Feature selections using stepwise regression and robust regression were applied to utilize only significant gases. It was observed that the proposed hybrid feature selection-artificial intelligence-gravitational search algorithm technique yields reasonable accuracy although fewer types of dissolved gases were used.

II. METHODOLOGY

This project utilized 6 types of gases; hydrogen (H_2), acetylene (C_2H_2), methane (CH_4), carbon monoxide (CO), ethane (C_2H_6) and ethylene (C_2H_4) with 100 input data each and the output was categorized into 4 different classes, namely low intensity and high intensity, thermal fault and no fault. The output data were classified into 16 low intensity cases, 16 high intensity cases, 18 thermal fault cases and 50 no fault cases. These data were divided into training and testing sets. From 100 data sets, 70% data were used as training data and 30% data were used as testing data. To minimize data usage for training of artificial intelligence, the ratio of training data and test data was reduced from 70:30 to 50:50. Table 1 shows the input and output data used for classification.

TABLE I. INPUT AND OUTPUT DATA USED FOR CLASSIFICATION

Input data	Output data
Hydrogen (H_2) Acetylene (C_2H_2) Methane (CH_4) Carbon monoxide (CO) Ethane (C_2H_6) Ethylene (C_2H_4)	Discharge of low intensity Discharge of high intensity Thermal fault No fault

The input and target data of the gas compositions with respect to the transformer fault were obtained from DGA analysis. The gas compositions were used as the input data and the transformer faults were used as the target data. The algorithm started with loading the training data for training purpose. Then, all training data were normalized to ease the step of classification. After that, SVM was trained.

Originating from Vapnik and Chervonenkis, SVM is a discriminating classifier used widely in statistical learning. Classification and data prediction are its main functions. SVM is a powerful approach to deal with the problem with small amounts of training data and large amounts of input data and non-linear data set. Hence, it has the ability to cope with large feature spaces, making it optimal for large amounts of classification data while still achieving high accuracy and efficiency in various fields ranging from biomedical to image classification. SVM develops a separating hyperplane in high dimensional space for classifying data by maximizing the margin data points or support vectors as shown in Fig. 1. Support vectors and training data are represented by filled circles and unfilled circles respectively.

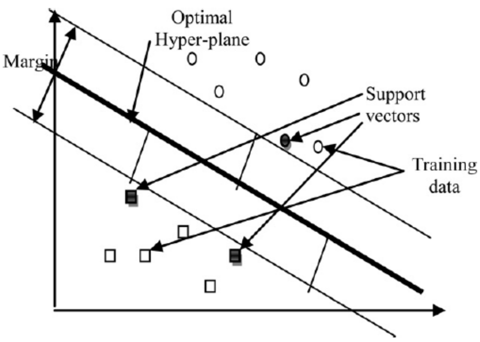

Fig. 1. Separation of two classes by SVM [4]

The accuracy of SVM is affected by two factors; kernel function and training samples. Linear Kernel function, polynomial kernel function, Gaussian radial basis function and sigmoid kernel function are the common kernel functions used in SVM. Through supervised learning, labelled training data are used with kernel functions that outputs an optimal hyperplane and is able to classify non-linearly separable data inputs. This project utilizes (Gaussian) radial basis function (RBF) kernels for multi category classifications. The RBF kernel is defined as

$$K(x,x') = \exp\left(-\frac{||x-x'||^2}{2\sigma^2}\right) \quad (1)$$

Support vectors are represented by x and x', σ is a RBF kernel parameter, which needs to be determined in optimization method. c is a misclassification parameter, which is another parameter needs to be determined. Both σ and c are important parameter used to control the classification performance and accuracy of SVM. Thus, optimization techniques are applied in this work to seek out the optimize value of these parameters to achieve desirable outcome. Trade-off between size of slack variables and margin is controlled or determined by misclassification parameter or c. A large c value produces a smaller-margin hyperplane but classifies all the training points correctly. It causes SVM classifies the training data stricter and induce overfitting. However, a small c value generates a larger-margin hyperplane trade off with more slack points are appearing in condition where underfit will happen.

Kernel function was used in SVM to seek the optimal hyperplane that is able to classify non-linearly separable data inputs. This project utilized (Gaussian) radial basis function (RBF) kernel for multi category classifications. Hence, the kernel parameters σ and misclassification parameter c are needed to be optimized by the optimization methods to avoid the overfitting and underfitting process in SVM.

After finding the optimized parameters, multi-layer SVM classifiers were built for classification purpose. To identify four groups of faults, low intensity and high intensity, thermal fault and no fault, three SVMs; SVM-1, SVM-2 and SVM-3 were built. The output of SVM-1 is set to +1 when the fault detected is normal case, otherwise -1. SVM-2 classifies all fault cases into thermal faults and discharge faults. Output SVM-2 is set to +1 if the fault belongs to thermal fault; otherwise -1. Discharge faults are classified into low intensity and high intensity by SVM-3. Output SVM-3 is set to +1 if the fault is low intensity, otherwise -1.

Testing data were used to test the trained SVM. The output of the trained SVM was compared with the target output to determine the accuracy of SVM. After that, the accuracy of each different parameters was compared to sort out the best accuracy of the algorithm.

To compare the performance of SVM, ANN was also used to classify the fault type in power transformers based on DGA data. The performance of ANN model is affected by input, output and network topology. There are two stages of designing ANN model; training and testing stages. Firstly, the algorithm starts with loading the training data for training purpose. Then, all the training data was normalized to ease the step of classification. After that, it goes to training stage of ANN.

III. RESULTS AND DISCUSSION

Table 2 to 4 shows the classification accuracy results of transformer fault type based on DGA data using SVM based on different ratio of training to testing data. From these tables, it can be seen that when the values of c and σ were varied randomly for different training:testing data, the classification accuracy does not change much. However, when a comparison is made between different training:testing data as shown in Table 5, the average classification accuracy decreases slightly when the training data are reduced but testing data are increased.

Also, referring to Table 5, comparison between SVM and ANN shows that when the ratio of training to testing data is reduced, the classification accuracy obtained by SVM is always higher than ANN. This is due to the performance of ANN is not optimized with suitable number of hidden layers and number of neurons in each hidden layer.

TABLE II. CLASSIFICATION RESULTS OF SVM USING 70:30 TRAINING:TESTING DATA

c	*σ*	**Error (%)**	**Accuracy (%)**
8.8644	0.7948	0.00	100.00
8.2724	0.3682	0.00	100.00
5.7985	0.9581	0.33	99.67
0.663	5.1681	0.33	99.67
1.5638	7.3747	0.33	99.67
5.9659	0.1085	0.00	100.00
6.5825	0.708	0.33	99.67
1.3853	0.1285	0.00	100.00
6.4089	0.3042	0.33	99.67
5.9134	0.9742	0.00	100.00
0.5744	0.1304	0.00	100.00
0.955	6.4756	0.33	99.67
3.4007	0.7205	0.00	100.00
1.4923	8.1359	0.33	99.67
1.6861	8.2138	0.33	99.67
1.2808	7.7179	0.33	99.67
1.2659	7.6929	0.00	100.00
7.9992	0.7228	0.00	100.00

TABLE III. CLASSIFICATION RESULTS OF SVM USING 60:40 TRAINING:TESTING DATA

c	*σ*	**Error (%)**	**Accuracy (%)**
6.0777	2.6772	0.00	100.00
5.9104	3.1401	2.50	97.50
7.1107	3.3251	0.00	100.00
7.1871	6.5118	0.25	99.75
9.5144	8.8366	0.25	99.75
7.9773	7.2465	0.25	99.75
7.4491	1.8442	0.50	99.50
7.4771	1.1942	0.00	100.00
7.3375	4.3701	2.50	97.50
7.2102	5.4762	0.50	99.50
4.6972	5.2599	0.25	99.75
9.2938	8.32	0.50	99.50
9.0877	5.385	0.50	99.50
8.7904	6.7007	0.25	99.75
6.3255	2.8789	0.00	100.00
8.6262	7.4823	0.50	99.50
6.8834	2.7051	0.00	100.00
8.7318	4.9231	0.50	99.50

TABLE IV. CLASSIFICATION RESULTS OF SVM USING 50:50 TRAINING:TESTING DATA

c	*σ*	**Error (%)**	**Accuracy (%)**
8.255	8.2197	2.00	98.00
5.3158	9.3768	2.00	98.00
8.3928	9.0573	2.00	98.00
6.8747	7.5908	4.00	96.00
7.7443	3.502	2.00	98.00
9.7088	8.1214	4.00	96.00
8.8654	9.0504	4.00	96.00
8.7672	9.453	2.00	98.00
5.3644	6.2654	4.00	96.00
7.0395	9.3178	4.00	96.00
8.4454	9.0278	2.00	98.00
9.1463	8.8867	2.00	98.00
2.3044	7.0808	2.00	98.00
3.2076	5.5518	2.00	98.00
5.8486	9.0171	2.00	98.00
6.5024	6.1426	2.00	98.00
2.923	5.5641	2.00	98.00
5.1726	7.2809	4.00	96.00

TABLE V. AVERAGE CLASSIFICATION ACCURACY USING DIFFERENT INPUT:OUTPUT DATA RATIO

AI Method	**Ratio**		
	70:30	**60:40**	**50:50**
SVM	99.84	99.49	97.33
ANN	97.33%	94.33%	78.6%

IV. CONCLUSIONS

Classification of power transformer fault type based on dissolved gas analysis (DGA) data has been successfully performed using support vector machine (SVM). From the results obtained, the classification accuracy of fault identification in power transformers based on DGA data using SVM yields reasonable accuracy under different ratio of training and testing data. From the comparison of the results from SVM with artificial neural network (ANN), SVM yields higher classification accuracy than ANN when different training to testing data ratio. It can be recommended that SVM can be used for the application of power transformer fault type identification based on DGA data in the actual practice.

ACKNOWLEDGMENT

The authors thank the Malaysian Ministry of Education and Universiti Malaya for supporting this work through research grant of HIR (H-16001-D00048), Faculty Research Grant (GPF077A-2018), Impact-Oriented Interdisciplinary Research Grant (IIRG001-2020IISS) and UM Living Labs Grant (LL046-2019), and Centre of Advanced Manufacturing & Material Processing (AMMP Centre).

REFERENCES

[1] A. Singh and G. Upadhyay, "Dissolved Gas Analysis of power transformer using K-means and Support Vector Machine," International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, pp. 1-5, 2016.

[2] S. Koroglu, "A Case Study on Fault Detection in Power Transformers Using Dissolved Gas Analysis and Electrical Test Methods" J. Electrical System, vol. 12, no. 3, pp. 442-459, 2016.

[3] J. Faiz and M. Soleimani, "Dissolved Gas Analysis Evaluation in Electric Power Transformers using Conventional Methods" University of Tehran Iran, vol. 24, no. 2, April 2017.

[4] K. Bacha, S. Souahlia, M. Gossa "Power transformer fault diagnosis based on dissolved gas analysis by support vector machine" Electric Power System Research, vol. 83, pp. 73-79, Feb 2012.

[5] F. Zakaria, D. Johari and I. Musirin, "Artificial neural network (ANN) application in dissolved gas analysis (DGA) methods for the detection of incipient faults in oil-filled power transformer," 2012 IEEE International Conference on Control System, Computing and Engineering, Penang, pp. 328-332, 2012.

[6] S. Fei, M. Wang, Y. Miao, J. Tu and C. Liu, 'Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil', Energy Conversion and Management, vol. 50, no. 6, pp. 1604-1609, 2009.

[7] H. Illias, X. Chai, A. Abu Bakar and H. Mokhlis, 'Transformer Incipient Fault Prediction Using Combined Artificial Neural Network and Various Particle Swarm Optimisation Techniques', PLOS ONE, vol. 10, no. 6, p. e0129363, 2015.

[8] Y. Zhang, H. Wei, R. Liao, Y. Wang, L. Yang and C. Yan " A New Support Vector Machine Model Based on Improved Imperialist Competitive Algorithm for Fault Diagnosis of Oil-immersed Transformers" J. Electr Eng Tech, vol. 12, no.2, pp 830-839, 2017.

[9] H.A. Illias, K.C. Chan, H. Mokhlis, "Hybrid feature selection–artificial intelligence–gravitational search algorithm technique for automated transformer fault determination based on dissolved gas analysis," IET Generation, Transmission & Distribution, Vol. 14, No. 8, pp. 1575-1582, 2020.